



**TÜRKİYE CUMHURİYETİ
ADANA ALPARSLAN TÜRKES SCIENCE AND TECHNOLOGY
UNIVERSITY**

**GRADUATE SCHOOL
INDUSTRIAL ENGINEERING DEPARTMENT**

**ESTIMATION OF GLOBAL INNOVATION INDEX SCORES WITH
MACHINE LEARNING**

RABİA SULTAN YILDIRIM

Ph.D.



**TÜRKİYE CUMHURİYETİ
ADANA ALPARSLAN TÜRKES SCIENCE AND TECHNOLOGY
UNIVERSITY**

**GRADUATE SCHOOL
INDUSTRIAL ENGINEERING DEPARTMENT**

**ESTIMATION OF GLOBAL INNOVATION INDEX SCORES WITH
MACHINE LEARNING**

RABİA SULTAN YILDIRIM

Ph.D.

**THESIS ADVISOR
PROF. DR. NEŞE YALÇIN**

**SECOND ADVISOR
ASSOC. PROF. DR. MURAT OTURAKÇI**

ADANA, 2024

ÖZET

MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE KÜRESEL İNOVASYON ENDEKSİNİ TAHMİNLEME

Rabia Sultan YILDIRIM

Doktora, Endüstri Mühendisliği

Danışman: Prof. Dr. Neşe YALÇIN

II. Danışman: Doç. Dr. Murat OTURAKÇI

Haziran 2024, 196 sayfa

Günümüzün dinamik küresel ekonomisinde inovasyon, büyüme ve başarıyı yönlendirmede hayati bir rol oynamaktadır. İnovasyonu ölçmek zor bir süreçtir, ancak Küresel İnovasyon Endeksi (KİE), inovasyonun temel göstergelerini belirleyen değerli bir araç olarak öne çıkmaktadır. Ancak, KİE'yi hesaplama süreci karmaşık ve zaman alıcıdır, bu da önemli bir zorluk teşkil etmektedir. Bu çalışmada, KİE'yi hesaplamak için gereken parametre sayısını azaltmak amacıyla bir özellik seçme yöntemi kullanılmıştır. 2012-2022 yılları arasında 125 ülkenin KİE puanlarını tahmin etmek için yapay sinir ağı (YSA) yöntemleri ve çeşitli regresyon teknikleri uygulanmıştır. Bulgular, seçilen 38 parametrenin KİE puanını doğru bir şekilde hesaplamak için yeterli olduğunu ve YSA modelinin en başarılı tahminleme yöntemi olduğunu göstermektedir. KİE puanlarını tahmin etmek için bir prototip arayüz tasarlanmış ve bu da tahmin sürecini daha kolay ve verimli hale getirmiştir. Bu çalışma, parametre sayısını azaltarak ve tahmin doğruluğunu artırarak araştırma literatürüne önemli bir katkı sağlamaktadır; bu da KİE'yi hesaplama sürecini basitleştirerek dünya genelindeki kuruluşlar ve politika yapıcılar için daha erişilebilir hale getirebilir.

Anahtar Kelimeler: küresel inovasyon endeksi ,öznitelik seçimi, makine öğrenmesi

ABSTRACT

ESTIMATION OF GLOBAL INNOVATION INDEX SCORES WITH MACHINE LEARNING

Rabia Sultan YILDIRIM

Ph.D., Department of Industrial Engineering

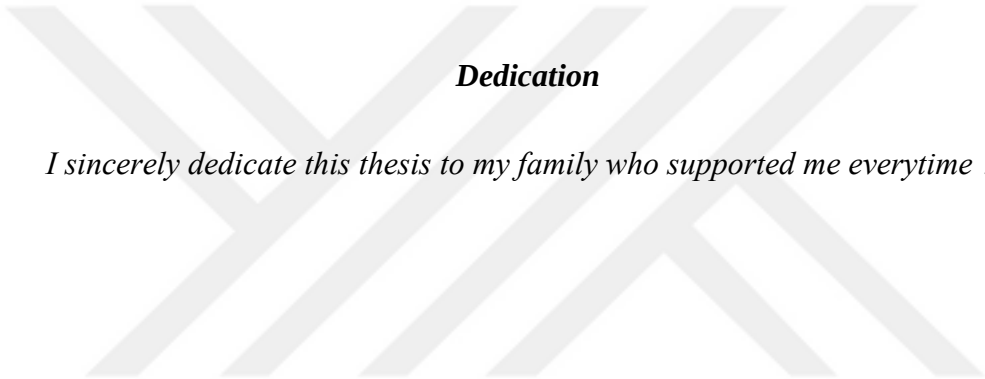
Supervisor: Prof. Dr. Neşe YALÇIN

Co-Supervisor: Assoc. Prof. Dr. Murat OTURAKÇI

June 2024, 196 pages

In today's dynamic global economy, innovation plays a critical role in driving growth and success. Measuring innovation is challenging, but the Global Innovation Index (GII) is a valuable tool identifying essential indicators of innovation. However, the process of calculating the GII is complex and time-consuming, which poses a significant challenge. To streamline the process, this study focused on reducing the number of parameters required to calculate the GII using a feature selection method. The study used an artificial neural network (ANN) methods and various regression methods to estimate GII scores for 125 countries from 2012-2022. The study's findings suggest that the selected 38 parameters were sufficient to accurately calculate the GII score, and the ANN model was the most precise method. A prototype interface was designed to predict GII scores, allowing for easier and more efficient estimation. This study makes an important contribution to the research literature by reducing the parameter count and improving estimation accuracy, which could simplify the process of calculating the GII and make it more accessible to organizations and policymakers worldwide.

Keywords: global innovation index, feature selection, machine learning



Dedication

I sincerely dedicate this thesis to my family who supported me everytime ...

ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Dr. Neşe YALÇIN and Assoc. Prof. Dr. Murat OTURAKÇI for their inspiring and guiding presence, and for their expertise and experience which have always helped me during my doctoral studies.

I also want to thank my friend Lecturer Yiğit ALİŞAN for his support and assistance during the interface preparation stages of my thesis.

Lastly, I am grateful to my friend Fide SEVGİ for her help and support in organizing the formal aspects of my thesis.

TABLE OF CONTENTS

CERTIFICATION OF APPROVAL

Hata! Yer işareti tanımlanmamış.

ÖZET **i**

ABSTRACT **ii**

Dedication **iii**

TABLE OF CONTENTS **v**

LIST OF FIGURES **viii**

LIST OF TABLES **ix**

LIST OF ABBREVIATIONS **x**

LIST OF SYMBOLS **xii**

1. INTRODUCTION **1**

1.1. Innovation 1

1.1.1. Types of Innovation 4

1.1.1.1. Product Innovation 4

1.1.1.2. Process Innovation 5

1.1.1.3. Marketing Innovation 6

1.1.1.4. Organizational Innovation 7

1.1.2. Indexes Measuring Innovation Performance 8

1.1.3. Information Systems and Analytics 9

1.1.4. Decision Support System 11

1.1.4.1. Artificial Intelligence 13

1.1.4.2. Machine Learning 16

1.2. Global Innovation Index 18

1.2.1. The Innovation Input Sub-Index 20

1.2.1.1. Institutions 20

1.2.1.2. Human capital and research 22

1.2.1.3. Infrastructure 25

1.2.1.4.	Market sophistication	28
1.2.1.5.	Business Sophistication	31
1.2.2.	The Innovation Output Sub-Index	34
1.2.2.1.	Knowledge and technology outputs	35
1.2.2.2.	Creative outputs	38
1.3.	Aim and Structure of the Study	42
2.	LITERATURE REVIEW	44
3.	MATERIAL AND METHOD	58
3.1.	Materials	58
3.2.	Methods	61
3.2.1.	Classification Method	61
3.2.1.1.	Decision Trees	62
3.2.1.2.	Ensemble Methods	66
3.2.1.3.	Support Vector Machines	69
3.2.1.4.	Linear and Logistic Regression	70
3.2.1.5.	Artificial Neural Networks	75
3.2.1.6.	K-Nearest Neighbors	77
3.2.1.7.	Others	79
3.2.2.	Feature Selection Method	83
3.2.3.	Artificial Neural Network Method	92
3.2.4.	Regression Methods	100
3.2.4.1.	Linear Regressions	101
3.2.4.2.	Regression Trees	105
3.2.4.3.	Support Vector Machine	106
3.2.4.4.	Kernel Approximation Models	109
3.2.4.5.	Ensembles of Regression Trees	111
3.2.4.6.	Neural Network Regression	114

3.2.5. Decision Support System Method	118
4. FINDINGS AND DISCUSSION	121
4.1. Findings	121
4.1.1. Classification Findings	121
4.1.2. Feature Selection Findings	122
4.1.3. Repeated Classification Results Based on Feature Selection Findings	123
4.1.4. Artificial Neural Network Findings	125
4.1.5. Regression Findings	128
4.1.6. The Comparison Findings of the Artificial Neural Network and Regression Findings	131
4.1.7. Decision Support System Application Findings	132
4.2. Discussion	138
5. CONCLUSION	168
6. RECOMMENDATIONS	170
REFERENCES	171
APPENDIX	195
CURRICULUM VITAE	196

LIST OF FIGURES

Figure 1.1. A Framework from model-driven to knowledge-driven DSS.	12
Figure 1.2. The AI Applications.....	14
Figure 1.3. Various types of ML techniques	17
Figure 1.4. The Framework of the GII	20
Figure 3.1. The Methodology of This Research.....	61
Figure 3.2. Classification Techniques	62
Figure 3.3. Feature selection methodology applied in this study	92
Figure 3.4. General structure of an ANN with two hidden layers	93
Figure 3.5. ANN methodology applied in the study	98
Figure 3.6. LR methodology applied in the study	117
Figure 3.7. A flow diagram of the DSS application.....	119
Figure 4.1. ANN Network.....	125
Figure 4.2. The Regression Plot of Levenberg-Marquardt Model	128
Figure 4.3. Predicted and True Response Plot	130
Figure 4.4. Interface Screens Hierarchy	133
Figure 4.5. Main screen to predict the GII score.....	134
Figure 4.6. Data entry screen	135
Figure 4.7. The error displayed on the data entry screen	135
Figure 4.8. The prediction result screen	136
Figure 4.9. 2023 GII Real and Predicted Score Comparison	137

LIST OF TABLES

Tablo 2.1. The Summary of the Studies about GII	55
Tablo 3.1. Indexes Codes and Names	58
Tablo 3.2. The Number of Countries and Parameters.....	60
Tablo 3.3. Regression Models.....	101
Tablo 4.1. Classification Results.....	121
Tablo 4.2. Feature Selection Results.....	122
Tablo 4.3. Repeated Classification Results.....	123
Tablo 4.4. The error values of the hybrid of CfsSubsetEval and Best First method.....	124
Tablo 4.5. Selected Parameters	124
Tablo 4.6. ANN Prediction Performance Results	126
Tablo 4.7. The Results of Regression Models	129
Tablo 4.8. The Performance Results of the Best Two Methods	131
Tablo 4.9. Selected and Not Selected Parameters	138
Tablo 4.10. Percentage of Selected Parameters	141
Tablo 4.11. The Related Studies of the Selected Parameters.....	163

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
ANN	: Artificial Neural Network
BEA	: Bureau of Economic Analysis
BIC	: Bayesian Information Criterion
BRIC	: Brazil, Russia, India and China
CART	: Classification and Regression Tree
ccTLD	: Country-Code Top-Level Domain
CSE	: ClassifierSubsetEval
DSS	: Decision Support System
EBOPS	: Extended Balance of Payments Services
EIS	: European Innovation Scorecard
EI	: E-Participation Index
EPI	: Environmental Performance Index
ES	: Expert Systems
FDI	: Foreign Direct Investment
fsQCA	: Fuzzy Set Qualitative Comparative Analysis Approach
FTE	: Full-Time Equivalent
GCI	: Global Competitiveness Index
GII	: Global Innovation Index
IBk	: Instance-Based k-Nearest Neighbors
INSEAD	: European Institute of Business Management
IS	: Information System
ISCO	: International Standard Classification of Occupations
ISIC	: International Standard Industrial Classification
K-NN	: K-Nearest Neighbors
KRR	: Kernel Ridge Regression
LPI	: Logistics Performance Index
LWL	: Locally Weighted Learning
MABAC	: Multi-Attributive Border Approximation area Comparison
MAE	: Mean Absolute Error
ML	: Machine Learning

MLP	: Multilayer Perceptron
MSE	: Mean Squared Error
MSTI	: Main Science and Technology Indicators
NRI	: Network Readiness Index
OECD	: Organization for Economic Cooperation and Development
PCT	: Patent Cooperation Treaty
PISA	: Programme for International Student Assessment
PPP	: Purchasing-Power-Parity
R&D	: Research And Development
RAE	: Relative Absolute Error
REPTree	: Reduced Error Pruning Tree
RMSE	: Root Mean Square Error
RRSE	: Root Relative Squared Error
SAARC	: South Asian Association for Regional Cooperation
SCG	: Scaled Conjugate Gradient
SITC	: Standard International Trade Classification
SVM	: Support Vector Machines
TLD	: Top-Level Domain
WEF	: World Economic Forum
WIPO	: World Intellectual Property Organization
WTO	: World Trade Organization

LIST OF SYMBOLS

R	: regression analysis
N	: Number
m	: number of features
α_i and α_j	: Lagrange multipliers
b	: bias term
E_i	: error
K_{ij}	: kernel values
x_i and x_j	: training instances
$x_1, x_2, \dots, x_n$	: independent variables
y	: dependent variable
$\beta_0, \beta_1, \dots, \beta_n$	: coefficients (or parameters) of the model
ϵ	: error term
f_i	: smooth function of the predictor variable X_i
k	: number of nearest neighbors
Merit(S)	: merit score
S	: <i>Subset</i>
AvgPerf	: average performance metric
Subset_i	: Performance
w_k	: weight vector
J_k	: Jacobian matrix
λ	: Levenberg-Marquardt parameter
I	: identity matrix
e_k	: error
k	: iteration number
$P(w D)$	: weight distribution given the data
$P(D w)$	: likelihood of the data under the weights
$P(w)$	: prior distribution of the weights
$P(D)$	: probability of the data
w_{k+1}	: updated weights
a_k	: step size

$f(x)$	: <i>function</i>
θ	: parameters of the neural network
$L(y, \hat{y})$	: loss function



1. INTRODUCTION

The thesis is divided into several chapters, each of which serves a distinct purpose. The first chapter provides an overview of the thesis and its purpose and scope. It introduces the basic information that the readers need to understand the topic in the following chapters. In the second chapter, existing studies in the literature are discussed, and the contribution that this thesis aims to make to the field is highlighted. This chapter also discusses the gaps in the literature that this thesis aims to address. The subsequent chapter presents an elaborate account of the dataset employed in the thesis. It elaborates on the pre-processing steps that were taken to create the data set, the feature selection methods that were employed, and the estimation methods used. This chapter also discusses the challenges faced during the data collection process and how they were overcome. In the fourth section, the findings obtained by classifying and selecting the data set are explained in detail. The prediction success of machine learning (ML) algorithms is presented, and the results of the thesis applications are discussed. This section also compares the steps taken in the thesis with the literature and offers suggestions for future research. Finally, the thesis proposes a prototype DSS for calculating the Global Innovation Index (GII) score. The prototype DSS is designed to automate the GII calculation process and provides a user-friendly interface for users to input data. The prototype DSS 's features and functionality are presented, and potential future developments are discussed.

1.1. Innovation

The concept of innovation was first evaluated by Schumpeter (1934). Schumpeter (1939) developed the concept of innovation based on the notions of creative destruction and entrepreneurship, building on his studies on economic development theory.

The word 'innovation' has its roots in the Latin word 'innovatus', signifying the introduction of new methods and ways for development. In today's world, where competition is on the rise, innovation has become a buzzword in the marketplace. Its importance cannot be overstated, especially for businesses (Alkan, 2014). As competition and the number of products and services increase, innovation provides a way for companies to set themselves apart and grow without being hindered. This holds true not only for established businesses but also for startups and small enterprises looking to survive and thrive in a competitive environment.

As technology continues to advance in this century, innovation has become a critical factor for the success of businesses, the competitiveness of organizations, and the growth and strengthening of the economy. To survive and thrive, businesses must constantly innovate in a sustainable manner. This means creating new opportunities and resources for generating wealth, as stated by Drazin (1985), and gaining an edge in the competitive landscape by developing new technologies and business methods, as described by Porter (1985).

In today's global economy, innovation plays a crucial role in determining a country's level of development. The success of innovation is closely linked to the concept of entrepreneurship, and it is essential to give entrepreneurs the necessary support and attention to achieve success. The benefits of innovation are not limited to the creation of new products or goods, but also extend to innovative services. A prime example is Yemeksepeti.com, Turkey's first online food delivery service, which represents a significant innovation in the service industry (Demirkaya and Zengin, 2014). As businesses continue to innovate, they can expect to reap greater profits and drive economic growth.

Innovation presents a substantial competitive advantage for both developed and developing nations. As businesses within a country prosper and generate profits, the economy benefits greatly. It catalyzes organizations to remain relevant in the market, stay abreast of advancements, and keep pace with the ever-evolving technological landscape. In essence, innovation entails the creation of novel products and services, which are effectively introduced to the market. Henceforth, in the present and future, innovation will become a ubiquitous aspect of organizational life across all sectors (Akbey, 2014).

Innovation encompasses a range of commercial activities, including production design and management, utilizing new equipment and processes for commercial purposes, with the goal of introducing new products and services to the market (Cem and Keskin, 2013). The key consideration is that any new product or service must be sustainable and proprietary to the company. With this in mind, a business can maintain a strong foothold in the market and effectively compete with other businesses. In today's rapidly evolving technological landscape, innovation is critical for organizations to remain relevant and competitive.

In today's global competitive landscape, businesses must adapt and evolve their production methods to ensure their continued success. This ongoing process of change and improvement is known as innovation (Tüysüz, 2020). Innovation can take many forms, from the introduction of new products or services to the implementation of novel organizational or marketing methods. The key requirement for innovation is that it represents a significant improvement for the business. To foster innovation, businesses must explore all areas of scientific, financial, technological, and commercial development, including research and development (R&D) as a critical tool to drive progress (Oecd, 2005).

When companies engage in innovative activities, it's crucial to consider what kind of impact those innovations will have on the business. Innovation can lead to increased demand, reduced costs and improved overall performance. For forward-thinking companies, developing new products or processes or making significant changes to existing ones can provide a competitive edge in the marketplace. By driving innovation, businesses can gain a cost advantage over their rivals or boost demand for their products and services. This increased flexibility can translate into higher sales and market share, as well as improvements in operational efficiency and quality (Oecd, 2005).

According to Kongar (1995), innovation activities serve three major purposes for businesses.

- Firstly, sustaining the business in a highly competitive market with diverse product or service offerings requires frequent renewal and innovation activities to maintain competitiveness and survival.
- Secondly, businesses that closely monitor market conditions, adequately perceive consumer needs and demands and respond to technological innovations and market demands are likely to become market leaders. The company that emerges as a leader in the market is in a strong position to control and shape the market by leveraging competitive conditions to its advantage.
- Thirdly, innovation activities may initially appear to reduce profitability, but in the long term they contribute significantly to increasing profits. Successful innovation activities can lead to cost savings, reduced production times and improved performance and efficiency. Although innovation is not a quick fix, it is a practice that bears fruit over time and can have substantial direct and indirect benefits for businesses.

1.1.1. Types of Innovation

In today's global landscape, competition knows no bounds, making innovation a crucial pursuit for companies across industries and sectors. This has led to a broadening of the scope of innovation, impacting industries and individuals alike through product and process advancements. By prioritizing innovation, businesses have seen success in intra-industry competition, leading to a breakdown of the main field of innovation into various subsections (Mayda, 2019).

The literature does not provide a definitive agreement on the types of innovation. Most scholars have identified product, process, marketing and organizational innovations as the most common innovation types. However, there are other types such as technological, social, open and radical innovations that were mentioned in the literature (İmamoglu and Özdemir, 2023). Oecd (2005) provides a framework for grouping innovation types into four general categories: product, process, marketing and organizational innovations.

1.1.1.1. Product Innovation

According to Oecd (2005), product innovation involves the introduction of a product in a completely new form or making significant modifications to an existing product/service. These changes may manifest in the form of functional enhancements in technical features, software, user-friendliness, components, and material.

Product innovation can be achieved in different ways, such as by utilizing new information and technologies or by exploring new usage areas for existing ones. To create new products, businesses often make changes to the intended use or features of their previous offerings. This can involve minor modifications, like using existing chemical compositions for different purposes, or significant improvements, such as incorporating new components or materials to enhance performance. Examples include microprocessors and digital cameras developed with new technologies, as well as ABS braking systems in vehicles. Ultimately, product innovation is about creating changes that add value to consumers and drives business growth (Elverdi, 2019).

Companies that prioritize product innovation can gain a significant advantage in both local and global markets. By introducing new and innovative products, businesses can increase profits,

capture new markets and shape the competitive landscape in their favor. For instance, the introduction of color televisions marked a significant innovation in the industry and helped companies remain competitive amidst technological advancements. Therefore, it is crucial for businesses to prioritize innovation activities if they wish to remain active in the global marketplace and stay ahead of the competition (Adiyaman and Hayaloglu, 2020).

In the realm of service product innovation, new services can be introduced by enhancing existing ones with improved speed, efficiency and added features. This can be exemplified by the ease and swiftness of internet banking for customers, compared to the previous options available (Elverdi, 2019). In the realm of product and service innovations, there are two distinct types: radical and incremental. Radical innovations are those in which a previously non-existent product is introduced to the market, causing significant changes in competitive structures and market conditions. On the other hand, incremental innovations involve continuous improvements to existing products (İmamoglu and Özdemir, 2023).

1.1.1.2. Process Innovation

Process innovation pertains to modifications in production or delivery procedures that are either new or significantly altered. It encompasses substantial modifications to equipment, software or techniques. The domain of process innovation encompasses a plethora of activities such as curbing production/delivery costs, generating novel or vastly improved products, and enhancing quality (Oecd, 2005).

The process concept encompasses a series of activities that occur over a specific duration of time, during which inputs are converted into outputs. Business processes are a collection of activities that are intended to generate a particular outcome for both the market and customers. Process innovation involves the application of new technologies to all stages involved in product production. For instance, the development of alternative fuel production techniques to replace petroleum fuel utilized in vehicles can be cited as an example (Tüysüz, 2020).

To comprehensively assess process innovation, it can be viewed as a horizontal process that converts inputs into services or products. The defining characteristic of innovation is the sustained application of the process (Şahin, 2020). Process innovation extends beyond just

production and distribution methods, encompassing novel or substantially improved methods for the creation and delivery of services (Elverdi, 2019).

1.1.1.3. Marketing Innovation

Marketing innovation encompasses advancements in product design, packaging, promotion, presentation, pricing and positioning that have undergone significant improvement. The objectives of marketing innovation can be summarized as follows (Oecd, 2005):

- Boosting business sales
- Achieving greater success and positive outcomes by accurately identifying customer desires
- Exploring new markets and expanding business activity
- Introducing the product to the market in a fresh and unique manner.

Innovation in marketing arises from a well-defined vision and its implementation results in a valuable outcome. The sources of this innovation are diverse and may include insights from employees, customers, universities, research institutions, and the environment, among other stakeholders (Tüfekci and Tüfekci, 2014).

Marketing innovation studies involve three key stages: discovery, development, and distribution. During the discovery process, the target audience is identified and their problems are analyzed. This leads to the generation of suggestions for solving their problems. The development process involves creating tools that can effectively address the identified issues. Finally, the distribution process ensures that the solution is made readily available to the target audience (İmamoglu and Özdemir, 2023).

Innovations in the realm of marketing encompass various functions such as physical distribution, exchange, and functional facilitation. The focus of these innovations is to differentiate communication with potential customers throughout the entire purchasing process. This involves evaluating customer-seller relationships through a fresh and innovative lens and moving away from traditional methods. The goal of these approaches is to gain a competitive edge by increasing sales, rather than simply the production of more products. To achieve this, businesses employ innovative solutions that allow them to establish a sustainable competitive advantage. Online shopping is a prime instance of marketing innovation. Brands offering their

products for sale through digital channels are making a significant contribution to marketing innovation. This approach also helps them gain an edge over their competitors in the market (Turan, 2018).

1.1.1.4. *Organizational Innovation*

This pertains to the implementation of a fresh and distinct approach to organizing a company's operations, external connections, or workplace structures. What sets this innovation apart from other organizational modifications is that it was not previously utilized by the business and was introduced as a strategic decision made by management. In the realm of organizational innovation, there are several potential benefits to consider. These include lowering transaction and administrative costs, gaining access to non-commercial assets like unregulated external information, enhancing workplace satisfaction (and thus boosting labor productivity), reducing material expenses, and ultimately improving overall business performance. Such contributions have been evaluated by reputable sources like the Oecd (2005).

One instance of innovation outlined by Tüysüz (2020) includes the implementation of a monitoring system to unify the operations of production, finance, marketing and strategy teams. Additionally, the establishment of interdepartmental work teams promotes the seamless sharing of information and better access to resources. The introduction of quality control measures for external contractors and suppliers also contribute to improved service and product standards. Finally, organizational changes are another example of innovative practices implemented in businesses.

Organizational innovation refers to a business's ability to leverage its competitive advantage by restructuring work processes and tools. Numerous studies have shown that the impact of organizational innovation on a business's ability to sustain its competitive advantage over the long term is significant. There are two key aspects of organizational innovation that help businesses gain an edge over their competitors. Firstly, it serves as a catalyst for technical product and process innovation. Without organizational innovation, it's challenging for businesses to implement innovation successfully. Secondly, businesses can utilize organizational innovations to establish a long-term competitive advantage and develop their capabilities (Burmaoğlu and Şeşen, 2011).

1.1.2. Indexes Measuring Innovation Performance

In today's global economy, a nation's innovation performance is a critical factor in its development and competitiveness. The Organization for Economic Cooperation and Development (OECD) is recognized as the leading authority on measuring innovation around the world. Other global organizations, such as the European Union, European Institute of Business Management (INSEAD), the Economist Intelligence Unit and the World Bank are also dedicated to this important work. Typically, a country's innovation performance is measured by five key indices that are widely accepted around the world:

- Global Competitiveness Index (GCI)
- European Innovation Scorecard (EIS)
- World Bank Knowledge Economy Index (KEI)
- Most Innovative Economies Ranking (MIER)
- Global Innovation Index (GII)

Since 1979, the World Economic Forum (WEF) has published the GCI report, which enables countries to gauge their competitiveness. This report plays a crucial role in illuminating the primary drivers of economic growth, thus providing insight into the factors that underlie the success of some countries. It also offers valuable information to political leaders and business executives seeking to develop more advanced and robust economies (Ovali, 2014). According to the WEF, the GCI measures a country's economic productivity and efficiency by evaluating key institutions, policies, and factors. The index is divided into three categories, including basic needs, productivity enhancers, and innovation and diversity factors, with each category further subdivided into 12 distinct subgroups. WEF releases its annual analysis of the GCI to gauge a country's level of development and welfare (Dutta, Lanvin, and Wunsch-Vincent, 2020).

EIS was crafted by the European Commission in 2001 to compare innovation-related efforts among EU member countries. The basic indicators used in the assessment are derived from raw data obtained from Eurostat, and countries such as Turkey, the USA, Japan, and BRIC (Brazil, Russia, India and China) are also included in the analysis. A total of 25 sub-indicators are utilized, which are categorized under three main headings: Facilitators, companies' activities, and outputs (Karaata, 2012).

The Knowledge Economy Index, developed by the World Bank Institute in 1995, assesses the suitability of information for economic growth and development. It provides a collective index of a country's progress towards a knowledge economy, based on four main indicators: economic incentives and institutional structure, education and human resources, innovation process, and information and communication technology. The index is generated by taking the simple average of performance ratings across these indicators (Bashir, 2013).

The "Most Innovative Economies Ranking" is an annual publication by Bloomberg, a leading source of information in the world of business and economics. Since 2014, it has become increasingly popular due to its straightforward and simple indicators, which sets it apart from other indexes. This index is highly regarded for providing valuable information and opinions to foreign investors about various countries (Pekdemir, Erdoğan, Yasemin, and Selçuk, 2015). It examines over 200 countries and comprises of 7 indicators, including R&D intensity, productivity, high technology intensity, researcher intensity, value-added production, higher education activity, and patent activity.

The GII, which is the main subject of this study, is explained in detail in the following Section 1.2.

1.1.3. Information Systems and Analytics

An Information System (IS) serves the purpose of gathering, organizing, retaining, examining and disseminating information with a specific goal in mind. Its primary aim is to deliver valuable information to the intended recipients, in the appropriate amount, format and at the ideal moment. To ensure that ISs deliver useful insights, it is imperative to distinguish between two interconnected concepts- data and information (Rainer and Prince, 2021).

ISs play a crucial role in providing organizations with the necessary information to make informed decisions, identify issues and challenges, manage operations effectively, and innovate new products and services. ISs accomplish these tasks through three fundamental processes, namely input, processing, and output. These processes are critical in ensuring that the information generated is accurate, reliable, and relevant to organizational needs. As such,

understanding the input, processing, and output tasks of ISs is essential for practitioners and scholars working in the field of ISs (Zachman, 1987).

The initial stage of data management involves the collection of raw data, known as input. Following this stage, the processing phase transforms the unprocessed input into useful information. Finally, the processed data is presented as information to those who require it, or to the tasks that require it, known as output. This foundational process of input, processing, and output is critical for the successful handling of data in numerous academic and professional settings (Laudon and Laudon, 2004).

The combination of ISs and analytics have significant potential in aiding decision-making processes. Analytics involves the collection, organization, and analysis of data to draw conclusions and make informed decisions. It requires the use of statistical and analytical techniques to examine data, identify patterns, and reveal trends. In the context of ISs, analytics can be applied to analyze stored data and gain insights that can lead to better decision-making. For instance, a retailer can employ analytics to go over sales data stored in their IS, thus identifying trends and predicting future sales. This approach has significant potential in enhancing decision-making processes and contributing to the success of an organization (Sharda, Delen, and Turban, 2014).

There are three primary types of analytics: descriptive, predictive, and prescriptive analytics (Sharda et al., 2014).

Descriptive analytics refer to the process of understanding what is happening within an organization and identifying the underlying trends and causes of such events. The first step in this process is to consolidate the data sources and ensure that all relevant data is available in a suitable format that allows for appropriate reporting and analysis. Often, the development of this data infrastructure is part of data warehouses. From this data infrastructure, appropriate reports, queries, alerts and trends can be developed using various reporting tools and techniques.

Predictive analytics is an analytical approach that is centered on determining what may occur in the future. It entails the utilization of statistical and ML methods to scrutinize data and make forecasts regarding upcoming results. Predictive analytics enables the identification of trends

and patterns that may persist, thereby facilitating informed decision-making based on such projections.

Prescriptive analytics comes under the scope of operations research or management sciences and aims to optimize system performance by recognizing what's happening, making predictions, and providing recommendations or decisions for the best possible outcomes. It generates various forms of suggestions, including specific yes/no decisions on a problem, particular quantities like airfare prices, or complete sets of production plans. The output can be presented as a report or used directly in an automated decision rule system like airline pricing systems. Because of its ability to provide actionable recommendations, prescriptive analytics is also known as decision analytics.

1.1.4. Decision Support System

DSSs are essential tools for improving the quality of managerial decisions when dealing with semi-structured problems. By combining the intellectual resources of individuals with the capabilities of computers, they provide powerful support for decision-makers (Keen and Scott Morton, 1978).

DSS is an interactive, versatile, and adaptive computer-based IS that helps in decision-making. It is specifically designed to tackle unstructured management problems and uses data, user insights, and provides an easy-to-use interface. A DSS also includes models, which are developed through an interactive and iterative process and can support all stages of decision-making. Moreover, it can include an information component and can be used by a single user or many people in different locations, either online or offline (Sharda et al., 2014).

DSSs are systematically categorized based on their functions and the nature of decisions they facilitate. These categorizations commonly manifest as data-driven DSS, document-driven DSS, communication-driven DSS, model-driven DSS, and information-driven DSS within the existing literature (Ada and Ghaffarzadeh, 2015).

The Data-Driven DSS predominantly focuses on handling data, converting it into information, and delivering this information to decision-makers. Such systems are intricately designed to enable users to retrieve and analyze data, thereby furnishing insights crucial for facilitating

decision-making processes. Document-Driven DSSs find application across diverse user groups and are meticulously tailored for this purpose. These systems primarily aim to navigate web pages and locate pertinent documents through the application of specific keywords or search terms. Communication-Driven DSS allows multiple people to collaborate on a common task (Sharda et al., 2014).

This study systematically investigates DSSs within the framework of two distinct categories, namely knowledge-driven and model-driven DSS, as illustrated in Figure 1.1 (Hızıroğlu, Pişirgen, Özcan, and İlder, 2022).

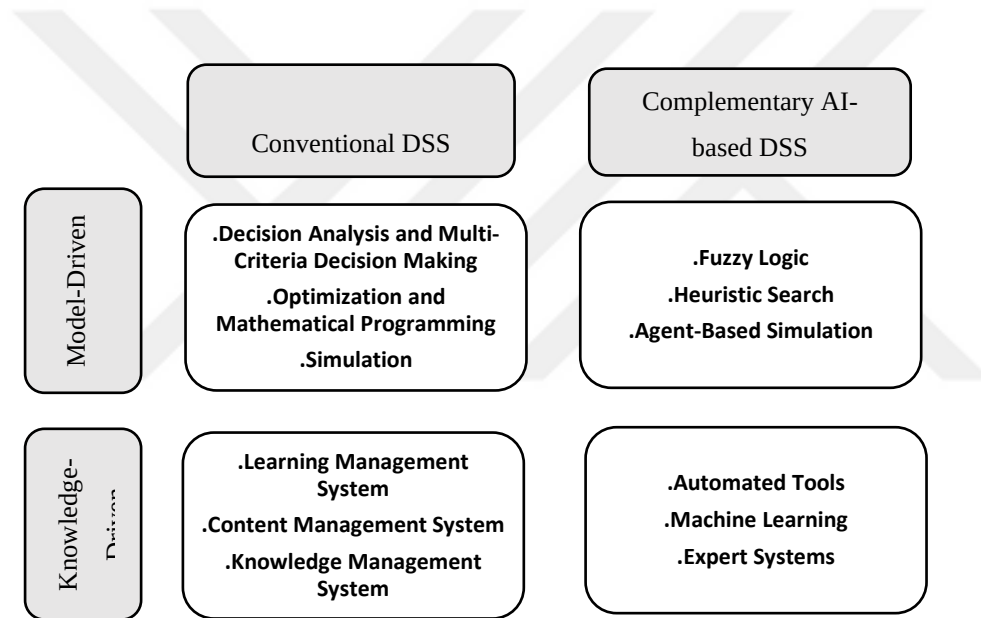


Figure 1.1. A Framework from model-driven to knowledge-driven DSS.

Model-driven DSS leverage input data to enhance decision outcomes through optimization or simulation techniques. The constraint on data entry in these systems underscores the critical role of computational methodologies and the assessment of uncertainty. The establishment of model-driven systems primarily revolves around three key approaches: decision analysis, mathematical programming, and simulation (Power and Sharda, 2007). Furthermore, the integration of artificial intelligence (AI) has propelled the evolution of these techniques, leading to the development of sophisticated methodologies. Advanced fuzzy systems, heuristic search

methods, and agent-based simulations represent notable advancements in the application of AI in this context. These enhancements not only augment the intricacy and efficacy of DSSs but also underscore the academic progress in leveraging cutting-edge technologies to refine and extend traditional modeling approaches (Hızıroğlu et al., 2022).

Knowledge-driven DSS, mirroring intelligent human behavior, are designated as computer-based reasoning systems. Within this paradigm, decision-makers serve as facilitators, engaging with these DSSs to address specialized problem domains. This distinctive interaction establishes them as integral components of what can be termed "person-computer systems." This conceptualization underscores the symbiotic relationship between human expertise and computational intelligence, framing a scholarly perspective on the role of knowledge-driven DSS in problem-solving scenarios (Burstein and Holsapple, 2008). Nevertheless, these systems currently do not involve AI. The integration of AI transforms the landscape, making automated tools, ML, and expert systems (ES) the favored techniques for knowledge-driven DSSs. This shift indicates a departure from traditional methods and emphasizes the preference for advanced computational approaches in supporting decisions within intricate problem domains (Davenport and Kalakota, 2019). This research employs AI- DSS based on complementary knowledge-driven approaches.

1.1.4.1. Artificial Intelligence

AI is a collection of theories and methodologies directed towards the development of intelligent systems. It spans diverse domains and finds applications across various fields. The expansive branches of AI have emerged from a synthesis of disciplines, including computer science, engineering, management science, psychology, philosophy, and linguistics (Reisman, 1996).

AI encompasses a set of principles and notions associated with the creation of intelligent systems, with applications across various domains. To grasp the extent of AI, it is imperative to examine a cluster of domains collectively referred to as the AI family. Figure 1.2 shows the primary branches of applications within the field of AI (Sharda et al., 2014).

Various significant applications of AI are outlined in the literature, encompassing areas such as computer vision, natural language processing, robotics, genetic algorithms, fuzzy logic, and ML.

Computer vision is an interdisciplinary domain that integrates aspects of computer science, engineering, physics, mathematics, and cognitive psychology. It encompasses two primary categories of systems: conventional systems and systems based on deep learning techniques (O'Mahony et al., 2020). Conventional systems rely on manually designed features like edges, corners, and textures for object recognition in images. In contrast, deep learning systems employ neural networks to autonomously acquire features from images. The increasing popularity of deep learning systems in recent years stems from their capacity to excel across a diverse spectrum of computer vision tasks (Greenspan, Van Ginneken, and Summers, 2016).

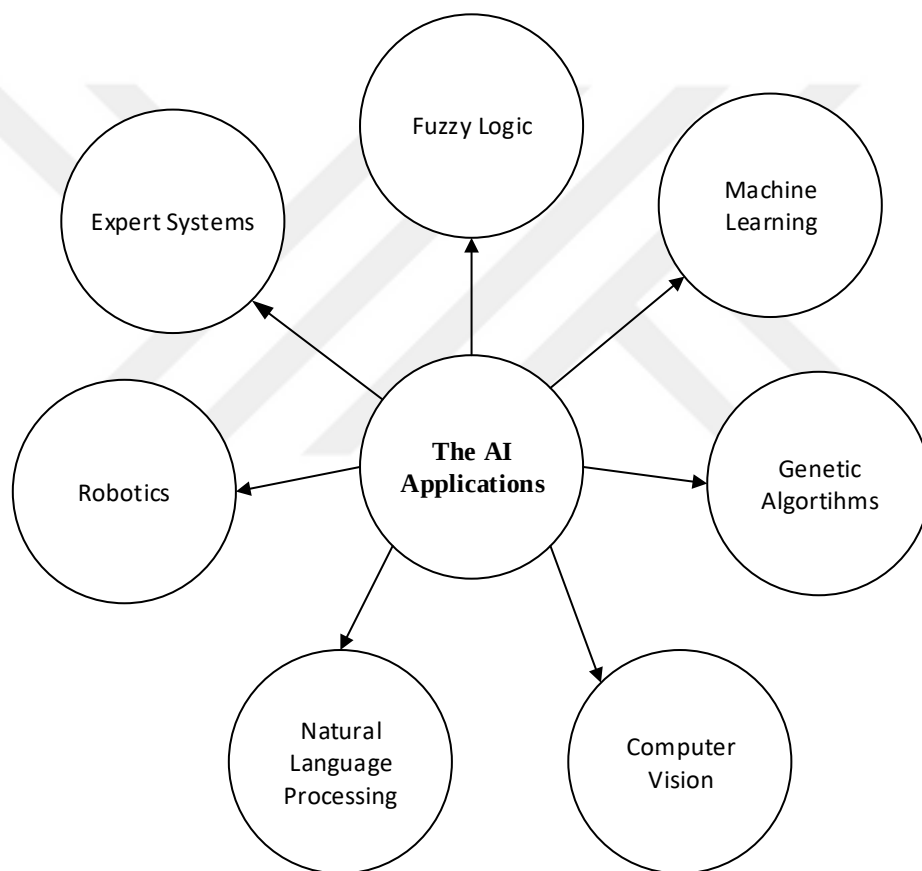


Figure 1.2. The AI Applications

Natural Language Processing plays an important role in text mining and falls within the realms of AI and computational linguistics. It addresses the challenge of comprehending human language, aiming to convert representations of natural language, like textual documents, into more structured forms, typically numerical or symbolic data. This transformation facilitates easier manipulation by computer programs. Natural Language Processing's objective is to

advance beyond mere syntax-based text manipulation, often referred to as "word counting," and achieve a genuine understanding and processing of natural language. This involves considering grammatical and semantic constraints, along with contextual nuances. Natural Language Processing has found success across diverse domains and tasks, utilizing computer programs to automatically handle natural human language processing tasks that were previously exclusive to human capabilities. The most popular of these tasks include: question answering, automatic summarization, natural language generation, natural language understanding, machine translation, foreign language reading, foreign language writing, speech recognition, text-to-speech, text proofing and optical character recognition (Sharda et al., 2014).

Robotics is dedicated to the development, design, and functionality of robots, integrating principles from computer science, engineering, physics, mathematics, and control theory. It encompasses various functions, including perception, reasoning, planning, and control. The field entails the creation of algorithms, models, and systems that empower robots to interact with and navigate the physical world. Specializations within robotics include industrial robots, service robots, mobile robots, aerospace robots, and medical robots (Zhou et al., 2014).

Genetic algorithms form a subset of global search methods employed to identify approximate solutions for optimization problems deemed too complicated for conventional optimization approaches, which assure the identification of the best solution for a given problem. Genetic algorithms have proven effective in addressing a diverse array of complex real-world challenges, such as vehicle routing (Baker and Ayechev, 2003), forecasting bankruptcy (Shin and Lee, 2002), and enhancing Web search capabilities (Nick and Themis, 2001).

Computer-based ISs known as ESs leverage expert knowledge to achieve advanced decision-making capabilities within specific and well-defined problem domains. These systems find application in various fields such as taxation, credit analysis, equipment maintenance, help desk automation, environmental monitoring, and fault diagnosis. ES has gained popularity in large and medium-sized organizations as a sophisticated tool for enhancing productivity and quality. Key concepts within ES involve identifying experts, defining expertise, extracting and transferring knowledge from individuals to computers, and replicating the reasoning processes employed by human experts within the ES (Sharda et al., 2014).

Fuzzy logic is a method of reasoning that maintains logical consistency and is capable of handling uncertain or incomplete information. This approach is reflective of both human thinking processes and the design principles of ES. Fuzzy logic proves valuable for modeling intricate systems and making decisions in the presence of noisy data. Its applications span various fields, including control systems, image processing, natural language processing, and decision-making. Fundamental components of fuzzy systems include the fuzzy set, which denotes the degree of membership of an input to a specific category, and the fuzzy rule, portraying the connection between inputs and outputs. The inference mechanism amalgamates fuzzy sets and rules to facilitate decision-making (Meyers, 2009).

This research focuses on ML within the broader field of AI, with a specific emphasis on two supervised ML techniques:

1.1.4.2. Machine Learning

ML is a branch of AI centered on creating algorithms and statistical models. These tools empower computers to learn, predict, or make decisions without requiring explicit programming for a predefined task. In the field of ML, the term 'learning' pertains to the procedure of instructing a computer an algorithm to identify patterns and formulate predictions using data. This training process commonly entails fine-tuning the algorithm's parameters to minimize the disparity between projected outcomes and accurate results within the training data. This adjustment process is referred to as optimization or model fitting. Following the training phase, the algorithm becomes capable of making predictions on fresh data that shares similarities with the training dataset.

ML algorithms are primarily categorized into four groups: Supervised learning, Unsupervised learning, Semi-supervised learning, and Reinforcement learning (Mohammed, Khan, and Bashier, 2016), as illustrated in Figure 1.3.

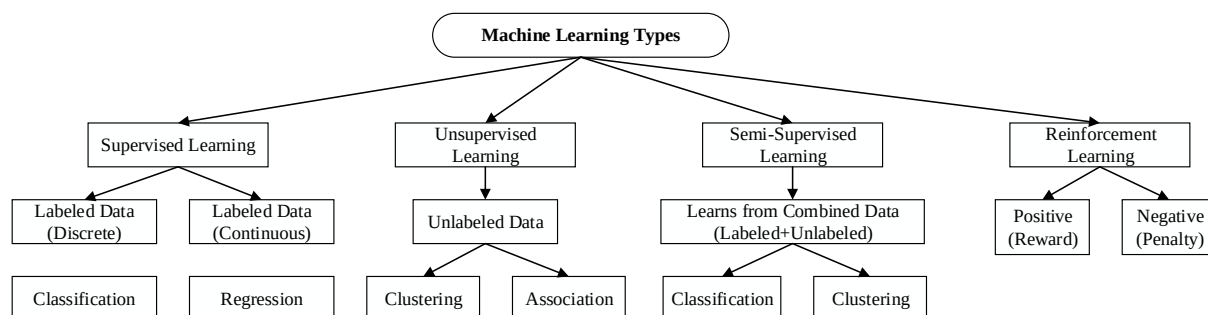


Figure 1.3. Various types of ML techniques

Supervised learning typically involves the ML task of acquiring knowledge about a function that correlates input to output based on provided input-output pairs (Huang and Yu, 2022). It relies on labeled training data and a collection of training samples to deduce a function. In supervised learning, specific objectives are defined for achieving desired outcomes from a given set of inputs (Sarker et al., 2020), indicating a task-oriented approach. The most prevalent tasks encompass 'classification,' which categorizes data, and 'regression,' which models the data. Predicting the class label or emotion of textual information, such as a tweet or product review (text classification), is an example of supervised learning.

Unsupervised learning involves the examination of datasets lacking explicit labels, relying on a data-driven approach without human intervention. It is extensively employed for extracting valuable features, discovering significant trends and structures, grouping outcomes, and exploratory analysis. Common unsupervised learning tasks include clustering, density estimation, feature learning, dimensionality reduction, identifying association rules, anomaly detection, and similar endeavors (Han, Pei, and Tong, 2022).

Semi-supervised learning is a fusion of supervised and unsupervised methods, operating on both labeled and unlabeled data (Han et al., 2022; Sarker et al., 2020). It stands between 'unsupervised' and 'supervised' learning in its approach. In practical scenarios, obtaining labeled data can be challenging in various contexts, while there is often an abundance of unlabeled data suitable for semi-supervised learning. The primary objective of a semi-supervised learning model is to enhance predictive performance compared to using labeled data alone. Applications of semi-supervised learning include machine translation, fraud detection, data labeling, and text classification (Mohammed et al., 2016).

Reinforcement learning stands as a category of ML algorithms that empowers software agents and machines to autonomously assess optimal behavior within a specific context or environment, emphasizing an environment-centric approach. This learning paradigm relies on a system of rewards or penalties, with the ultimate objective of leveraging insights from interactions with the environment to take actions that enhance rewards or minimize risks. Reinforcement learning proves highly effective for training AI models aimed at enhancing automation and optimizing operational efficiency in intricate systems such as robotics, autonomous driving, manufacturing, and supply chain logistics. However, it is not the preferred choice for addressing basic or straightforward problems (Mohammed et al., 2016).

1.2. Global Innovation Index

The GII serves as a valuable tool for countries to assess their innovation capacities and benchmark their competitiveness. Its primary objective is to identify a country's strengths and weaknesses in innovation-related policies and practices. The GII is a comprehensive approach that encompasses measuring innovation performance and analyzing innovation development and progress. Continuous evaluation of the factors affecting innovation is crucial, and the GII provides up-to-date data for this purpose. It enables an annual comparison of countries and offers a detailed analysis of their innovation landscape (Savaşçı, 2017).

The GII index, which promotes innovation activities across various countries and economies, was first introduced by the INSEAD in 2007. Later, the World Intellectual Property Organization (WIPO) and Cornell University also contributed to the development of this index, which comprises around 80 indicators related to innovation. The GII index is regularly updated to provide an instantaneous analysis of these indicators in different countries, and is evaluated by approximately 140 countries each year. The data for this index is sourced from a variety of expert organizations, public and private sector institutions, and surveys conducted by the WEF (Turkmen and Aynaoglu, 2018).

The GII was launched in 2007 by INSEAD in France, led by Professor Soumitra Dutta, and later backed by Cornell University, the WIPO, and numerous knowledge partners. GII provides a thorough analysis of innovation activities across countries and goes beyond traditional indicators such as scientific publications and R&D spending. Its primary objective is to gauge

countries' innovation potential holistically and evaluate their performance accordingly (ELVERDİ and Hayriye, 2020).

The GII strives to go beyond the basic metrics of innovation measurement. To achieve this, new variables must be integrated with a careful balance between variable quality and comprehensive economic coverage. The core focus is on improving innovation measurement in knowledge-intensive services, user and public sector innovation, and policy support for innovative entrepreneurship and venture capital. Additionally, the GII aims to improve innovation linkages, particularly those on an international level, and assess innovation outputs and impacts on a broader scale (Guadagno and Wunsch-Vincent, 2020).

Over the past few years, GII has emerged as a crucial tool for evaluating innovation on a global scale. By taking into account the human factors that drive innovation and prioritizing innovation-based economic growth and development, it has become clear that innovation is a vital driver of prosperity. The GII framework, which is applicable to both developed and developing economies, goes beyond traditional measures of innovation such as (R&D) levels, incorporating a broader range of indicators (Toker and Görener, 2020).

The GII is composed of two sub-indices, namely the Innovation Input Sub-Index and the Innovation Output Sub-Index, both of which are built around specific pillars. The Innovation Input Sub-Index comprises of five pillars that capture elements of the national economy which facilitate innovative activities, while the Innovation Output Sub-Index focuses on the results of innovative activities within the economy. The Output Sub-Index, despite consisting of only two pillars, holds equal importance in the calculation of the overall GII scores as the Input Sub-Index. The overall GII score is determined by averaging the scores of the Input and Output Sub-Indices. Additionally, each pillar is subdivided into three sub-pillars, each of which comprises individual parameters, with an average score of 80 throughout the year (Soumitra, Lanvin, and Wunsch-Vincent, 2020). Out of the 80 parameters, 57 are hard data variables, while 18 are composite indicators sourced from third-party data providers. Additionally, 4 are survey questions taken from the WEF's Executive Opinion Survey (EOS).

The GII's framework is presented in Figure 1.4 which depicts the sub-indexes that form the GII, and also the elements that form the sub-indexes. Each sub-index is divided into three sub-

indices, each with its own set of parameters. There are parameters that make up each of these 3 sub-indexes. Individual parameters may differ or be missing from year to year depending on the country. Also, the Global Innovation Index 2020 Report provides all indexes, the sub-indexes constituting them, and the parameters they consist of, detailed explanations. Furthermore, the report encompasses the references utilized for parameter calculations (Dutta et al., 2020). The section following this one provides a detailed explanation of the indexes and sub-indexes that constitute the entire framework.

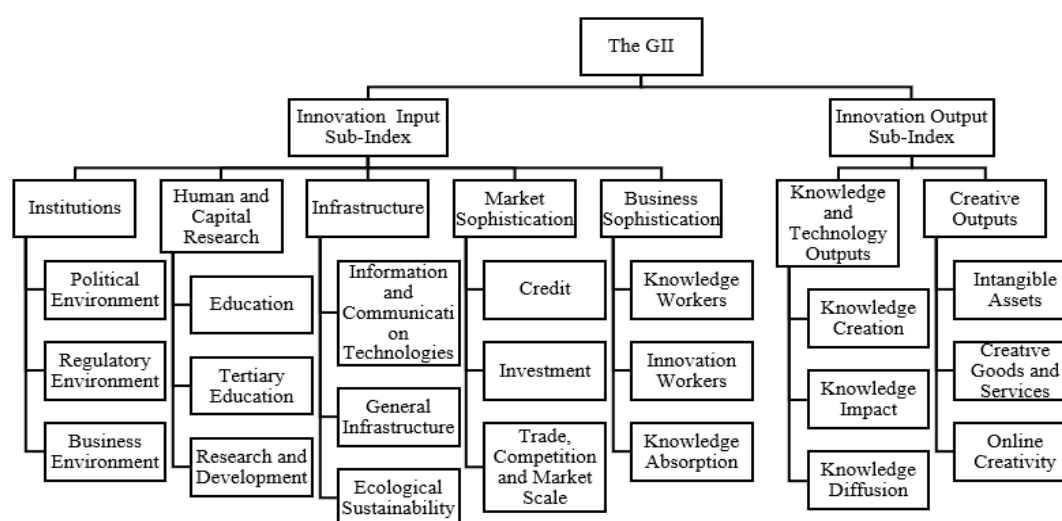


Figure 1.4. The Framework of the GII

1.2.1. The Innovation Input Sub-Index

The GII's first sub-index comprises five enabler pillars that collectively contribute towards innovation within an economy. These pillars are: institutions , human capital and research , infrastructure , market sophistication, and business sophistication. Each enabler pillar defines certain aspects of the environment that foster innovation.

1.2.1.1. Institutions

Fostering economic growth and attracting businesses requires an institutional framework that provides appropriate levels of protection and incentives, while ensuring good governance. The institutions pillar involves analyzing the institutional framework of an economy. The sub-index for institutions comprises three fundamental constituents: the political environment, the regulatory environment, and the business environment.

The sub-pillar of Political environment comprises of two indices which are political and operational stability and government effectiveness. The political and operational stability index is available to assess the probability and severity of political, legal, operational, or security risks that may affect business operations. The scores are standardized and calculated annually. The effectiveness rating of the government is established by evaluating the perceived standard of public services delivered by the government, the caliber of the civil service, and its degree of autonomy from political influences. Additionally, the assessment considers the quality of policy development and execution, along with the government's trustworthiness in upholding these policies. The scores are standardized.

The sub-pillar of regulatory environment comprises of three indices which are regulatory quality, rule of law and cost of redundancy dismissal. The index for regulatory quality is based on how the government is perceived in terms of its ability to create and execute effective policies and regulations that encourage and enable the development of the private sector. The scores received are standardized. The index of the rule of law is based on how much people trust and follow the laws of a society, such as contract enforcement, property rights, police, and courts. It also considers the chances of crime and violence. The scores obtained are standardized. The cost of redundancy dismissal associated with notifying and compensating a redundant worker, in terms of weeks of salary, are referred to as redundancy costs. These costs are calculated based on the number of weeks of salary that are required as advance notice and severance pay, and are averaged for workers with one, five, and ten years of service. A month is equivalent to 4 and 1/3 weeks. If the entire redundancy cost is 8 weeks' salary or less, it is given an 8 score, although the actual number of weeks is still stated. If the cost exceeds 8 weeks of salary, the score is equivalent to the number of weeks that are required to compensate the worker.

The sub-pillar of business environment comprises of two indices which are ease of starting a business and ease of resolving insolvency. The ease of starting a business ranking is determined by sorting the ratings of the economies. These scores are the average of the component indicators' scores. The World Bank records all the procedures that an entrepreneur must follow to initiate and operate an industrial or commercial business officially. It also records the time, cost, and the minimum capital requirement to complete these procedures. These stages involve getting all essential licenses and permissions as well as completing any necessary notices,

verifications, or registrations for the company and its employees with the appropriate authorities. The data are collected by limited liability companies located in the largest economic cities. For 11 economies, such as Bangladesh, Brazil, China, India, Indonesia, Japan, Mexico, Nigeria, Pakistan, the Russian Federation, and the United States, the data are also collected for the second-largest business city. The ease of resolving insolvency for economies is determined by ranking them based on their scores, which are the average scores of the recovery rate and the strength of insolvency framework index. The recovery rate is calculated as the amount in cents secured creditors receive through debt enforcement, reorganization, or liquidation proceedings. The final outcome, whether the business survives or the assets are sold off piece by piece, is also considered. The expenses incurred during the proceedings are decreased by 1 cent for each percentage point of the debtor's estate. Furthermore, the decrease in the value of the asset due to the duration it is held in insolvency proceedings is also considered, including the devaluation of hotel furniture. The insolvency framework index's strength is determined by four other indices: the creditor participation index, reorganization proceedings index, commencement of proceedings index, and management of debtor's assets index.

1.2.1.2. *Human capital and research*

Assessing a country's potential for innovation is greatly dependent on the quality and extent of education and research conducted within its economy. This aspect seeks to evaluate the intellectual resources of economies by using a range of parameters (Dutta et al., 2020).

The first sub-pillar education gauges achievements in elementary and secondary education through parameters such as education expenditure, school life expectancy, and government funding per pupil for secondary education. Results from the OECD Programme for International Student Assessment (PISA) and pupil-teacher ratios are used to measure the quality of education. Higher education is critical for economies to advance beyond simple production processes and products. The education sub-pillar consists of five indices: expenditure on education as a percentage of GDP, government funding per pupil at the secondary level, school life expectancy in years, PISA scales in reading, maths, and science, and pupil-teacher ratio at the secondary level. The total expenditure on education, expressed as a percentage of GDP, encompasses current expenses, capital expenses, and transfers, and is covered by local, regional, and central government. The amount of money provided by the government for education per student, which encompasses all levels of government (local,

regional, and central) and all types of funding (current and capital), including scholarship grants to students, but not including international transfers to the government for education, is computed. This is then expressed as a percentage of GDP per capita in US. The term "school life expectancy" refers to the number of years of schooling that a child of a specific age might expect to obtain in the future. This is based on the existing enrolment ratio at that age. The enrollment ratio represents the likelihood of a child being enrolled in school at a specific age. To calculate a child's school life expectancy, age-specific enrollment rates from primary to tertiary education levels are included. Any enrollment that is not age-based is split by the school-age population for primary to tertiary education, where they are currently enrolled, and then multiplied by the duration of that education level. This figure is then added to the total of age-related enrollment rates. A higher number indicates that youngsters will spend longer years in school and do better overall in the educational system. Due to grade repetition, the predicted years spent in school may not match the expected number of grades completed. PISA is an assessment program conducted by the OECD. The performance of 15-year-old students in reading, mathematics, and science is assessed by this program. The results obtained from PISA provide information about the quality and fairness of learning outcomes achieved globally. The 2018 PISA survey marks the seventh triennial assessment conducted so far. The PISA indicator is generated by averaging each country's reading, math, and science scores. PISA scores are calculated by comparing the results achieved by all test takers in a given country. There is no set minimum or maximum score in PISA. The results are converted to nearly normal distributions, with mean scores of around 500 points and standard deviations of around 100 points. The secondary school student-to-teacher ratio (regardless of teaching area) is computed by dividing the number of secondary school students by the number of secondary school teachers. If data is unavailable for some nations, the upper-secondary ratios are published; if there is still a lack of data, the lower-secondary ratios are considered. A high pupil-to-teacher ratio indicates that each instructor is responsible for a larger number of students. Thus, a higher pupil/teacher ratio indicates restricted availability to teachers for students.

The tertiary education sub-pillar focuses on two main areas: enrollment in tertiary education and the sectors related to innovation. This comprises the percentage of tertiary graduates from science, engineering, manufacturing, and construction. It also addresses the inflow and mobility of tertiary students, which facilitates the exchange of ideas and skills required for innovation. The sub-pillar of tertiary education consists of three indices: tertiary enrollment, % gross;

graduates in science and engineering, %; and tertiary inbound mobility, %. The enrolment of tertiary education students, regardless of age, is compared to the population of the age group that corresponds to tertiary education. For admission to tertiary education, which may or may not include advanced research qualifications, the completion of secondary education is typically a minimum requirement. School enrolment ratios may surpass 100% due to grade repetition or the inclusion of students who are either older or younger than the typical age range for their grade level. This can occur because of early or late entry into the education system. The percentage of all graduates at the tertiary level is calculated from students who have specialized in natural sciences, mathematics, statistics, information and technology, manufacturing, engineering, and construction. To determine the tertiary inbound mobility, the number of foreign students studying in a country is calculated as a percentage of the total tertiary-level enrolment.

The final sub-pillar of R&D focuses on assessing R&D activities, levels of researchers (measured in full-time equivalence), gross expenditures, R&D expenditures of leading global R&D investors, and the caliber of scientific and research institutions. The quality of scientific and research institutions is gauged based on the average ranking of the top three universities in the 2019 QS World University Ranking. To evaluate the R&D expenditures of the top three companies in a given economy, the average expenditure of these firms are examined, which are among the top 2,500 R&D spenders globally. The QS university rankings indicator offers the average scores of the top three universities in the economy, which are part of the top 700 universities worldwide. It is important to note that these indicators are not meant to assess the average performance of all institutions within an economy. The sub-pillar of R&D comprises of four indices which are researchers, FTE/mn pop.; gross expenditure on R&D, % GDP; global corporate R&D investors top 3; mn US\$ and QS university ranking, top 3. Professionals involved in creating new knowledge, products, processes, methods, or systems and managing related projects are considered as researchers in R&D. The total number of such researchers per million population is used as a measure for comparison. This measure also includes postgraduate PhD students engaged in R&D. The data is sourced from UNESCO Institute for Statistics, Eurostat, and OECD Main Science and Technology Indicators. The percentage of GDP spent on R&D within a particular unit or sector of the economy during a specific period is known as the intramural R&D expenditure. The source of funding does not matter for this expenditure. This data is collected from multiple sources such as the UNESCO Institute for

Statistics, Eurostat, and OECD Main Science and Technology Indicators. The top three global companies' R&D expenses are averaged to determine the average expenditure. The amount for a country with less than three global corporations listed is either the average of the two listed companies or the total of a single listed company. A score of 0 is given to countries with no listed companies. If the QS ranking of the global top 1000 universities lists fewer than three universities, the total scores of the listed universities are combined and then divided by three to determine the average score of the top three universities per country. This method implies a score of zero for universities that are not listed in the ranking.

1.2.1.3. Infrastructure

The third pillar consists of three sub-pillars: Information and Communication Technologies (ICTs), General Infrastructure, and Ecological Sustainability. The enhancement of environmentally friendly communication, transportation, and energy infrastructures promotes productivity, efficiency, and market access, ultimately contributing to sustainable growth (Dutta et al., 2020).

The ICTs sub-pillar is comprised of four parameters which are ICT access, ICT use, government's online service and E-participation. The ICT access index is a combination of five ICT indicators, each carrying a weight of 20%. These indicators include: (1) fixed telephone subscriptions per 100 inhabitants, (2) mobile cellular telephone subscriptions per 100 inhabitants, (3) international internet bandwidth (bit/s) per internet user, (4) the percentage of households with a computer, and (5) the percentage of households with internet access. The index was previously part of the ITU ICT Development Index. The ICT use index is a combination of three different ICT indicators, each with a weight of 33%. These indicators include the percentage of people who use the Internet, the number of fixed (wired) broadband internet subscriptions per 100 residents, and the number of active mobile broadband subscriptions per 100 inhabitants. The ICT use index was previously a part of the ITU ICT Development Index. The E-Government Development Index includes the Online Services Index component, which is a measure of how governments use ICTs to provide public services at the national level. In 2018, the Online Service Questionnaire (OSQ) consisted of 140 questions. 206 online researchers from 89 countries who are United Nations Volunteers (UNV), covering 66 languages, evaluated each country's national website in the native language to determine the Online Service Index values for 2018. They assessed the national portal, e-

services portal, e-participation portal, and the related websites of the ministries of education, labor, social services, health, finance, and environment. The earned points were normalized to a range of 0 to 1 for each country. The online index value of a specific country is determined by subtracting the lowest total score from the actual total score and dividing by the range of total score values for all countries. The United Nations E-Government Survey has a supplementary index known as the E-Participation Index (EI). This index focuses on the use of online services by the government for e-information sharing, e-consultation, and e-decision-making with its citizens. A country's EI displays the e-participation methods established by its government in contrast to other countries. The measure seeks to provide insights on how different countries employ online resources to encourage engagement between the government and its residents, as well as among individuals, for the benefit of society. The EI is an evaluation metric that considers the availability and relevance of participatory services on government websites. The country rankings are solely for illustrative purposes, demonstrating broad trends in promoting citizen engagement. The EI is not intended to be an absolute measure of e-participation, but rather an attempt to represent countries' e-participation performance relative to each other at a given point in time. The indicator runs from 0 to 1, with 1 indicating increased e-participation. To normalize the EI mathematically, the lowest total score of any nation in the survey is subtracted from the total score value of a certain country and the result is divided by the range of total score values for all countries.

The general infrastructure sub-pillar is determined by the per capita electricity output, efficiency of logistics, and gross capital formation, encompassing investments in fixed assets and the economy's net inventories. This encompasses improvements to land, purchases of machinery and equipment, construction of infrastructure such as roads, railways, schools, hospitals, private residences, and commercial and industrial buildings. The general infrastructure sub-pillar is comprised of three indices which are electricity output, GWh/mn pop., logistics performance and gross capital formation, % GDP. Electricity production is the measurement of all alternator sets in a station. This measurement encompasses hydropower, coal, oil, gas, and nuclear power generation, as well as geothermal, solar, wind, tidal, and wave energy, combustible renewables, and waste. In order to generate electricity, combined heat and power plants and dedicated electric plants must both produce electricity. The electricity output in gigawatt-hours (GWh) is adjusted based on the population. The Logistics Performance Index (LPI) is a comprehensive evaluation of logistics performance that ranks 160 countries based on

six core performance components. These components include customs performance, infrastructure quality, and timeliness of shipments. A survey of logistics professionals regarding their experiences operating abroad provides the data needed for the ranking. The LPI's six components are (1) customs and border management clearance efficiency ("customs"); (2) quality of trade and transport infrastructure ("infrastructure"); (3) ease of arranging competitively priced shipments ("international shipments"); (4) competence and quality of logistics services ("services quality"); (5) ability to track and trace consignments ("tracking and tracing"); and (6) frequency of shipments reaching consignees within scheduled or expected delivery times ("timeliness"). The LPI includes both subjective and objective criteria, enabling the development of country profiles that assess their logistics friendliness. The ratio of total investments made in the current local currency to GDP made in the same currency is known as the gross capital formation. This ratio is based on the System of National Accounts (SNA) of 1993 and is used to measure investment for a unit or sector. It includes gross fixed capital formation, changes in inventories, and acquisitions minus disposals of valuables.

The sub-pillar on ecological sustainability is based on three indices: GDP/unit of energy use, environmental performance, and ISO 14001 environmental certificates/bn PPP\$ GDP. The amount of purchasing power parity gross domestic product (PPP\$ GDP) per kilogram of oil equivalent energy use is being measured. Total primary energy supply (TPES) is comprised of production plus imports minus exports, minus international marine bunkers, minus international aviation bunkers, and the net change in stock. The Environmental Performance Index (EPI) for 2022 is a comprehensive report that utilizes data to present a summary of the current state of sustainability worldwide. The EPI assesses 180 countries' performance based on 40 indicators across 11 categories, including climate change, environmental health, and ecosystem vitality. These indicators help determine how close a country is to achieving their established environmental policy goals. The EPI provides a report card that identifies both high and low-performing countries in terms of environmental sustainability and offers practical advice to countries looking to achieve a more sustainable future. On a scale of 0 to 100, with 100 being the best performance, the index measures each country's environmental performance. ISO 14001:2015 outlines the prerequisites for an environmental management system that organizations can adopt to improve their environmental performance. Any organization, regardless of its size, type, or nature, can employ ISO 14001 to manage its environmental obligations in a systematic approach that supports sustainability's environmental dimension.

The standard enables organizations to achieve their environmental management system's planned outcomes, which create value for the environment, the organization, and other interested parties. The anticipated consequences of an environmental management system are consistent with the organization's environmental policy and include increased environmental performance, compliance with legal and other requirements, and the attainment of environmental targets. ISO 14001 pertains to an organization's activities, products, and services that it has control or influence over from a life cycle perspective. It does not define specific environmental performance criteria. ISO 14001 can be utilized to systematically improve environmental management wholly or partially. However, claims of ISO 14001 compliance are not accepted until all of its requirements are integrated into an organization's environmental management system and met without exception. The figures are presented per billion PPP dollar GDP.

1.2.1.4. *Market sophistication*

Businesses require credit availability, investment support, international market access, fair competition, and market scale to thrive and foster innovation. The market sophistication pillar is composed of three sub-pillars, which are credit, investment and trade, diversification, and market scale, that focus on market conditions and the overall transaction volume (Dutta et al., 2020).

The credit sub-pillar evaluates the accessibility of credit, including the protection of borrowers' and lenders' rights through collateral and bankruptcy laws, as well as the regulations and practices related to credit information coverage and accessibility. It assesses transactions based on the total value of domestic credit and the gross loan portfolio of microfinance institutions, making the model more appropriate for emerging markets. The credit sub-pillar consists of three indices: ease of getting credit, domestic credit to private sector, and microfinance gross loans as a percentage of GDP. The ease of obtaining credit is determined by ranking economies based on scores calculated from the legal rights index (ranging from 0 to 12) and the depth of credit information index (ranging from 0 to 8). One set of indicators is used to analyze borrowers' and lenders' legal rights, while another set is used to assess credit information reporting. The first set of indicators evaluates the lending process by analyzing specific collateral features and bankruptcy regulations. The second set of indicators assesses the comprehensiveness, breadth, and accessibility of credit data offered by credit reporting service

providers such as credit bureaus or credit registries. Although Doing Business collects data on the percentage of persons covered by public registers and private bureaus for credit purposes, these measures are not used to rate businesses. "Domestic credit to the private sector" refers to the financial resources provided to private sector entities by financial institutions, which may include loans, purchases of non-equity securities, or trade credits and other accounts receivable that establish a repayment obligation. In certain countries, these obligations may also include credit extended to public enterprises. The financial institutions include central banks and commercial banks, as well as other financial entities (including ones that do not accept transferable deposits but do have obligations such as time and savings accounts) where data is available. Other financial businesses include finance and leasing companies, money lenders, insurance companies, pension funds, and foreign exchange companies. The percentage of the current gross loan balances of microfinance institutions in a particular country combined with its current GDP is known as Microfinance gross loans, % GDP.

The Investment sub-pillar comprises of three indices that evaluate the ease of protecting minority investors, market size and dynamism, and venture capital deals. These indices provide hard data metrics on the transaction volume and evaluate whether the market size is matched by market dynamism. The ease of protecting minority investor's index is a combination of two indexes that assess the extent of regulation of conflicts of interest and shareholder governance. The regulation of conflicts of interest index evaluates the measures taken to safeguard shareholders from directors who use company assets for personal gain. It evaluates three dimensions of regulation, including transparency of transactions involving related parties, the ability of shareholders to hold directors liable for self-dealing and the ease of shareholder suits. The shareholder governance index measures the extent of shareholder rights in corporate governance using three dimensions of good governance. It evaluates the involvement of shareholders in significant corporate decisions, protections against excessive board control and entrenchment, and transparency in areas such as ownership stakes, compensation, audits, and financial prospects. The index also determines whether a specific set of rights and safeguards are available in private enterprises. The data for this index is obtained by a survey given to corporate and securities lawyers. The responses are based on securities legislation, company laws, civil procedure codes, and court rules of evidence. The determination of market capitalization, also known as market value, involves the multiplication of the share price by the total number of outstanding shares (inclusive of different classes) of domestically listed

companies. This calculation excludes investment funds, unit trusts, and companies that are primarily focused on holding shares of other listed companies. The data provided represents the average of end-of-year values over the past three years, with the exception of Bulgaria, Ghana, Jamaica, Kenya, Romania, and Serbia, for which averages for the years 2010 and 2011 were taken into account, and Zambia, for which only the 2011 data is available. Thomson Reuters Eikon data contains information about private equity transactions, such as the investment location, Investment Company, investor firms, funds, and crowdfunding, among other specifics. The data pertains to venture capital transactions occurring between January 1, 2019, and December 31, 2019. In 2019, a total of 17,960 transactions were conducted across 81 countries. The data also provides the number of transactions per billion PPP\$ GDP.

The final sub-pillar deals with trade, diversification, and market scale. This category includes three parameters: applied tariff rate, weighted avg., %, Intensity of local competition, and domestic market scale, bn PPP\$. The first parameters gauge the average tariff rate weighted by import shares to assess market conditions for trade. The second parameters consist of a survey question that reflects the level of competition in local markets, although efforts to gather solid data on competition have been unsuccessful so far. The last parameters encompass domestic market scale, measured by an economy's GDP in 2016, and consider the impact of an economy's size on its ability to introduce and test innovations in the marketplace. The "weighted mean applied tariff" represents the average of rates that have been effectively applied, weighted by the import shares of each partner country's products. The data is structured according to the Harmonized System of trade, either at the six- or eight-digit level. The information on tariff lines has been compared with Standard International Trade Classification (SITC) revision 3 codes to categorize goods and assess import volumes. Whenever feasible, specific duties have been converted into ad valorem equivalent rates and included in the calculation of average tariffs based on weight. The weighted average tariff rates at the six- and eight-digit product level are computed for each group of goods. If the applied rate is unknown, the most favored nation rate is used in its place. The index that measures the intensity of competition in local markets is determined based on the average response to the following survey question: "In your country, how intense is competition in the local markets?" The response scale ranges from 1, indicating no intensity at all, to 7, indicating extreme intensity. The gross domestic product, which is calculated using the GDP of the country's PPP and expressed in current international dollars (billions), determines the size of the domestic market.

1.2.1.5. Business Sophistication

The final enabler of the Business Sophistication pillar intends to assess how advanced businesses are and how they contribute to innovation. The human capital and research pillar argues that education, especially higher education, and the prioritization of R&D activities, are crucial for innovation to happen. According to this logic, highly skilled professionals and technicians can help businesses enhance their productivity, competitiveness, and innovation potential. The business sophistication pillar is composed of three sub-pillars, which are knowledge workers, innovation linkages and knowledge absorption (Dutta et al., 2020).

The first sub-pillar knowledge workers contains four quantitative parameters on knowledge workers, such as knowledge-intensive employment, %; firms offering formal training, %; GERD performed by business, % GDP; GERD financed by business, % and females employed w/advanced degrees, %. The percentage of people employed in categories 1 to 3, which are defined by the International Standard Classification of Occupations (ISCO), is referred to as employment in knowledge-intensive occupations. This percentage is calculated by adding up the number of people in those categories and dividing it by the total number of people employed. The ISCO-08 classification includes three categories: managers, professionals, and technicians and associate professionals. The data provided covers the period from 2009 to 2018. If ISCO-08 data were not available, the ISCO88 data were used instead. The ISCO-88 classification includes the following categories: legislators, senior officials and managers, professionals, and technicians and associate professionals. The World Bank's Enterprise Survey collected data from various companies in different countries to determine the percentage of firms that provide formal training programs to their permanent, full-time employees. The data for Botswana, Chile, Costa Rica, Jamaica, Mexico, Paraguay, and Trinidad and Tobago is specifically from 2009. The percentage of GDP spent on business enterprise R&D is known as gross expenditure. The percentage of total gross expenditure on R&D that is financed by business enterprise. The proportion of women who are employed and have advanced degrees in relation to the total number of employed individuals. The employed individuals encompass all people of the working age who were in one of two categories during a brief specified period: (1) Working for a wage, either on the job or employed but not working, and (2) working for oneself, either operating a business or self-employed but not working. Data is analyzed by level

of education, which indicates the highest level of education attained, and is categorized according to the International Standard Classification of Education.

Innovation is greatly influenced by public/private/academic partnerships and linkages. Industrial or technological clusters and networks have resulted in pockets of wealth in emerging markets, while other parts of the territory may still suffer from poverty. The sub-pillar of innovation linkages depends on both qualitative and quantitative data, including university-industry R&D collaboration, cluster development and depth, GERD financed by abroad, % GDP, joint venture/strategic alliance deals/bn PPP\$ GDP, and patent families/bn PPP\$ GDP. Collaboration between universities and industries for research. The survey question's average response: To what extent do businesses and universities collaborate on R&D in your country? The response scale ranges from 1, which indicates no collaboration, to 7, which indicates extensive collaboration. Regarding the development of clusters in the state, the survey question on the role of clusters in the economy yields an average response. The question asks about the prevalence of well-developed and deep clusters in a particular field, which are geographic concentrations of firms, suppliers, producers of related products and services, and specialized institutions. Respondents are asked to rate the prevalence on a scale of 1 to 7, where 1 signifies non-existence and 7 denotes prevalence in many fields. With respect to GERD, the percentage of gross expenditure on R&D that is financed by foreign sources, as a proportion of the country's GDP, and the total amount of national currency spent on R&D, with foreign financing as a percentage of GDP. The following information pertains to joint ventures and strategic alliances. Thomson Reuters data on these types of deals shows the number of deals, fractional counting per billion PPP\$ GDP, and details on the country of origin of partner firms. The data pertains to transactions completed between January 1, 2019 and December 31, 2019. In total, 10,535 deals were announced in 2019 with enterprises based in 122 GII participant economies. The score for each country is equivalent to $1/n$, with n being the number of participating nations per company in a deal. This scoring system ensures that all country scores add up to 10,535. The data is reported per billion PPP\$ GDP. A "patent family" is a collection of patent applications filed in one or more jurisdictions to protect a single invention. Patent families that include applications submitted in at least two different offices and seek protection for the same invention in multiple countries are classified as a subset. When "patent families data" was discussed, it referred to patent families that contained applications filed in at least two IP offices, scaled by PPP\$ GDP (billions). A "patent" is a collection of legal rights granted only

to applicants for innovations that are novel, non-obvious, and industrially useful. Patents are only valid for a short time (often 20 years) and within a specific territory. The patent system tries to stimulate innovation by offering inventors exclusive legal rights for a short time, allowing them to gain economically from their unique work.

The sub-pillar knowledge absorption includes five parameters related with high-tech industries or those critical to innovation. These metrics include intellectual property payments as a percentage of total trade (averaged over three years), high-tech imports as a percentage of total imports, imports of communication, computer, and information services as a percentage of total trade, and net inflows of foreign direct investment (FDI) as a percentage of GDP. In 2016, the percentage of research talent in business was incorporated to enhance the sub-pillar. This addition aimed to measure the presence of professionals involved in conceiving or creating new knowledge, products, processes, methods, and systems, including business management. The absorption of knowledge sub-pillar consists of five parameters: payments for intellectual property, % total trade, High-tech imports, % total trade, ICT services imports, % total trade, FDI net inflows, % GDP and Research talent, % in businesses. The fees for the use of intellectual property that are not included in other payments are determined by using the average of the three most recent years or accessible data. The value is determined using the 2010 Extended Balance of Payments Services (EBOPS) Classification. It refers to the charges for the use of intellectual property that is not otherwise classified, represented as a proportion of overall trade. Total trade equals the sum of total imports of goods under code G and commercial services under code SOX, excluding government goods and services not classified elsewhere, and total exports of goods under code G and commercial services under code SOX, excluding government goods and services not classified elsewhere, divided by 2. The phrase "goods" in the sixth edition of the International Monetary Fund's Balance of Payments Manual refers to general merchandise, net exports of goods under merchanting, and non-monetary gold. "Government goods and services not included elsewhere" is subtracted from "services" to arrive at the category "commercial services". This category includes receipts for the use of proprietary rights, including trade secrets, industrial processes and designs, patents, trademarks, copyrights, and franchises, between residents and non-residents. Along with copyrights on books and manuscripts, computer software, cinematographic works, and sound recordings, these receipts also cover licenses to reproduce or distribute intellectual property found in produced originals or prototypes. They also cover related rights for live performances and television, cable, or

satellite broadcasts. Imports of high-technology products as a proportion of total trade. According to the Eurostat classification, high-technology exports and imports consist of technical goods with a high R&D intensity. This classification is based on the SITC Revision 4 and the definition of the OECD. The commodities are classified into the following sectors: aerospace; computers & office equipment; electronics; telecommunications; pharmaceuticals; scientific instruments; electrical machinery; chemistry; non-electrical machinery; and armaments. The percentage of total trade that comprises imports of Telecommunications, computers, and information services is being referred to. This pertains to the classification of Telecommunications, computer and information services under the OECD's EBOPS Classification 2010, which is coded as SI. Foreign direct investment refers to the average of the net inflows of investment in the past three years. This investment is made to acquire a lasting management interest in an enterprise that operates in an economy other than that of the investor. The investment is calculated by adding equity capital, reinvestment of earnings, other long-term capital, and short-term capital, as shown in the balance of payments. The series shows the net inflows in the reporting economy from foreign investors, and this is then divided by GDP. The resulting figure represents the amount of investment that is received from foreign investors after taking into account any disinvestment made. "Full-time equivalent (FTE) researchers in the business enterprise sector" are professionals who are involved in creating or conceiving new knowledge, products, processes, methods, and systems, as well as managing these projects. The researchers are categorized based on the fields they are involved in, such as business enterprise, government, higher education, and private non-profit organizations. Business enterprise sector in R&D statistics encompasses all organizations, institutions, and firms that are involved in the market production of goods or services (excluding higher education) that they sell to the general public at a significant price, as well as the private non-profit institutions that primarily cater to them. Private enterprises are the main component of this sector, which also includes public enterprises.

1.2.2. The Innovation Output Sub-Index

The outcomes of innovative activities in an economy are referred to as innovation outputs. The Output Sub-Index constitutes the same weight as the Input Sub-Index in calculating the overall GII scores, despite its inclusion of only two pillars. Knowledge and technology outputs and creative outputs are the two pillars that comprise the Output Sub-Index.

1.2.2.1. *Knowledge and technology outputs*

This pillar covers variables that are commonly believed to be the outcomes of innovations or inventions. The Knowledge and technology outputs pillar is composed of three sub-pillars, which are Knowledge creation, Knowledge impact and Knowledge diffusion (Dutta et al., 2020).

The first sub-pillar is about knowledge creation and comprises five indices that result from innovative and inventive activities: Patents by origin/bn PPP\$ GDP, Patent Cooperation Treaty (PCT) patents by origin/bn PPP\$ GDP, Utility models by origin/bn PPP\$ GDP, Scientific and technical articles/bn PPP\$ GDP and Citable documents H index. The first-named applicant's country of residence filing an application with an Intellectual Property office is known as a resident patent application. For example, if a resident of Japan files an application with the Japan Patent Office, it is considered a resident application for Japan. Similarly, when an applicant residing in any of the European Patent Office (EPO) member states submits an application to the EPO, such as Germany, it is deemed a resident application for that member state (Germany). PPP\$ GDP (billions) is used to adjust the data. A PCT application is an international patent application that is filed through the PCT. This system allows inventors to seek patent protection for their invention simultaneously in multiple countries by filing a single application. The first-named applicant's residence determines the origin of the PCT application. The information is only accessible for economies that are PCT Contracting States. The data is adjusted based on PPP\$ GDP (in billions). A distinct form of patent right is referred to as a "utility model". The requirements for obtaining a utility model differ slightly from those for regular patents, with shorter protection terms and less stringent patentability criteria. In some countries, a utility model may be referred to as "petty patents", "short-term patents", or "innovation patents". A submission to an IP office on behalf of the first-named applicant's country of residence is referred to as a resident UM application. For example, if a resident of Germany submits an application to the German IP office, the application is considered a resident application for Germany. The data is measured using PPP\$ GDP (in billions). The information below pertains to the quantity of scientific and engineering papers released across a range of disciplines, including but not limited to agriculture, astronomy, astrophysics, automation control systems, biochemistry molecular biology, biodiversity conservation, biotechnology applied microbiology, cell biology, chemistry, computer science, construction

building technology, dentistry oral surgery medicine, engineering, environmental sciences, ecology, evolutionary biology, food science technology, general internal medicine, life sciences biomedicine and other topics, marine freshwater biology, materials science, mathematical computational biology, mathematics, metallurgy and metallurgical engineering, meteorology atmospheric science, microbiology, nuclear science and technology, physics, plant sciences, radiology nuclear medicine medical imaging, reproductive biology, research experimental medicine, science technology and other topics, telecommunications, transportation, and veterinary sciences. The information is sourced from a collection of journals included in the Science Citation Index and the Social Sciences Citation Index (SSCI). The publications are sorted by publication year and attributed to individual countries or economies based on the institutional address (es) listed in the publication. The count of publications is based on a counting method, meaning that for publications with collaborative institutions from multiple countries or economies, each country or economy receives recognition based on its participating institutions. The data is presented per billion PPP\$ GDP. The number of articles (H) that have received at least H citations is expressed by the H-index. It measures both the scientific productivity and scientific impact of a journal. The H-index is calculated by dividing the number of citations received in subsequent years by the number of articles published in a given year.

The second sub-pillar, on the impact of knowledge, includes Labor productivity growth, %, New businesses/th pop., Software spending, % GDP, ISO 9001 quality certificates/bn PPP\$ GDP and High-tech manufacturing, %. The average growth rate of real GDP per person employed, calculated using constant 1990 PPP\$, is a measure of labor productivity. This metric measures output per unit of labor input, and is derived by dividing GDP by the total employment in the economy. The GDP per person employed is expressed in millions of U.S. dollars that are constant in 1990 and GK PPP. However, it does not account for part-time work because it only counts employed people. As a result, countries with a higher percentage of part-time workers, which are mostly OECD countries, may underestimate GDP per person employed to some extent. The measurement of new business density is the number of newly registered corporations per 1,000 working-age (15–64 years old) population in the private, formal sector companies with limited liability. In 2018, Brazil expanded the scope of data, while data corrections were made for Panama relative to the 2016 survey. Malawi uses data for 2009. Spending on computer software involves the entire worth of acquired or rented pre-packaged

software, which may include operating systems, database systems, programming tools, utilities, and applications. It excludes expenses for in-house software development and outsourced custom software development. The figures are a blend of factual numbers and approximations. The information is presented as a proportion of the GDP. The specifications for a quality management system were established in 2015 for organizations to demonstrate their consistent ability to deliver products and services that meet customer and relevant legal requirements. The goal is to improve customer satisfaction by effectively implementing the system, including processes for enhancing the system and ensuring compliance with customer and relevant legal requirements. All requirements outlined in 2015 are general and applicable to any organization, regardless of its size, type, or the nature of its products and services. The data reported is based on per billion PPP\$ GDP. The information provided indicates the proportion of high-tech and medium-high-tech output, in relation to total manufacturing output. The classification of Technology Intensity Definition, as defined by the Organization for OECD, is used as the basis for this calculation. The International Standard Industrial Classification (ISIC) Revisions 4 and 3 serve as the foundation for the classification scheme. When available, ISIC Revision 4 data was preferred, but when unavailable, ISIC Revision 3 data was used. If a particular country had any ISIC three-digit classification codes missing, but had data available at the four-digit level, the three-digit values were computed by adding up all the available four-digit codes.

Four statistics that are essential to innovation or related to high-tech industries are included in the third sub-pillar: Intellectual property receipts, % total trade, High-tech exports, ICT services exports and foreign direct investment net outflows. The percentage of total trade that pertains to charges for the use of intellectual property not included elsewhere receipts is being measured. This is an average of the three most recent years or available data. The measurement is conducted using the EBOPS Classification 2010—that is, code SH charges for the use of intellectual property not included elsewhere. Payments for the use of proprietary rights, such as patents, trademarks, copyrights, industrial processes, and designs, are known as receipts. In addition to covering licenses to reproduce or distribute intellectual property included in produced originals or prototypes, such as copyrights on books and manuscripts, computer software, cinematographic works, and sound recordings, these payments take place between residents and non-residents. They also cover allied rights, like those for broadcast on television, cable, or satellite and live performances. The percentage of total trade represented by high-technology exports minus re-exports remains a crucial factor to consider. The proportion of

trade represented by telecommunications, computer, and information services is indicated by the EBOPS Classification 2010, coded SI: Telecommunications, computer, and information services. The term "foreign direct investment" represents the average of direct investment equity flows within an economy for the past three years. This includes the total amount of equity capital, earnings reinvestment, and other types of capital. Direct investment is a sort of cross-border investment in which a resident of one economy has control or major influence over the administration of a business in another economy. To determine the presence of a direct investment link, the criterion of owning 10% or more of the ordinary shares of voting stock is employed. The series depicts the amount of investment that leaves the reporting economy and flows to the rest of the world. The amount is then split by GDP.

1.2.2.2. *Creative outputs*

The importance of innovation in creativity is frequently disregarded in innovation measurements and policy talks. From the beginning, the GII has consistently emphasized creativity as a component of its Innovation Output Sub-Index. The ultimate pillar, which relates to creative results, includes three sub-pillars: Intangible assets, Creative products and services, and Online creativity (Dutta et al., 2020).

The first sub-pillar is divided into three parameters: trademarks by origin/bn PPP\$ GDP, industrial designs by origin/bn PPP\$ GDP, and ICTs and organizational model creation. The Intangible assets sub-pillar offers information on trademark applications submitted by residents to the national office. This year, it also includes an indicator to determine which economies have the most valuable brands. This unique statistic aggregates the values of each economy's top 5,000 most valuable brands and adjusts them for GDP. Within this pillar, industrial designs used in regional or national office applications replace one survey question on organizational models, a new area associated with process innovations in the literature. A "trademark" is a symbol that the owner of a specific product or service uses to distinguish it from others. Trademarks can include words, slogans, logos, images, characters, numbers, sounds, and moving images. National and regional IP Offices supervise trademark registration procedures. The rights to a trademark are restricted to the jurisdiction of the IP office that registered it. Trademarks can be registered by submitting an application to a national or regional office, or via the Madrid System. A resident trademark application is one that is submitted with the applicant's home country's intellectual property authority. This statistic is based on the number

of classifications of goods and services listed in trademark applications filed by locals. The data is updated to reflect PPP\$ GDP (in billions). The term industrial design refers to a system of legal rights that safeguard the aesthetic or decorative qualities of items. It is only usable for a short period and within a specific area. A resident industrial design application is one that is filed with the intellectual property office in the applicant's country of residence. For example, an application submitted by a resident of Japan with the Japan Patent Office is considered a resident application for Japan. Similarly, an applicant residing in any of the member states of the Office for Harmonization in the Internal Market, such as Italy, is considered to have a resident application for that member state (Italy) when filing with the Office for Harmonization in the Internal Market. The measure is derived from the design count, which represents the overall number of designs featured in the domestic industrial design applications. The data is adjusted based on PPP\$ GDP (in billions). The inquiry is about the relationship between ICTs and the formation of organizational models. The question asks for an average rating on a scale of 1 to 7, indicating the extent to which ICTs make new organizational models possible (such as virtual teams, remote working, and telecommuting) within businesses in the country.

The second sub-pillar of creative goods and services involves using indicators to assess the creativity and the creative products of an economy. The sub-pillar for Creative goods and services consists of five parameters: Cultural and creative services exports as a percentage of total trade, the number of national feature films per million people, the entertainment and media market per thousand people, printing and other media as a percentage of manufacturing, and creative goods exports as a percentage of total trade. The percentage of creative services exports compared to total exports is calculated using the Extended Balance of Payments Services Classification EBOPS 2010. This covers information services under code SI3, advertising, market research, and public opinion polling services under code SJ22, audiovisual and related services under code SK1, and heritage and recreational services under code SK23. Data pertaining to the United States of America (U.S.) was sourced from the Bureau of Economic Analysis (BEA). The categories considered from the BEA for this purpose encompass audio-visual and related products (such as movies and television programming, books and sound recordings, and broadcasting and recording of live events), information services, advertising, and sports and performing arts. The information concerns the quantity of national feature films created per million individuals aged 15 to 69. A feature film is a movie lasting 60 minutes or more, encompassing fiction, animation, and documentaries, intended for commercial screening

in cinemas, and excluding films made solely for television broadcast, newsreels, and advertising films. The data for Paraguay and Cameroon is based on the year 2009. The Outlook offers a detailed overview containing worldwide analyses and projections for consumer and advertising expenditure in 14 entertainment and media categories across 53 regions for a five-year period. The report covers territories in North America, Western Europe, Central Europe, the Middle East and North Africa, Latin America, and Asia Pacific. The rankings and scores for the Global Media Expenditures in these 53 territories are calculated based on advertising and consumer data, encompassing both digital and non-digital media, measured in US\$ millions using the average 2019 exchange rates. These results are presented in normalized values per thousand population, specifically for individuals aged 15-69. To derive the data for Algeria, Bahrain, Jordan, Kuwait, Lebanon, Morocco, Oman, Qatar, the Islamic Republic of Iran, Malta, Tunisia, and Yemen, the report approximated the total for Middle East and North Africa countries and utilized a breakdown of total GDP (current US\$) to establish reference percentages. The percentage of manufacturing output related to the production and duplication of recorded media was calculated. The sum of the current US dollar value of exports of creative goods, divided by the total value of trade. Creative goods are defined according to the 2009 UNESCO Framework for Cultural Statistics, which is based on the 2007 Harmonized System (HS 2007) for international trade of cultural goods and services.

The third sub-pillar on digital creativity includes four measurements: Generic top-level domains (TLDs)/th pop., Country-code TLDs/th pop., Wikipedia edits/mn pop., and Mobile app creation/bn PPP\$ GDP. The measurement on mobile app creation was refined in 2019 to more accurately track app downloads based on the location of the developer or company headquarters. This enhancement offers deeper insights into the progress of innovation, production, and trade of digital creative products and services within an innovation-driven economy. A TLD is a category of domains maintained by the Internet Assigned Numbers Authority for use on the Internet. The categories of TLDs include generic (TLDs), country-code top-level domain (ccTLDs), and sponsored TLDs. While generic TLDs can be unrestricted, meaning they can be used by anyone, some have eligibility criteria. The five generic domains .com, .info, .org, .net, and .biz are included in the statistic. However, .name, .pro, and sponsored domains (.arpa, .aero, .asia, .cat, .coop, .edu, .gov, .int, .jobs, .mil, .museum, .tel, .travel, and .xxx) are not included. The total number of registered domains as of December 2019 is represented in the statistic, which includes existing domains, new registrations, and expired

domains. The information was gathered from a 4% random sample of domain names for each top-level domain. A domain's location is determined by the domain name owner's registration address, which is discovered through a whois inquiry. Registration information is organized by nation and postal code, and then integrated at several geographic levels, such as county, city, or country/economy. The original raw data was adjusted based on a thousand people aged 15–69. For privacy reasons, only standardized values are disclosed. While the relative positions remain unchanged, the actual figures are not. A ccTLD is a type of TLD managed by Internet Assigned Numbers Authority for use on the internet. These TLDs consist of two letters and are designated for use by a specific country, economy, or autonomous territory. As of June 2020, there are 3916 ccTLDs available in various alphabets and characters. The statistic represents the total number of registered domains, including existing domains, new registrations, and expired domains as of December 2019. The data is collected by the registry responsible for each ccTLD, representing the total number of domain registrations within the ccTLD. Each ccTLD is associated with a specific country and is not based on the registration address of the registrant. ZookNIC reports that between 85-100% of domains for the ccTLDs it covers are registered in the same country, except for ccTLDs that have been licensed for global commercial use. Data is reported per thousand populations between 15-69 years old, and only normalized values are reported for privacy reasons. While relative positions are preserved, magnitudes are not disclosed. The information presented in this document is obtained from the internal data sources of Wikimedia Foundation. The data provided pertains to the economies that have more than 100,000 edit counts in the year 2019. However, the data does not include contributions that can be identified in the data sources. The data is presented per million population aged between 15 to 69 years old. Data from China is not available and is considered as “n/a”. Global mobile app downloads, categorized by the origin of the developer/firm's headquarters, have been scaled by PPP\$ GDP (billions). These global downloads have been compiled using App Annie Intelligence, public data sources, and the company's proprietary forecast model. The data used to create the model was collected from the Google Play Store and iOS App store in each country, between January 1, 2019, and December 31, 2019. Due to the unavailability of data for China on the Google Play Store, and its availability only on the iOS App store, the data from China has been treated as missing and marked as "n/a".

1.3. Aim and Structure of the Study

In the modern economy, growth models are mostly determined by innovation. Countries and businesses that renew their product and service range are at the forefront of global competition in the rapidly globalizing world. To manage innovation, it needs to be measured. Measurable information is obtained through the GII, which identifies inputs and outputs that are indicators of innovation. The GII ranks countries based on their innovative capacity. In the current literature, studies on the GII typically take a strategic perspective, or they use statistical or ML-based approaches to analyze the structure and effects of GII. However, these studies are limited in scope, often focusing on specific periods, countries, or sub-indexes. This lack of comprehensive research has motivated the current study. Most studies in the literature have only estimated GII for a few years, and the selection criteria used in the estimation process have not been chosen by any formal selection method. Calculating an average of 79 different parameters for each country is a time-consuming and expensive process. Hence, in response to this gap in the literature, this study aims to calculate GII using fewer parameters and estimate the scores of countries. By identifying the most important parameters and calculating GII scores based on them, countries can make more focused progress in guiding their innovation policies. The proposed research could offer a significant contribution to the field of innovation studies and help policymakers to better assess their national innovation capabilities.

To conduct a thorough analysis, the first step was to acquire and compile detailed score tables of countries from the GII reports published between 2012-2022. In the next stage, a feature selection process was implemented to identify the most relevant parameters from a pool of 79 common parameters. This was necessary since changes were noticed in the parameters and numbers used for GII score calculation over the years. Moreover, to ensure that the selected parameters were consistent with the literature studies, a comprehensive review of the compatibility of the parameters with the literature was also conducted. In the next stage, estimation was made using ANN and regression methods over the selected parameters. The results were thoroughly scrutinized and analyzed to draw meaningful conclusions. During the last phase of the project, a highly efficient DSS was developed with the primary objective of providing accurate and reliable GII estimation results for different countries across the globe. The system was designed by incorporating advanced analytical techniques and algorithms that enable users to input a wide range of parameter scores associated with various innovation

dimensions. Upon entering the required parameter scores, the system generates the corresponding GII estimation results for the respective country based on a comprehensive analysis of the input parameters. The DSS has been designed to provide a user-friendly interface that is easy to navigate and understand, ensuring that users can obtain the desired results with utmost ease.

The aforementioned study highlights the importance of evaluating the parameters that significantly contribute to the development of the GII. The study suggests that this should be a focus area for researchers and policymakers alike, in order to better understand the factors that drive innovation in nations worldwide. In addition, the research provides valuable insights into the characteristics that should be given greater attention when analyzing a nation's innovation performance. This can be viewed as an important step towards developing more effective strategies and policies for promoting innovation and economic growth. The study also makes a theoretical contribution by highlighting the differences between the findings produced with one-year data and those obtained from a comprehensive data study. This can help researchers and policymakers to better understand the limitations of one-year studies and the benefits of conducting more comprehensive research on innovation and related topics. Overall, the text emphasizes the need for continued research and analysis in the field of innovation, in order to promote sustainable economic growth and development.

The study is structured in a comprehensive manner to provide a clear understanding of the research conducted. The first chapter serves as an introduction, providing an overview of the research topic and relevant studies. In the subsequent section, the material and methodology used in the study are discussed in detail, providing a sound understanding of the techniques utilized. The findings and discussion section is a crucial part of the study where the results obtained from the feature selection method, as well as the prediction methods using the ANN and regression analysis, are presented and analyzed. This section provides insights into the effectiveness of the chosen algorithms and their ability to predict outcomes accurately. Additionally, the designed prototype DSS is also explained in detail, providing an overview of the system architecture and how it can be utilized in real-world applications. Overall, the study aims to provide an in-depth analysis of the topic, making it a valuable resource for researchers, professionals, and anyone interested in the subject matter.

2. LITERATURE REVIEW

In the literature since 2009, there have been several studies conducted on GII, or the Global Innovation Index. These studies have delved into various aspects of innovation across countries, including factors that influence innovation and how countries can improve their innovation performance. At the end of this section, a table is provided that summarizes the key insights from these studies.

Ersöz (2009) has examined innovation indicators for Turkey and Europe by using two different analysis methods. Hierarchical Clustering Analysis was applied to innovation indicators for Turkey, while multivariate analysis was applied to innovation indicators for Europe. As a result, the author has identified the cluster to which Turkey belongs and revealed that the most influential variables in innovation indicators are the number of scientists and engineers, the number of higher education graduates, public R&D expenditures, information and communication technology expenditures, and exports of high technology products.

Ünlükaplan (2009) has studied the relationships between economic development, competitiveness, and innovation in EU member countries, using canonical correlation analysis. The author has concluded that there is a high correlation between economic development and innovation and competitiveness characteristics.

Kijek and Kijek (2010) conducted research to evaluate how input factors impact innovation outputs in the European Union. The study examined the components that determine the GII of 22 EU countries in 2005 and 2006. Through canonical correlation analysis, researchers were able to identify and quantify the associations between input and output factors. The findings indicate that R&D expenditures, ICT, and human capital are crucial innovation inputs that impact innovation and patent propensity, as well as new-to-market sales.

Chen, Hu, and Yang (2011) present a comparison of R&D efficiency among different nations using diverse output-oriented R&D efficiency indices that are developed through the data envelopment analysis approach. The study compares the R&D efficiency of twenty-four countries for the period of 1998 to 2005. Empirical findings reveal that countries have similar R&D efficiency in terms of patents and royalties, while their performance regarding journal

publications varies significantly. The study further examines how the national innovation system, particularly the innovation environment, affects these output-oriented R&D efficiency indices.

Ye Sheng and Wong (2012) studied that the GII conceptual framework was utilized to present a comprehensive collection of factors that influence a nation's innovation and development. The GII, specifically for the year 2011, was employed to evaluate the innovation capabilities of 124 countries. To delve deeper into the connections between the different innovation inputs and outputs, a canonical correlation analysis was conducted at a broader level on the five Innovation Inputs and two Innovation Outputs. The primary aim was to assess the strength of the associations between the innovation input scales (independent variables) and the innovation output scales (dependent variables).

Karagöz (2013) conducted that Turkey's rank in the GII between 2007-2012 was not satisfactory when compared to other countries. The study indicates that Turkey's innovation level is not sufficient, and there hasn't been enough focus on innovation. Moreover, the research suggests that the perception of innovation in Turkey is not fully developed. The study concludes that innovation is crucial for better service delivery in the public sector, highlighting the need for prioritizing service innovation in the public sector.

Mahroum and Al-Saleh (2013) began by stating that innovation indices have been developed by researchers over the years to evaluate the effectiveness of government intervention and compare countries' innovation policies and performance. The study utilized data from the GII to examine the relationship between countries' innovation enablers and innovation performance. The focus was on comparing the innovation performance of developed (or high-income) countries to that of developing (or low-income) countries, with a sample that included 141 countries - 86 classified as high-income and 55 as low-income. The GII was chosen for its extensive coverage of countries, use of input and output measures, and connection of its measures to key elements of National Innovation Systems.

Şimşit et al. (2014) have used principal component analysis to determine five conceptual factors for GII, which was published in 2013. They have further applied data mining methods to cluster 142 countries. This study has provided a new perspective to the countries evaluated within the

framework of a 4-cluster solution, allowing them to better understand their performance, a global position more realistic and to draw their strategies in this direction.

Kononova (2015) conducted a study to analyze the correlation between the e-index and the GII for a total of 96 countries. The e-index is an indicator of the scientific research output of a country, while the GII measures the innovation performance of countries based on various factors such as R&D, infrastructure, and human capital. The study aimed to explore how these two indices are related and whether there is a direct or indirect impact between scientific research output and innovation performance. The study findings could help policymakers and researchers to identify areas for improvement and enhance the innovation ecosystem in countries.

Tezcan (2015) has compared Turkey and the EU in terms of innovation performance using cluster analysis and multidimensional scaling analysis. The study has revealed that Turkey is in the same group as the countries called "Unpretentious Innovators"; however, Sweden, Finland, Denmark, and Germany are separated from other countries and are in the group called Innovation Leaders.

Usman and Liu (2015) have measured innovation capacity and productivity for South Asian Association for Regional Cooperation (SAARC) countries in GII 2014 and argued that these countries should strengthen their education systems for economic development.

Albulescu and Drăghici (2016) conducted a study that utilized panel data analysis method and focused on 34 developed and developing countries within the scope of the GII from 2009 to 2012. The main objective of the research was to investigate how innovation developments can impact a country's economic growth in a positive way. They concluded that opportunity- or innovation-driven productions had a positive effect on both GII and innovation output. Additionally, the study highlighted the significance of innovation-driven initiatives in generating and disseminating knowledge. Furthermore, the quality of institutions, infrastructure, and human capital played an important role for the GII input component.

Crespo and Crespo (2016) have applied a fuzzy-set qualitative comparative analysis on the GII for 2015. They have demonstrated that countries divided into two groups as high-income and

low-income can achieve high innovation performance with a combination of causal conditions such as institutions, human capital and research, infrastructure, market complexity, and business complexity, which are the five inputs providers of GII. The study has shown that low-income countries need to come up with more versatile solutions than high-income countries to achieve better innovation performance. It has also emerged that to improve the level of innovation, the political decision-making mechanisms should be different for both country groups.

Hancıoğlu (2016) employed canonical correlation analysis to analyze the link between innovation input and output sub-index variable clusters in the GII of OECD nations from 2011 to 2015. The study has emphasized the importance of determining the policies that will form the basis for developing countries to increase their innovation performance.

Sohn, Kim, and Jeon (2016) conducted research to examine how the innovation capability and efficiency levels of individual countries can be observed by using input and output factors via the GII. However, the GII does not account for possible structural relationships among the factors that influence a country's innovation performance. Therefore, this study aims to propose a Structural Equation Model (SEM) based on the conceptualized national innovation framework, which includes seven factors that represent inputs (institutional framework, human capital and research, infrastructure, market sophistication, and business sophistication) and outputs (knowledge and technology outputs, and creative outputs). By analyzing GII data from 2013, the study found that business sophistication and infrastructure have the most significant direct and indirect effects on creative output, respectively. Additionally, a new ranking based on the fitted SEM has been generated, which provides valuable insights to enhance innovation capabilities. This scholarly pursuit contributes to the advancement of understanding and evaluating innovation dynamics at the national level, offering actionable feedback to promote innovation ecosystems.

Franco and de Oliveira (2017) have conducted a study that aims to analyze the innovation progress of the Brazil, Russia, India, China, and South Africa countries as a block and its individual members from 2008 to 2013. This is based on the importance of innovation to global competitiveness and the significance of the BRICS countries to the world economy. The data source used for this study is the GII, which is recognized for its importance in measuring

innovation by inputs and outputs. The study includes two analyses. First, the BRICS innovation index was examined as a whole, demonstrating a significant drop in the innovation index during the economic crises of 2008-2009, followed by a slow rebound. Second, each country's inputs and outputs were analyzed using a regression model. The regression analysis by nation yields a low R^2 , indicating that the relationship between inputs and outputs cannot be fully explained. The BRICS group exhibits a significant correlation ($R^2 = 64\%$).

Jankowska, Matysek-Jędrych, and Mroczek-Dąbrowska (2017) studied how national innovation systems translate innovation input into output across countries. The researchers used the 2015 GII to discuss the definition of "innovation" and how it is understood on a national scale. The study was based on the notion that a higher innovation input would result in a higher innovation output in a country. The researchers tested this assumption on 228 nations using cluster analysis. After that, they carried out a more thorough examination of two nations, Bulgaria and Poland, where the research question was not answered. The researchers sought to confirm how and why national innovation systems produced ideas successfully or unsuccessfully using the cross-comparison approach.

Khedhaouria and Thurik (2017) conducted a research study from the perspective of the national innovation system to create an integrating framework that identifies the factors leading to high levels of national innovation capability outcomes. By employing a fuzzy set qualitative comparative analysis approach (fsQCA), the study distinguishes among five conditions - establishing national institutions, enhancing human capital and research systems, upgrading infrastructures, and enabling business and market conditions. The study examined data obtained from the GII database, which included 74 indicators and covered 133 countries from 2012 to 2015. The results show that there are three different combinations of variables that lead to high levels of innovation capability rather than a single pathway. Two configurations highlight the fact that a country can obtain high innovation capability by combining three distinct circumstances, one of which states that institutions and market conditions are "not necessary." The third arrangement highlights that in order for a nation to attain a very high capacity for innovation, all five requirements must be met. Some significant implications of these findings for theory and practice are examined in the paper.

Hamidi and Berrado (2018) have studied the relevance of the determinants suggested by the GII to predict national innovation outputs and develop the best possible model. They have proposed a mixed methodology of unsupervised and supervised data analysis models and segmented innovation determinants by using the 2017 GII dataset. Thus with the help of a reduced set of determinants, a better definition of factors to explain the variation in innovation intensity between countries is provided.

Pop and Pop (2018) have conducted research on the assessment and measurement of innovation at the global and European levels. This approach serves as a means of analyzing the innovative process and developing long-term strategies for sustainable economic growth. The measurement of innovation is conducted by specialized organizations through specialized studies, which use relevant indicators that have been modified over time to ensure their effectiveness. This paper proposes a comparative analysis of a country's innovation performance as assessed by two tools: the GII and the European Innovation Scoreboard. The GII for 2017 analyzed 127 functional economies and considered 81 indicators that characterize global innovation, such as the political environment, education, infrastructure, business sophistication, and more.

Turkmen and Aynaoglu (2018) conducted research on the importance of innovation in providing an advantage to economies in a changing competitive environment. Economies that place a greater emphasis on innovation retain their strength in competition with other economies. The study analyzed the relationship between the GCI and the GII, based on the interactive relationship between competition and innovation. The study aimed to determine how the parameters of the Global Competitiveness Index affect the GII for the top 29 countries in the GII rankings of 2009. The study used data from 2009 to 2017 for both indices. The study found that there is a high positive correlation between GII and GCI parameters such as higher education and training, innovation, and labor market efficiency. The macroeconomic environment parameter was found to be the least effective. The study also found a statistically significant relationship between the innovation and specialization factor group and GII in the model based on GCI factor groups.

Ayçin and Çakın (2019) have conducted research about that the correlation between a country's economic development and its level of R&D, innovation, technology, and science has become

increasingly apparent in recent years. This has resulted in countries continuously supporting R&D and innovation activities by fostering an innovation ecosystem and encouraging the participation of actors within it. To measure innovation performance, this study employed the Entropy and Multi-Attributive Border Approximation area Comparison (MABAC) methods and used the European Innovation Scoreboard 2018 indicators developed by the European Commission to assess the comparative efforts of EU member states in innovation. The criteria were weighted using the entropy method, and variables such as innovators, intellectual assets, finance, and support were deemed the most important. The MABAC method was used to calculate innovation performance values for the countries. The study concluded that the countries with the highest innovation performance were Switzerland, Sweden, and Denmark, while Ukraine, Romania, and Macedonia had the lowest performance.

Pençe, Kalkan, and Çeşmeli (2019) have estimated the ranking scores of 128 countries using ANN. They have used only the 2016 GII report and chosen 27 features in the report as input data. Their findings are significant since they are the first to use the GII 2016 data set to fit curves and estimate scoring processes. Their results have shown that the selected 27 features are sufficient to obtain the income points of the countries. It has also been revealed that increasing the number of features will contribute to the curve fitting process as it will decrease the Root Mean Square Error (RMSE) value. In their study, they have proceeded by averaging the 8-year GII report data and making feature selection, which will give more accurate results, instead of using a single year.

Altıntaş (2020) has conducted research on the GII components of 28 European Union countries for 2019. The research has four purposes, and the mathematical methods used to determine these goals are presented below as separate items. The first goal was to determine the importance of innovation components according to countries by using the Entropy method. The second goal was to determine the innovation effectiveness performance values of the countries. The third goal was to determine the innovation efficiency performance values of the countries using the OCRA method. Finally, the fourth goal was to determine the innovation efficiency performance values of the countries using the EATWIOS method. Based on the evaluation of the innovation effectiveness values of the countries, it was found that the top three countries with the highest innovation effectiveness are Sweden, Netherlands, and England. It can be concluded that countries with high innovation effectiveness perform more efficiently than other

countries by analyzing which innovation input/inputs will provide which innovation output/outputs more optimally.

Cui, Sun, Li, Yang, and Wu (2020) conducted a study aimed at assessing the correlation between various factors and determining changes in innovation capability levels through the use of data-driven approaches such as random forest and ANN methods. The researchers sought to predict GII for each country using key indicators, enabling countries to keep track of their development and improve their national innovative capacity in a swift and effective manner.

Gürtuna and Polat (2020) have conducted research on the impact of rapidly changing technology and the increased competition due to globalization, which highlights the significance of innovation efforts for companies and countries. In order to determine their competitiveness, analyze the current situation, and establish a roadmap for innovation, it is crucial for companies and/or countries to measure innovation. Several studies have been conducted on measuring innovation, as there are numerous factors that affect it. The GII, published annually, measures the innovation of countries using various indicators that affect innovation. To analyze the data of 126 countries, this study considered the three subheadings under each of the main indicators of innovation input and output subscale in the 2018 GII report by using clustering analysis. This study utilizes Ward's Technique and K-Means method.

Layos and Peña (2020) conducted a study on the relationship between innovation systems and their ability to tackle pandemics, using a two-step System Generalized Method of Moments approach. The research was conducted on the ASEAN-5 economies, and their respective innovation levels were measured using the GII for 2019. The study found a significant and positive relationship between a country's level of innovation and its ability to handle a crisis. It also discovered that search interest, which is a measure of market response in an innovation setting, had a substantial negative link with crisis management. The report presents early analysis and insights into these two main findings, as well as policy recommendations for enhancing innovation systems.

Omer, El-Amin, Nasr, and Ahmed (2020) have conducted research on Innovation Performance, including its fundamental concepts, evolution, and quantification methodologies. The study primarily relies on data from the GII, with a specific focus on assessing African Innovation

Performance. The research provides a brief summary of the existing literature on the use of ML in modeling innovation performance. Through simple ML techniques, the analysis aims to examine two key aspects. Firstly, it predicts and evaluates the "Mobile App Creation Indicator" from the GII, drawing insights from the stack-overflow developers survey. Secondly, the research designs and compares models aimed at forecasting the Innovation Output Sub-index, which is also derived from the GII.

Doğruel and Firat (2021)'s research focused on two globally significant indices, the Network Readiness Index (NRI) and the GII, both of which have strong academic infrastructures to determine innovation levels of countries. To compare and analyze the indicators within the dimensions below the subscales of the GII score and NRI index, the study employed supervised ML, specifically regression tree analysis and linear regression analysis. The study drew its data set from NRI indicator values and GII scores of 123 countries in 2016. The primary objective of the regression tree application was to estimate GII based on NRI indicators and determine the best discriminating GII indicators, with the Classification and Regression Tree (CART) algorithm being used for analysis. The analysis results identified NRI indicators that were used in the GII scores and country ranking estimation. The study also conducted linear regression analysis with the same data set and compared the regression tree obtained by the CART algorithm with the linear regression model.

Famalika and Sihombing (2021) carried out research on the subject. The k-means method and k-medians method are techniques used for clustering countries based on GII for 2018. In the k-means method, Cluster 1 comprises of 48 Countries, Cluster 2 comprises of 45 Countries, and Cluster 3 comprises of 33 Countries, where the average value of seven variables is the highest. In the k-medians method, Cluster 1 comprises of 33 Countries, where the average value of seven variables is the highest, Cluster 2 comprises of 53 Countries, and Cluster 3 comprises of 40 Countries. The clustering results obtained using k-means method and k-medians method indicate that k-medians is superior to k-means method because the variance value of k-medians is smaller than k-means.

Huang and Yu (2022) carried out a research on The GII, which is a widely used measure of a country's innovation competitiveness. The purpose of this study was to analyze the complete relationships of the GII. The GII structure comprises of several layers of variables. In a research

structure, as the number of variables and layers increases, the number of relationships can grow exponentially. To screen out the insignificant relationships, this study employed both fuzzy set qualitative comparative analysis and structural qualitative association as the research methods and proposed a systematic approach. The data used in this study was GII 2020, and the findings revealed two connections for Positive GII and two for Negative GII. When all antecedents except Market Sophistication are positive, the GII is often positive. In contrast, when all antecedents save Market Sophistication are negative, the GII is typically negative. This study also provides a new ranking of countries based on empirical findings. Although a country's new rating may differ from its previous one, the top countries in the original ranking remain at the top, while the bottom countries remain at the bottom.

Bate, Wachira, and Danko (2023) conducted a research on innovation and found that despite the lack of research on this subject, it is still necessary to clarify the main factors that determine innovation performance. Moreover, there is still a need for a comparative analysis of the determinants of innovation performance across countries with differing income levels. To address this research gap, this study analyzed the innovation performance of 63 countries. The participating countries were chosen from the GII dataset of 2020. The research included multistage and multimodal analyses such as hierarchical regression, multiple linear regressions, and ANOVA. The goal was to investigate variations in innovation performance and identify significant variables in each category of countries. The findings revealed that human capital, research, infrastructure, and business sophistication were the primary factors influencing countries' innovation success.

Dobrovolska, Sonntag, Kachula, Hubaryk, and Savanchuk (2023) have conducted a research on the causal relationship between financial and investment indicators and the level of innovation development in leading GII countries and Ukraine. The aim of the paper is to analyze the correlation among the following indicators for the years 2011-2020: GII value, foreign direct investment (net inflows), domestic credit to the private sector, ease of getting credit, protecting minority investors, and real interest rate for a sample of 10 leaders in GII-2022 and Ukraine. According to the findings of the correlation research, foreign direct investment and domestic lending to the private sector are the key drivers of increased innovation development and may have the most influence. In comparison to the GII leaders, only foreign direct investment has been identified as a driver of innovation development in Ukraine.

Monteiro et al. (2023) have carried out research on Brazil's performance in the GII in recent years. The objective of the research is to identify the factors and points that have significantly influenced the country's performance. To achieve this goal, the methodology involves collecting and selecting data from Brazil's official annual data in the global innovation performance ranking (GII) from 2020 to 2011. The data will be analyzed longitudinally, focusing on the seven pillars of the GII. The study will also highlight discrimination at the sub-pillar level and, in some cases, describe its indicators. The analysis, investigation, and evaluation of the data will be carried out using longitudinal survey techniques. The study aims to contribute to the understanding of the factors and points that influence innovation performance in Brazil. It is expected to provide guidance to stakeholders, policymakers, and those involved in the innovation system and innovation-oriented public policies. The focus will be on improving strategies to strengthen the relevant and significant points identified in the research. Ultimately, the aim is to create a more conducive environment for innovation and improve the performance of the Brazilian economy.

Yu and Huarng (2023) have conducted research on The GII, which is widely used as a source of data to evaluate national innovation competence by measuring various aspects of the innovation ecosystem in countries. While different studies have focused on various perspectives of GII, this research aims to address the research gap by examining the elements of Innovation Input and Innovation Output. To achieve this, the study has analyzed the GII yearly dataset spanning from 2016 to 2020 using fsQCA analysis of multiple years of data. The study has identified two types of solutions: one that combines all the antecedents except Market sophistication, and another that combines all antecedents for both outcomes, i.e., Knowledge and technology outputs and Creative outputs, respectively, through the fsQCA analysis. The findings of this study provide valuable insights into GII and justify.

Macit (2024) studied to evaluate and rank the innovation performances of 39 Asian countries using CILOS and CoCoSo methods, based on data and components from the GII 2022 report. The evaluation criteria include seven components, with infrastructure being the most significant. The CoCoSo method reveals that Singapore, Korea, Japan, and China have the highest innovation performances. The study also finds a positive correlation between the rankings obtained from the analysis and the GII report.

The previously mentioned section discussed studies conducted on GII since 2009. The data pertaining to GII used in these studies, the year of study, and the methodology employed have been summarized in the table provided below. Studies on GII since 2009 have been mentioned above. Table 2.1 below summarizes the GII data used in these studies, the method, and the year the study was conducted.

Tablo 2.1. The Summary of the Studies about GII

Author(s)	Data Used	Method(s)	Year
Ünlükaplan (2009)	2006 Economic Development Data in European Union Member Countries	Canonical Correlation	2009
Ersöz (2009)	2007 European Innovation Scoreboard Comparative Analysis Report	Hierarchical Clustering Analysis	2009
Kijek and Kijek (2010)	2005 and 2006 GII Data	Statistical analysis	2010
Chen et al. (2011)	From 1998 to 2005 National Innovation System Data	Statistical analysis	2011
Ye Sheng and Wong (2012)	2011 GII Data	Canonical correlation analysis	2012
Mahroum and Al-Saleh (2013)	2011 GII Data	correlation analysis	2013
Karagöz (2013)	From 2007 to 2012 GII Data	Data mining	2013
Şimşit et al. (2014)	2013 GII Data	Data mining	2014
Tezcan (2015)	Innovation Union Scoreboard 2012	cluster analysis and multidimensional scaling analysis	2015
Usman and Liu (2015)	2014 GII Data	Data mining	2015
Kononova (2015)	2013 GII Data	Correlation analysis	2015
Sohn et al. (2016)	2013 GII Data	Statistical analysis	2016
Hancioğlu (2016)	From 2011 to 2015 GII Data	Canonical correlation analysis	2016
Crespo and Crespo (2016)	2015 GII Data	fuzzy-set qualitative comparative analysis	2016
Albulescu and Drăghici (2016)	From 2009 to 2012 GII Data	panel data analysis	2016
Franco and de Oliveira (2017)	From 2008 to 2013 GII Data	Regression Analysis	2017
Khedhaouria and Thurik (2017)	From 2012 to 2015 GII Data	Statistical analysis	2017

Turkmen and Aynaoglu (2018)	From 2009 to 2017 GII Data	Regression Analysis	2018
Jankowska et al. (2017)	2015 GII Data	Cluster analysis	2017
Hamidi and Berrado (2018)	2017 GII Data	Cluster analysis and feature selection	2018
Pop and Pop (2018)	2017 GII Data	Theoretical	2018
Pençe et al. (2019)	2016 GII Data	ANN Method	2019
Ayçin and Çakın (2019)	2018 GII Data	Entropy and MABAC	2019
Gürtuna and Polat (2020)	2018 GII Data	Cluster analysis	2020
Layos and Peña (2020)	2019 GII Data	Statistical analysis	2020
Altıntaş (2020)	2019 GII Data	Entropy	2020
Cui et al. (2020)	From 2019 to 2020 GII Data	Random forest/feature selection and artificial neural network	2020
Omer et al. (2020)	From 2014 to 2019 GII Data	Gaussian Process model, Extremely Gradient Boosted Trees (XGBoost) model, Support Vector Machine (SVM) model, Random Forests model	2020
Doğruel and Fırat (2021)	2016 GII Data	Regression tree analysis, and linear regression analysis	2021
Famalika and Sihombing (2021)	2018 GII Data	K-means method and k-medians method	2021
Huarng and Yu (2022)	2020 GII Data	fuzzy set qualitative comparative analysis and structural qualitative association	2022
Monteiro et al. (2023)	From 2011 to 2020 GII Data	Longitudinal survey techniques	2023
Yu and Huarng (2023)	From 2016 to 2020 GII Data	Fuzzy sets/Qualitative Comparative Analysis	2023
Dobrovolska et al. (2023)	From 2011 to 2020 GII Data	Correlation analysis	2023
Bate et al. (2023)	2020 GII Data	multistage and multimodal analyses such as hierarchical regression, multiple linear regressions, and ANOVA	2023
Macit (2024)	2022 GII Data	CILOS and CoCoSo methods	2024

Upon conducting a thorough examination of studies in the literature, it was observed that the majority of them were carried out using 1-year GII data, while only a few were conducted between specific years. In contrast, our study involved the analysis of GII data from 2012, when the GII score was first calculated and reported in great detail, up until 2022. Furthermore, while previous studies in the literature focused on certain parameters, our study was conducted comprehensively by incorporating all parameters with only a few exceptions. Unlike previous literature, the parameters were selected and the GII score estimated using multiple methods. In addition, an interface for predicting the GII score was designed, which is not a widely addressed topic in the academic literature.



3. MATERIAL AND METHOD

3.1. Materials

The data utilized in this study was sourced from the yearly reports of the GII ("<https://www.globalinnovationindex.org> "). The study meticulously gathered GII data for all countries spanning the years 2012-2022. It is noteworthy that the GII, being a comprehensive global comparative study, encompasses all countries, albeit some nations were not featured in specific years.

The GII is composed of two sub-indexes, namely the innovation input sub-index and the innovation output sub-index. These sub-indexes depend on different enablers or foundations. In the preceding sections, a framework of GII is demonstrated in Figure 1.4. The indexes that form GII are coded to make them easier to study. GII consists of 7 sub-indexes coded from C1 to C7. Three sub-indexes make up each of the 7 sub-indexes and they are coded according to which sub-index they are linked to. For instance, the sub-indexes that make up the C1 index are C1.1., C1.2. and C1.3. The sub-indexes or indicators that constitute these sub-indexes are coded based on the sub-index code to which they are linked. For instance, the sub-indexes that make up the C1.1. index are C1.1.1. and C1.1.2. Table 3.1 below shows the coding of all parameters in this study. The index names in this table have been standardized, which means that although the definition of that index was the same in some years, its name was different. These differences were identified by examining the annual GII reports in detail, and the index names were made the same.

Table 3.1. Indexes Codes and Names

Index Code	Index Name	Index Code	Index Name	Index Code	Index Name
C1	Institutions	C3.3.	Ecological sustainability	C6	Knowledge and technology outputs
C1.1.	Political environment	C3.3.1.	GDP/unit of energy use	C6.1.	Knowledge creation
C1.1.1.	Political and operational stability	C3.3.2.	Environmental performance	C6.1.1.	Patents by origin/bn PPP\$ GDP
C1.1.2.	Government effectiveness	C3.3.3.	ISO 14001 environmental certificates/bn PPP\$ GDP	C6.1.2.	PCT patents by origin/bn PPP\$ GDP
C1.2.	Regulatory environment	C4	Market sophistication	C6.1.3.	Utility models by origin/bn PPP\$ GDP
C1.2.1.	Regulatory quality	C4.1.	Credit	C6.1.4.	Scientific and technical articles/bn PPP\$ GDP
C1.2.2.	Rule of law	C4.1.1.	Ease of getting credit	C6.1.5.	Citable documents H index
C1.2.3.	Cost of redundancy dismissal	C4.1.2.	Domestic credit to private sector	C6.2.	Knowledge impact
C1.3.	Business environment	C4.1.3.	Microfinance gross loans, % GDP	C6.2.1.	Labor productivity growth, %
C1.3.1.	Ease of starting a business	C4.2.	Investment	C6.2.2.	New businesses/th pop.

C1.3.2.	Ease of resolving insolvency	C4.2.1.	Ease of protecting minority investors	C6.2.3.	Software spending, % GDP
C2	Human capital & research	C4.2.2.	Market capitalization, % GDP	C6.2.4.	ISO 9001 quality certificates/bn PPP\$ GDP
C2.1.	Education	C4.2.3.	Venture capital recipients, deals/bn PPP\$ GDP	C6.2.5.	High-tech manufacturing, %
C2.1.1.	Expenditure on education, % GDP	C4.3.	Trade, diversification, and market scale	C6.3.	Knowledge diffusion
C2.1.2.	Government funding/pupil, secondary, % GDP/cap	C4.3.1.	Applied tariff rate, weighted avg., %	C6.3.1.	Intellectual property receipts, % total trade
C2.1.3.	School life expectancy, years	C4.3.2.	Intensity of local competition	C6.3.2.	High-tech exports
C2.1.4.	PISA scales in reading, maths and science	C4.3.3.	Domestic market scale, bn PPP\$	C6.3.3.	ICT services exports
C2.1.5.	Pupil-teacher ratio, secondary	C5	Business sophistication	C6.3.4.	Foreign direct investment net outflows
C2.2.	Tertiary education	C5.1.	Knowledge workers	C7	Creative outputs
C2.2.1.	Tertiary enrolment, % gross	C5.1.1.	Knowledge-intensive employment, %	C7.1.	Intangible assets
C2.2.2.	Graduates in science and engineering, %	C5.1.2.	Firms offering formal training, %	C7.1.1.	Trademarks by origin/bn PPP\$ GDP
C2.2.3.	Tertiary inbound mobility, %	C5.1.3.	GERD performed by business, % GDP	C7.1.2.	Industrial designs by origin/bn PPP\$ GDP
C2.3.	Research and development (R&D)	C5.1.4.	GERD financed by business, %	C7.1.3.	ICTs and organizational model creation
C2.3.1.	Researchers, FTE/mn pop.	C5.1.5.	Females employed w/advanced degrees, %	C7.2.	Creative goods and services
C2.3.2.	Gross expenditure on R&D, % GDP	C5.2.	Innovation linkages	C7.2.1.	Cultural and creative services exports, % total trade
C2.3.3.	Global corporate R&D investors, top 3, mn US\$	C5.2.1.	University-industry R&D collaboration	C7.2.2.	National feature films/mn pop.
C2.3.4.	QS university ranking, top 3	C5.2.2.	State of cluster development and depth	C7.2.3.	Entertainment and media market/th pop.
C3	Infrastructure	C5.2.3.	GERD financed by abroad, % GDP	C7.2.4.	Printing and other media, % manufacturing
C3.1.	Information and communication technologies (ICTs)	C5.2.4.	Joint venture/strategic alliance deals/bn PPP\$ GDP	C7.2.5.	Creative goods exports, % total trade
C3.1.1.	ICT access	C5.2.5.	Patent families/bn PPP\$ GDP	C7.3.	Online creativity
C3.1.2.	ICT use	C5.3.	Knowledge absorption	C7.3.1.	Generic top-level domains (TLDs)/th pop.
C3.1.3.	Government's online service	C5.3.1.	Intellectual property payments, % total trade	C7.3.2.	Country-code TLDs/th pop.
C3.1.4.	E-participation	C5.3.2.	High-tech imports, % total trade	C7.3.3.	Wikipedia edits/mn pop.
C3.2.	General infrastructure	C5.3.3.	ICT services imports, % total trade	C7.3.4.	Mobile app creation/bn PPP\$ GDP
C3.2.1.	Electricity output, GWh/mn pop.	C5.3.4.	FDI net inflows, % GDP		
C3.2.2.	Logistics performance	C5.3.5.	Research talent, % in businesses		
C3.2.3.	Gross capital formation, % GDP				

After all index names and codes were arranged, the data containing the index codes and scores were compiled in an excel file, with all years listed one under the other. When the data was examined, it was determined that some index scores were not calculated in some years. If the index was not calculated for most of the 11 years, it was excluded from the study. However, for most of the 11 years, these index scores were considered zero for the years that were not calculated. Additionally, it was determined that the GII scores of some countries were not calculated in some years. If this country score was not calculated for most of the 11 years, that country was excluded from the study. However, if calculated for most of the 11 years, index scores for countries without GII scores were considered zero. In this study, the index/parameters

in the lowest branch that constitutes GII in the hierarchy were used. The reason for this can be explained as follows: The parameter scores in the lowest branch form the upper branch. Parameter scores in a higher branch determine the parameter scores in the higher branch to which they are connected. The hierarchy goes like this; by using only the parameters in the lowest branch in the selection and estimation of parameters, the correct result were obtained without repeated applications. In addition, it was determined that there were spelling errors or different spellings in some country names in GII's annual reports in some years. For example, Korea was referred to as Republic of Korea in some years. In order to analyze the data correctly in the study, the country name was corrected as Republic of Korea in each report.

After the separation, a data set consisting of 79 indexes for 125 countries was prepared. In order to obtain more precise results, it was necessary to establish a set of common parameters and countries across all years. To this end, it was presented in Table 3.2 the number of countries and parameters for each given year.

Tablo 3.2. The Number of Countries and Parameters

Year	Number of Countries	Number of Parameters
2012	141	84
2013	142	83
2014	143	81
2015	141	79
2016	128	82
2017	127	81
2018	125	80
2019	129	80
2020	131	80
2021	132	81
2022	132	81

The study encompasses information gathered from 125 countries and 79 parameters, which were collected over 11 years. The study solely focused on the countries that were listed in Appendices A. In Appendix A, the 125 countries that were commonly used in the study are listed alphabetically from Albania to Zimbabwe. The GII reports spanning from 2012 to 2022 encompass scores assigned to 125 countries, with each nation being assigned a score ranging from 0 to 100.

3.2. Methods

This research involves the estimation of country innovation index scores from the GII 2012-2022 data using ANN and Regression methods. To identify the optimal parameters, after data acquisition and preparation, a range of classifier and feature selection methods were employed in Weka, ultimately yielding 39 key parameters. These parameters were then utilized in the Matlab version R2022b program to estimate GII scores for 125 countries.

To provide a comprehensive understanding of the research methodology employed, it was included a figure that outlines the various steps involved in the study. The method for each step specified in the Figure 3.1 is explained in detail in the following sections.

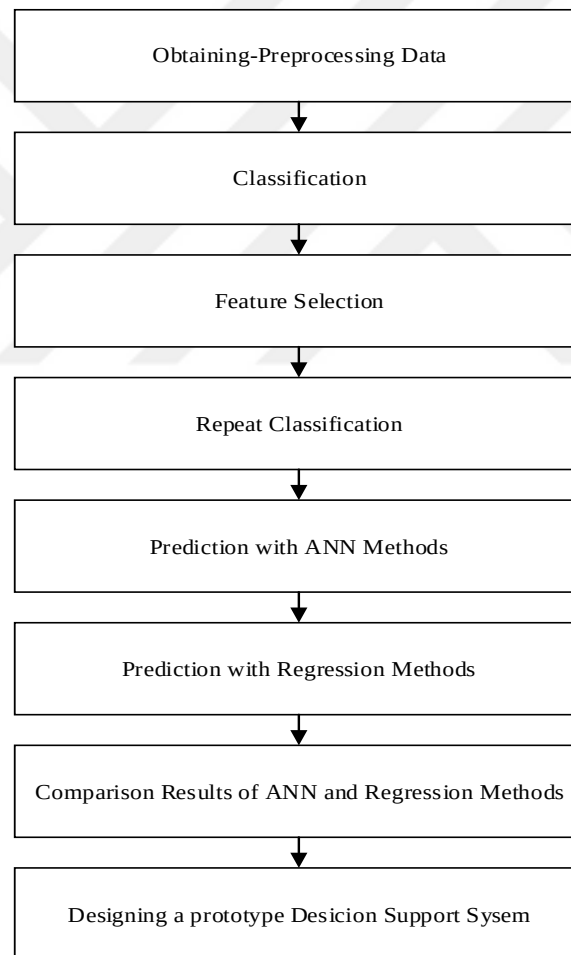


Figure 3.1. The Methodology of This Research

3.2.1. Classification Method

The definition of classification involves the process of categorizing input data instances into predefined classes or categories based on their features. This is achieved by training a classifier

using labeled training data, where each instance is associated with a known class label. The classifier learns patterns and relationships in the data to make predictions about the class labels of unseen instances. Classification is performed before feature selection in order to assign labels or categories to input data instances based on their features or attributes. Classification involves building a model that learns the relationship between the input features and the corresponding class labels, allowing it to predict the class label of new, unseen instances. The importance of classification before feature selection lies in the fact that feature selection aims to improve the performance and efficiency of the classification model by identifying and retaining the most relevant and discriminative features. However, without an initial classification step, it would be challenging to determine which features are relevant or useful for distinguishing between different classes (Hastie, Tibshirani, Friedman, and Friedman, 2009). The reason for classification in this study is to measure the performance of whether the selection was made correctly as a result of the feature selection. Figure 3.2 below shows the classification techniques and the algorithms associated with these techniques used in this study. A total of 7 classification techniques have been employed.

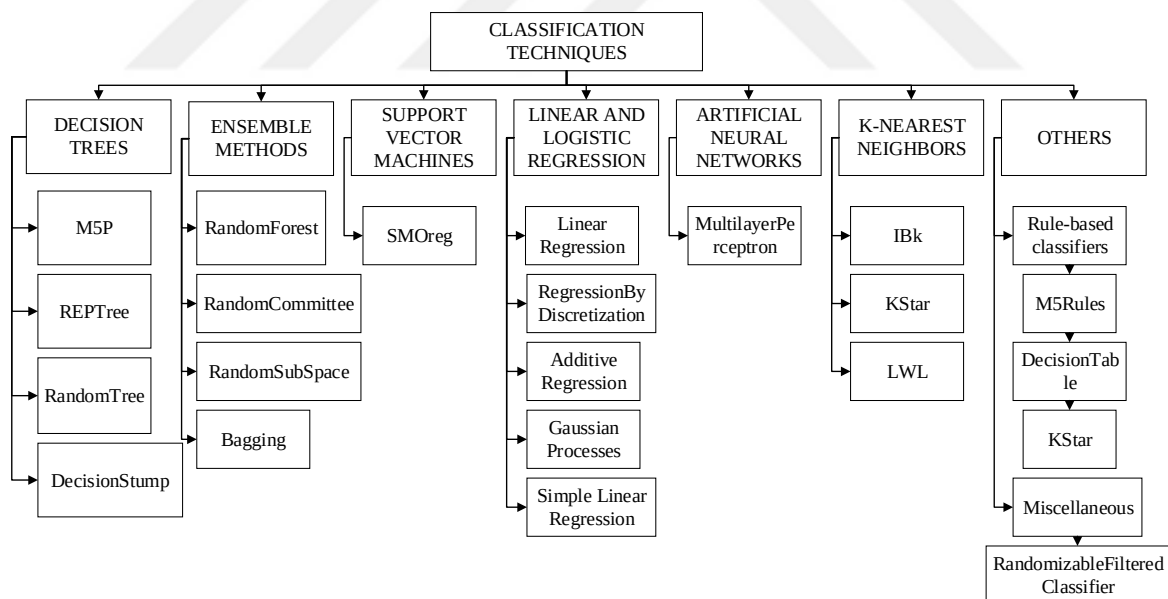


Figure 3.2. Classification Techniques

3.2.1.1. *Decision Trees*

Decision trees are widely used in ML for classifying and regressing data. They divide the feature space into different regions and assign a class label to each region. Every internal node in the tree corresponds to a feature, each branch represents a decision based on that feature, and

each leaf node represents a class label. (Quinlan, 1986). In this study, four different classifiers belonging to the decision tree method have been utilized: M5P, Reduced Error Pruning Tree (REPTree), Random Tree, and Decision Stump. Detailed information and operational algorithms of these classifiers are explained below in sequence.

The M5P algorithm was developed particularly for regression situations in which the goal is to predict a continuous numerical value. However, it can also be indirectly related to classification tasks through its association with decision trees and ensemble methods. The formula for the M5P algorithm involves the following steps (Quinlan, 1996):

1. Building the Tree:
 - Start with a decision tree structure.
 - At each node of the tree, select the attribute that maximizes the reduction in variance or another specified criterion.
 - Split the node based on the selected attribute.
2. Fitting Regression Models:
 - At the leaf nodes of the tree, fit regression models to predict the target variable.
 - Depending on the implementation, linear regression models or other regression techniques may be used.
3. Pruning and Optimization:
 - Optionally, prune the tree to prevent overfitting and improve generalization.
 - Optimize the tree structure and regression models based on performance metrics.
4. Combining Trees:
 - Aggregate the predictions from all trees to make the final regression prediction.
 - For example, in a simple average ensemble, the predictions from each tree are averaged to obtain the final prediction.

The REPTree classifier is a decision tree-based algorithm available in Weka. It is designed for both classification and regression tasks. REPTree constructs decision trees using a reduced-error pruning strategy, which helps to simplify the tree while maintaining predictive accuracy. The REPTree algorithm builds decision trees using the following steps (Frank and Hall, 2011):

1. Tree Construction:
 - Start with a decision tree structure.
 - At each node of the tree, select the attribute that maximizes the information gain or another specified criterion.
 - Split the node based on the selected attribute.
2. Reduced Error Pruning:
 - After the tree is fully grown, apply reduced error pruning to simplify the tree.
 - Pruning involves iteratively removing nodes from the tree while monitoring the change in predictive accuracy on a validation set.
 - Pruning continues until further simplification reduces predictive accuracy.
3. Handling Numeric Attributes:
 - REPTree handles numeric attributes by considering multiple split points and choosing the one that minimizes the error.
 - This allows REPTree to effectively handle both categorical and continuous attributes.
4. Handling Missing Values:
 - REPTree can handle missing attribute values by considering surrogate splits or using distribution-based methods.
5. Handling Class Imbalance:
 - Optionally, REPTree can adjust for class imbalance by using different error metrics or adjusting the class distribution during tree construction.

The Random Tree classifier is an algorithm available in Weka for classification tasks. This algorithm constructs decision trees using random feature selection and random split points. Unlike traditional decision tree construction criteria such as information gain, Random Tree does not rely on specific criteria for feature selection and splitting. Instead, it randomly selects features and split points, which introduces diversity into the decision tree ensemble (Ho, 1998).

The Random Tree algorithm doesn't have a straightforward mathematical formulation like some other algorithms such as linear regression or logistic regression. However, its key principles can be summarized as follows:

1. Random Feature Selection:
 - At each node of the tree, a random subset of features is selected.
 - This introduces diversity into the decision tree ensemble and prevents overfitting by reducing the correlation between individual trees.
2. Random Split Points:
 - For each selected feature, a random split point is chosen.
 - This randomness further enhances the diversity of the decision trees, making the ensemble more robust and less sensitive to noise in the data.
3. Decision Tree Construction:
 - Decision trees are constructed using the randomly selected features and split points.
 - The trees may vary in structure due to the random selection process, leading to a diverse set of classifiers in the ensemble.
4. Ensemble Combination:
 - Predictions from individual trees are combined to make the final prediction.
 - For classification tasks, this typically involves using majority voting, where the class with the most votes among the trees is selected as the final prediction.

The Decision Stump classifier is a classification algorithm available in Weka. This algorithm represents a very simple decision tree that operates on a single feature. It consists of only one node, and hence, it's called a "stump." Decision Stump builds a decision tree for each feature, and these trees can be combined to create more complex models. The formulation of the Decision Stump algorithm is straightforward since it represents a decision tree operating on a single feature. The formulation can be summarized as follows (Dietterich, 2002):

- Feature Selection: Select the best feature for splitting based on a certain criterion. For example, the feature that provides the highest information gain may be chosen.
- Split Point Selection: Determine the optimal split point for the selected feature. This could be, for example, the midpoint of feature values or a threshold value.
- Classification: Split the dataset into two groups based on the selected feature and split point. Determine the most common class for each group, and use these classes as predictions.

Decision Stump is a simple yet effective classifier that is frequently used as a foundation for more complicated models or as a baseline algorithm for comparing different classifiers.

3.2.1.2. Ensemble Methods

Ensemble methods combine multiple individual models to improve prediction accuracy. Common ensemble methods include Bagging, Boosting, and Random Forest. These methods leverage the diversity of individual models to make more accurate predictions (Dietterich, 2000). In this study, four different classifiers belonging to the decision tree method have been utilized: RandomForest, RandomCommittee, RandomSubSpace, and Bagging. Detailed information and operational algorithms of these classifiers are explained below in sequence.

Random Forest is a classification algorithm available in Weka. It is an ensemble method that combines multiple decision trees. The algorithm works by training multiple decision trees on the dataset, obtaining predictions from each tree, and then aggregating the results to make a final prediction. This approach helps to reduce overfitting and often performs better than a single decision tree.

The formula for the Random Forest algorithm involves the following steps (Breiman, 2001):

- Randomly select N samples from the dataset with replacement (bootstrap sampling).
- Grow a decision tree using the selected samples. At each node:
- Randomly select m features from the total features.
- Split the node using the feature that provides the best split according to a specified criterion (e.g., Gini impurity or information gain).
- Repeat steps 1 and 2 to create multiple decision trees.
- Aggregate the predictions from all decision trees to make the final prediction:
- For classification, use majority voting (the class with the most votes).
- For regression, use the average of the predictions.

The Random Forest algorithm allows for tuning parameters such as the number of trees (N), the number of features to consider at each split (m), and the criterion for splitting nodes. By adjusting these parameters, you can optimize the performance of the Random Forest classifier for your specific dataset and requirements.

RandomCommittee is an ensemble classifier available in Weka. This algorithm combines multiple base classifiers to form a decision-making mechanism. Each base classifier is typically trained differently using the same underlying algorithm, and then their predictions are aggregated. By combining the strengths and weaknesses of each base classifier, RandomCommittee aims to create a stronger and more generalized classifier. RandomCommittee is an ensemble method that utilizes multiple base classifiers. It creates diversity among the base classifiers by training them on different subsets of the data or by random selection of training instances. During classification, each base classifier predicts the class label of the input instance, and the final prediction is often determined through a voting mechanism, such as majority voting. The formulation of RandomCommittee relies on the basic principles of ensemble learning and typically involves the following steps (Dietterich, 2000):

1. **Base Model Selection:** RandomCommittee creates multiple base models using the same underlying algorithm but with different training data. These base models can be randomly selected or trained on different subsets of the data.
2. **Decision Making:** Each base model predicts the class label of the input instance independently. The final prediction is often determined by aggregating the predictions of all base models, such as through majority voting.

While RandomCommittee's formulation may not have a specific mathematical formula, its principles are grounded in ensemble learning techniques aimed at improving classification performance through model diversity.

RandomSubSpace is a classification algorithm available in Weka that belongs to the family of ensemble methods. This algorithm constructs multiple classifiers, each of which is trained on a random subset of features. By creating diversity among the classifiers, RandomSubSpace aims to improve the robustness and generalization performance of the ensemble. RandomSubSpace algorithm randomly selects a subset of features and then trains a classifier using only these selected features. It repeats this process multiple times, generating several classifiers, each operating in a different feature subspace. During classification, each classifier independently predicts the class label of the input instance. The final prediction is often determined through a voting mechanism, such as majority voting, where the most common class prediction among the classifiers is selected as the final output. The RandomSubSpace algorithm does not have a

specific mathematical formula per se, but its formulation involves the following steps (Ho, 1998):

1. Feature Subset Selection: RandomSubSpace randomly selects a subset of features from the original feature set.
2. Classifier Training: A classifier is trained using the selected feature subset. This process is repeated multiple times to generate several classifiers, each operating in a different feature subspace.
3. Classification: During classification, each classifier independently predicts the class label of the input instance using its corresponding feature subset.
4. Voting Mechanism: The final prediction is often determined through a voting mechanism, such as majority voting, where the most common class prediction among the classifiers is selected as the final output.

While there may not be a specific mathematical formula for RandomSubSpace, its principles are grounded in ensemble learning techniques aimed at improving classification performance through feature subset diversity.

Bagging (Bootstrap Aggregating) is an ensemble learning method that aims to improve the performance of ML models by reducing variance and increasing stability. It works by training multiple base learners on different bootstrap samples of the training data and then combining their predictions to make the final decision. The idea behind Bagging is to create diversity among the base learners, which helps to reduce overfitting and improve generalization. Bagging involves the following steps (Breiman, 1996):

1. Bootstrap Sampling: Multiple bootstrap samples are drawn with replacement from the original training dataset. Each bootstrap sample has the same size as the original dataset but may contain duplicate instances and omit some instances.
2. Base Learner Training: A base learner (often a simple model like decision trees) is trained on each bootstrap sample independently. This results in multiple base learners, each trained on a slightly different subset of the data.
3. Prediction Aggregation: During prediction, each base learner independently predicts the class label of the input instance. The final prediction is then determined by aggregating

the predictions of all base learners. This can be done through various methods such as majority voting for classification or averaging for regression.

3.2.1.3. Support Vector Machines

Support Vector Machine (SVM) is an effective supervised learning technique for classification tasks. It works by identifying the best hyperplane to split data points from distinct classes in the feature space. The SVM seeks to optimize the margin between classes while minimizing classification errors (Cortes and Vapnik, 1995). In this study, the SMOreg classifier belonging to the SVM method has been employed. Detailed information and the operational algorithm of this classifier are described below.

The SMOreg classifier in Weka is based on the Sequential Minimal Optimization (SMO) algorithm, which is commonly used for regression tasks within SVMs. The SMO algorithm, introduced by Platt (1999), is an efficient algorithm for training SVMs. It is specifically designed for solving the optimization problem associated with SVMs in a sequential and iterative manner. SMO breaks down the quadratic programming problem into smaller sub problems, making it computationally efficient and allowing it to handle large datasets. In the context of regression (SMOreg), SMO is used to train SVMs for regression tasks. The goal is to find the optimal hyperplane that best fits the training data while maximizing the margin and minimizing the error. This hyperplane represents the regression function, which can then be used to make predictions for unseen data points. SMOreg involves the following steps (Platt, 1999):

1. Initialization: Initialize the Lagrange multipliers α_i and bias term b to zeros or small random values.
2. Selection of Two Multipliers: Select two Lagrange multipliers α_i and α_j to optimize in each iteration. These multipliers are selected using a heuristic, such as the maximum step size or maximum error.
3. Compute Error and Kernel Values: Compute the error E_i for the selected multipliers α_i and α_j , and precompute the kernel values K_{ij} for the corresponding training instances x_i and x_j .
4. Update Multipliers: Update the selected Lagrange multipliers α_i and α_j using the optimization procedure. This involves selecting a pair of multipliers that violate the Karush-Kuhn-Tucker conditions, computing the optimal step sizes, and updating the

multipliers accordingly. Karush-Kuhn-Tucker conditions are a set of necessary and sufficient conditions for optimality in constrained optimization problems. These conditions define the relationships between the objective function, the constraints, and the Lagrange multipliers associated with the constraints.

5. Update Bias Term: Update the bias term b using the new Lagrange multipliers α_i and α_j and their corresponding support vectors.
6. Convergence Check: Check for convergence by monitoring the change in the Lagrange multipliers and bias term. If the change falls below a predefined threshold or the maximum number of iterations is reached, terminate the algorithm.
7. Repeat: Repeat steps 2-6 until convergence is achieved or until the optimization problem is solved within a predefined tolerance.
8. Output: Return the learned parameters, including the Lagrange multipliers α_i and bias term b , which define the optimal hyperplane for regression.

These steps describe the basic outline of the SMO algorithm for training SVMs for regression tasks. The algorithm efficiently solves the optimization problem associated with SVMs by iteratively updating pairs of Lagrange multipliers while maintaining the Karush-Kuhn-Tucker conditions.

3.2.1.4. Linear and Logistic Regression

Linear and logistic regression are statistical methods used for regression and classification tasks, respectively. Linear regression uses a linear equation to represent the relationship between dependent and independent variables, whereas logistic regression use a logistic function to model the likelihood of binary outcomes (Hosmer Jr, Lemeshow, and Sturdivant, 2013). In this study, five classifiers belonging to the Linear and Logistic Regression method have been utilized: Linear Regression, RegressionByDiscretization, Additive Regression, Gaussian Processes and Simple Linear Regression. Detailed information and operational algorithms of these classifiers are explained below in sequence.

Linear regression is a fundamental statistical method for describing the connection between a dependent variable and one or more independent variables. It seeks to identify the best-fitting linear equation that predicts the dependent variable based on the independent variables. Linear

regression models the relationship between the independent variable X and the dependent variable Y using a linear equation of the form (Montgomery, Peck, and Vining, 2021):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (3.1)$$

Where:

- y is the dependent variable.
- $x_1, x_2, \dots, x_n$ are the independent variables.
- $\beta_0, \beta_1, \dots, \beta_n$ are the coefficients (or parameters) of the model.
- ϵ is the error term, representing the difference between the observed and predicted values.

The goal of linear regression is to estimate the coefficients $\beta_0, \beta_1, \dots, \beta_n$ that minimize the sum of squared residuals (the squared differences between the observed and predicted values).

Linear Regression involves the following steps (Montgomery et al., 2021):

1. **Data Preparation:** Prepare the dataset by splitting it into independent variables (features) and the dependent variable (target).
2. **Model Training:** Fit a linear regression model to the training data using the method of least squares or other optimization techniques. This involves estimating the coefficients $\beta_0, \beta_1, \dots, \beta_n$ that minimize the sum of squared residuals.
3. **Model Evaluation:** Evaluate the model's performance using measures like mean squared error (MSE), R-squared, and modified R-squared. This step evaluates how well the model fits the data and how accurately it predicts the dependent variable.
4. **Prediction:** Use the trained model to make predictions on new or unseen data by applying the linear equation to the independent variables.
5. **Interpretation:** Interpret the model's coefficients to better understand the link between the independent and dependent variables. Positive coefficients imply a positive relationship, whilst negative coefficients suggest a negative one.

The RegressionByDiscretization classifier in Weka is a method that discretizes numeric attributes and then applies a regression algorithm to the discretized data. This method works especially well for regression jobs where there are nonlinear relationships between the

independent and dependent variables, or when there are anomalies in the data. The RegressionByDiscretization algorithm consists of the following steps (Hall et al., 2009):

1. Data Preprocessing: The algorithm starts by preprocessing the dataset, identifying numeric attributes that need to be discretized. Discretization involves partitioning the range of each numeric attribute into intervals or bins.
2. Discretization: Once the numeric attributes are identified, the algorithm discretizes them by dividing their range into a specified number of intervals or by using a specific discretization method such as equal-width or equal-frequency discretization.
3. Regression Modeling: After discretization, the algorithm applies a regression algorithm to the discretized data to build a predictive model. Common regression algorithms used in this context include linear regression, decision trees, or neural networks.
4. Model Evaluation: The performance of the regression model is evaluated using appropriate evaluation metrics such as MSE, RMSE, or coefficient of determination (R-squared). This step helps assess how well the model fits the data and how accurately it predicts the dependent variable.
5. Prediction: Once trained and assessed, the model can be used to make predictions about new or unknown data. The discretized numeric attributes in the new data are transformed using the same discretization intervals as in the training data, and then the regression model is applied to predict the target variable.

The Additive Regression classifier in Weka is a method used for fitting additive regression models to data. Additive regression models are a type of generalized linear model where the predictor variables are combined in a linear or nonlinear manner to predict the response variable. Additive Regression models aim to predict a response variable, Y based on a set of predictor variables $x_1, x_2, \dots, x_p$. The model can be represented as (Hastie et al., 2009):

$$Y = f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \epsilon \quad (3.2)$$

Where:

- f_i represents the smooth function of the predictor variable X_i .
- ϵ is the error term.

The goal is to estimate the smooth functions f_i that best fit the data while minimizing the error term. The Additive regression model consists of the following steps (Hastie et al., 2009) :

1. Data Preprocessing: Preprocess the dataset by handling missing values, scaling the features, and encoding categorical variables if necessary.
2. Initialization: Initialize the additive regression model with some initial estimates of the smooth functions f_i . This can be done using simple functions or by fitting separate models to each predictor variable.
3. Model Fitting: Fit the additive regression model to the data by iteratively updating the smooth functions f_i to minimize the loss function, such as the MSE. This is typically done using optimization techniques such as gradient descent or back fitting.
4. Variable Selection: Optionally, perform variable selection to identify the subset of predictor variables that have the most significant impact on the response variable. This can help simplify the model and improve its interpretability.
5. Model Evaluation: Assess the additive regression model's performance using relevant metrics such as MSE, R-squared, or cross-validated error. This step assesses the model's fit to the data and its ability to predict the response variable.
6. Prediction: Once the model is trained and evaluated, it can be used to make predictions on new or unseen data by applying the estimated smooth functions to the predictor variables.

Gaussian Processes classifier in Weka is a probabilistic method used for regression and classification tasks. Gaussian Processes model the relationship between input variables and output variables as a multivariate Gaussian distribution. Gaussian Processes model the relationship between input variables x and output variables y as a joint Gaussian distribution (Rasmussen and Williams, 2006):

$$p(y|x) = N(y|0, K) \quad (3.3)$$

Where:

- y is the vector of output variables.
- x is the matrix of input variables.

- K is the covariance matrix (kernel) that captures the relationships between input variables.

The kernel function K defines the similarity between pairs of input variables, and it is typically parameterized by hyperparameters that control the smoothness and complexity of the model. The Gaussian Process model consists of the following steps (Rasmussen and Williams, 2006) :

1. Data Preprocessing: Preprocess the dataset by handling missing values, scaling features, and encoding categorical variables as needed.
2. Model Training: Fit the Gaussian Processes model to the training data by estimating the parameters of the kernel function. This involves maximizing the likelihood of the observed data under the Gaussian Processes model.
3. Prediction: Use the trained Gaussian Processes model to make predictions on new or unseen data. Given a set of input variables x_{test} , the model predicts the corresponding output variables y_{pred} along with the uncertainty associated with each prediction.
4. Model Evaluation: Evaluate the performance of the Gaussian Processes model using appropriate metrics such as MSE, log likelihood, or predictive log likelihood. This step evaluates how well the model fits the data and how accurately it predicts the output variables.
5. Hyperparameter Tuning: Optionally, tune the hyperparameters of the kernel function to improve the performance of the Gaussian Processes model. This can be done using techniques such as cross-validation or maximum likelihood estimation.

The Simple Linear Regression classifier in Weka is a basic statistical method used for modeling the relationship between a single independent variable and a dependent variable. It aims to find the best-fitting straight line that predicts the value of the dependent variable based on the independent variable. Simple Linear Regression models the relationship between the independent variable x and the dependent variable y using a linear equation of the form (Montgomery et al., 2021):

$$y = \beta_0 + \beta_1 x + \epsilon \quad (3.4)$$

Where:

- y is the dependent variable.
- x is the independent variable.
- β_0 is the intercept (y-intercept) of the line.
- β_1 is the slope (gradient) of the line.
- ϵ is the error term, representing the difference between the observed and predicted values.

The goal of Simple Linear Regression is to estimate the coefficients β_0 and β_1 that minimize the sum of squared residuals (the squared differences between the observed and predicted values). The Simple Linear Regression model consists of the following steps (Montgomery et al., 2021) :

1. **Data Preparation:** Prepare the dataset by splitting it into independent variables (features) and the dependent variable (target).
2. **Model Training:** Fit a simple linear regression model to the training data using the method of least squares or other optimization techniques. This involves estimating the coefficients β_0 and β_1 that minimize the sum of squared residuals.
3. **Model Evaluation:** Evaluate the performance of the model using various metrics such as MSE, R-squared, or adjusted R-squared. This step helps assess how well the model fits the data and how accurately it predicts the dependent variable.
4. **Prediction:** Once the model is trained and evaluated, it can be used to make predictions on new or unseen data by applying the linear equation to the independent variable.
5. **Interpretation:** Interpret the model's coefficients to better understand the link between the independent and dependent variables. The slope β_1 indicates the change in the dependent variable with a one-unit change in the independent variable, whereas the intercept β_0 shows the value of the dependent variable when the independent variable is zero.

3.2.1.5. Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models inspired by the biological neural networks of the human brain. They consist of interconnected nodes organized in layers. Each node performs a simple computation, and information flows through the network from input to output layers, enabling complex pattern recognition and classification (Haykin, 1998). In this

study, the Multilayer Perceptron (MLP) classifier belonging to the ANN method has been employed. Detailed information and the operational algorithm of this classifier are described below.

The MLP classifier in Weka is a form of ANN formed up of numerous layers of interconnected neurons. It is a powerful and flexible ML algorithm capable of modeling complex relationships between input and output variables. The MLP is a feedforward neural network consisting of an input layer, one or more hidden layers, and an output layer. Each neuron in the network is connected to neurons in the adjacent layers, and each connection has an associated weight. The network learns by adjusting these weights during training to minimize the error between the predicted and actual outputs. The MLP model consists of the following steps (Bishop, 1995)

1. Data Preprocessing: Preprocess the dataset by handling missing values, scaling the features, and encoding categorical variables if necessary.
2. Model Initialization: Initialize the MLP model by specifying the number of input neurons, the number of neurons in each hidden layer, the activation function for each layer, and other hyperparameters such as learning rate, momentum, and regularization.
3. Forward Propagation: Propagate the input data forward through the network to compute the predicted output. This involves calculating the weighted sum of inputs at each neuron, applying the activation function to generate the neuron's output, and passing the output to the next layer.
4. Error Calculation: Calculate the error between the predicted output and the actual output using a suitable loss function such as MSE or cross-entropy loss.
5. Backpropagation: Backpropagate the error through the network to update the weights using gradient descent optimization. This involves computing the gradient of the loss function with respect to each weight and adjusting the weights in the direction that minimizes the error.
6. Repeat: Iterate steps 3-5 for a fixed number of epochs or until convergence criteria are met. During training, the network learns to adjust its weights to minimize the error on the training data.
7. Model Evaluation: Evaluate the performance of the trained MLP model using appropriate metrics such as accuracy, precision, recall, or F1-score on a separate

validation or test dataset. This step helps assess how well the model generalizes to unseen data.

8. Prediction: Once trained and assessed, the model may be used to predict new or unknown data by forward propagating the input through the trained network and retrieving the expected output from the output layer.

The ANN methodology will be elaborated in detail in a separate section that follows in this academic study.

3.2.1.6. *K-Nearest Neighbors*

K-Nearest Neighbor (K-NN) is a simple yet effective ML algorithm used for classification and regression tasks. The basic idea behind k-NN is to classify or predict a data point based on the majority vote or average of its k nearest neighbors in the feature space. It operates on the principle that similar instances tend to belong to the same class or have similar output values (Cover and Hart, 1967). In this study, three classifiers belonging to the K-NN method have been utilized: Instance-Based k-Nearest Neighbors (IBk), KStar, and Locally Weighted Learning (LWL). Detailed information and operational algorithms of these classifiers are explained below in sequence.

The IBk classifier in Weka is an instance-based learning algorithm that uses the k-nearest neighbors approach for classification tasks. It is a simple and effective non-parametric method for classifying data points based on the majority class of their nearest neighbors. The IBk classifier works based on the principle of similarity. It classifies a new data point by finding its k-nearest neighbors in the training dataset and assigning the majority class label among those neighbors to the new data point. The IBk consists of the following steps (Han et al., 2022):

1. Data Preprocessing: Preprocess the dataset by handling missing values, scaling the features, and encoding categorical variables if necessary.
2. Initialization: Specify the value of the parameter k , which determines the number of nearest neighbors to consider during classification.
3. Distance Calculation: Calculate the distance between the new data point and each data point in the training dataset using a distance metric such as Euclidean distance, Manhattan distance, or Minkowski distance.

4. Neighbor Selection: Select the k data points with the smallest distances to the new data point as its nearest neighbors.
5. Majority Voting: Determine the majority class label among the k nearest neighbors. If there is a tie, a simple tie-breaking rule (e.g., selecting the class label of the nearest neighbor) can be applied.
6. Classification: Assign the majority class label to the new data point as its predicted class label.
7. Repeat: Repeat steps 3-6 for each new data point in the test dataset.

The KStar classifier in Weka is an instance-based learning algorithm that uses the k -nearest neighbors approach with an entropy-based distance function for classification tasks. It is similar to the IBk classifier but differs in the distance metric used for measuring the similarity between instances. The KStar classifier works based on the principle of similarity. It classifies a new data point by finding its k -nearest neighbors in the training dataset and assigning the majority class label among those neighbors to the new data point. However, unlike the IBk classifier which uses traditional distance metrics, KStar utilizes an entropy-based distance function to measure the similarity between instances (Cleary and Trigg, 1995). The operational steps of the KStar classifier closely resemble those described for the IBk classifier above.

The LWL classifier in Weka is an instance-based learning algorithm that assigns weights to training instances based on their similarity to the test instance. It is a variant of the k -NN algorithm, where the contribution of each neighbor to the classification decision is weighted according to its proximity to the test instance. The LWL classifier works based on the principle of locally weighted learning, where the prediction for a test instance is determined by considering the k nearest neighbors in the training dataset. However, instead of treating all neighbors equally, LWL assigns weights to the neighbors based on their distance from the test instance. The working steps of the LWL classifier are described below (Moore and Schneider, 1995):

1. Data Preprocessing: Prior to model training, the dataset undergoes preprocessing steps such as handling missing values, feature scaling, and encoding categorical variables.
2. Initialization: The classifier is initialized by setting the value of the k parameter, which denotes the number of nearest neighbors to be considered during the classification process.

3. Distance Calculation: The distance between the test instance and each instance in the training dataset is computed using a distance metric like Euclidean distance, Manhattan distance, or Minkowski distance.
4. Weight Calculation: Each training instance is assigned a weight based on its distance from the test instance. Common weight functions include the inverse distance weighting function or the Gaussian weighting function.
5. Weighted Classification: The class label for the test instance is determined by considering the weighted contributions of the k nearest neighbors. A weighted majority voting scheme is employed, wherein closer neighbors exert a greater influence on the classification decision.
6. Iteration: Steps 3 to 5 are iteratively performed for each test instance in the dataset.

3.2.1.7. Others

This category includes various classification techniques such as Rule-based classifiers, and miscellaneous classifiers. These methods have their own unique characteristics and are suitable for different types of data and tasks (Witten, Frank, Hall, and Pal, 2005).

Rule-based classifiers are a type of ML algorithm that makes decisions based on a set of rules derived from the training data. These classifiers operate by generating a series of if-then rules that map input features to output classes. Each rule consists of conditions that describe patterns in the input data and corresponding actions that specify the predicted class label or value (Han et al., 2022). In this study, two classifiers belonging to the Rule-based classifiers have been utilized: M5Rules and Decision Table. Detailed information and operational algorithms of these classifiers are explained below in sequence.

M5Rules is a classification algorithm in Weka that generates rules from decision trees. It builds decision trees using the M5 model tree algorithm and then converts these trees into sets of rules. Each rule corresponds to a path from the root to a leaf node in the decision tree, where the conditions along the path define the rule and the leaf node specifies the predicted class label. The advantages of rule-based classifiers and decision trees are combined in M5Rules. It leverages the interpretability of rules while benefiting from the predictive power of decision trees. The algorithm follows these steps (Quinlan, 1992):

1. **Decision Tree Construction:** M5Rules begins by constructing decision trees using the M5 model tree algorithm. This algorithm builds regression trees that consist of linear models at each leaf node and splits nodes based on the attribute that maximizes the reduction in variance.
2. **Rule Extraction:** Once the decision trees are built, M5Rules extracts rules from these trees. Each rule corresponds to a path from the root to a leaf node in the decision tree. Along this path, the conditions on the attributes form the antecedent of the rule, and the class label associated with the leaf node serves as the consequent.
3. **Rule Pruning:** The extracted rules may be pruned to improve interpretability and generalization. Pruning involves removing redundant or irrelevant rules that do not significantly contribute to the predictive performance of the classifier.
4. **Rule Set Generation:** After pruning, M5Rules generates a set of rules that collectively cover the training data. Each rule captures a distinct pattern or decision boundary in the data, providing insights into the relationships between input features and class labels.
5. **Rule Application:** During classification, M5Rules applies the generated rule set to new instances by sequentially evaluating the conditions of each rule. The first rule whose conditions are satisfied determines the predicted class label for the instance.

The DecisionTable classifier in Weka is a rule-based classifier that builds a decision table from the training data to make predictions. It converts the training instances into a table format where each row represents an instance and each column represents a feature or attribute. The decision table consists of rules that define the conditions for predicting the class label of new instances. The DecisionTable classifier is based on the concept of decision tables, which are a tabular representation of decision rules. Each row in the decision table corresponds to a unique combination of attribute values, and each column represents a different attribute or feature. The last column typically represents the class label. The algorithm follows these steps (Grzymala-Busse and Siddhaye, 2004):

1. **Data Preprocessing:** Pre-process the dataset by handling missing values, feature scaling, and encoding categorical variables if necessary.
2. **Decision Table Construction:** Construct a decision table from the training data by enumerating all possible combinations of attribute values and determining the

corresponding class label for each combination. This process involves generating decision rules that map attribute values to class labels.

3. Rule Simplification: Simplify the decision table by removing redundant or irrelevant rules to improve interpretability and efficiency. This step may involve rule pruning or rule merging techniques.
4. Rule Application: During classification, apply the decision rules from the decision table to new instances to predict their class labels. For each instance, find the rule that matches its attribute values and assign the corresponding class label.
5. Rule Evaluation: Evaluate the performance of the DecisionTable classifier using metrics such as accuracy, precision, recall, or F1-score on a validation or test dataset. This step assesses the classifier's ability to generalize to previously unseen data.

RandomizableFilteredClassifier from Miscellaneous Classifiers is a classification algorithm available in Weka that combines the functionalities of a filter and a classifier. It allows for the application of both feature selection or feature transformation techniques and a base classifier. Additionally, it offers the capability to introduce randomness in the selection of filter parameters, enabling the exploration of different configurations to improve classification performance. RandomizableFilteredClassifier consists of the following components (Hall et al., 2009):

1. Filtering: Before classification, the input dataset can be preprocessed using a filter. Filters can perform tasks such as feature selection, feature transformation, or instance filtering. The filtered dataset is then used for classification.
2. Randomization: RandomizableFilteredClassifier introduces randomness into the filter's parameter selection process. This means that during each run, the filter parameters are randomly chosen from a predefined range or distribution. This randomization allows for the exploration of different filter configurations to find the most effective one.
3. Classification: After filtering, the preprocessed dataset is passed to a base classifier for training and prediction. The base classifier can be any classifier available in Weka, such as decision trees, SVMs, or neural networks.

In the initial stage of the research, a range of ML techniques were utilized to select the parameters for estimation. These methods were employed separately to construct various

classifiers, and their outcomes were assessed for comparison. The study employed 10-fold cross-validation to evaluate the performance of each classifier, utilizing metrics such as Correlation Coefficient, Relative Absolute Error (RAE), and Root Relative Squared Error to compare the model outcomes. The performance rankings of the classifiers are presented based on these evaluation metrics.

The definitions and formulas of the classification performance metrics R, RAE, and Root Relative Squared Error (RRSE) are as follows (Witten and James, 2013):

- The correlation coefficient, denoted by R, measures the strength and direction of a linear relationship between two variables. In the context of classification performance evaluation, R indicates the degree of linear relationship between the predicted and actual class labels. Higher values of R indicate better classification performance.

$$R = \sqrt{1 - \frac{\sum((y_i - x_i))^2}{\sum((y_i - y_{ort}))^2}} \quad (3.5)$$

Here, i represents the number of data, y represents the actual value, and x represents the predicted value.

- RAE: The accuracy of a classifier's predictions is measured by the RAE, which is computed as the mean absolute error divided by the mean absolute error of a basic predictor that consistently predicts the target variable's mean value. It offers a comparative assessment of forecast precision.

$$RAE = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |y_i - \bar{y}|} \quad (3.6)$$

where:

- x_i represents the predicted class label for instance i ,
- y_i represents the actual class label for instance i ,
- $\bar{y}$ is the mean of the actual class labels,
- n is the number of instances.

- **RRSE:** The square root of the MSE divided by the MSE of a simple predictor that consistently predicts the mean value of the target variable yields the RRSE, a measure of the relative accuracy of a regression or classification model. It offers a comparative assessment of forecast precision.

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.7)$$

where:

- y_i is the actual class label for instance i ,
- $\bar{y}_i$ is the predicted class label for instance i ,
- $\bar{y}$ is the mean of the actual class labels,
- n is the number of instances.

3.2.2. Feature Selection Method

Feature selection, also referred to as variable or attribute selection in the data, involves the process of selecting a subset of distinct features (variables, predictors) to be utilized in constructing ML and data science models. During predictive model design, feature selection is implemented to lower the number of input variables. This reduction serves two purposes: minimizing the computational cost of modeling and, in certain scenarios, optimizing the model's performance (Brownlee, 2019).

Supervised learning feature selection techniques can be categorized into Filter and Wrapper approaches. Filter approaches utilize measures such as consistency, distance, correlation, and dependency information, which are familiar characteristics of the training data. The learning algorithm and feature selection algorithm remain independent of each other, avoiding any bias. The primary benefit of the Filter approach is that it filters out unnecessary attributes before induction, making it useful for datasets with multiple features. On the other hand, the Wrapper approach applies the induction algorithm to determine feature subset quality. Once selected, the features are induced into the classifier for prediction. The Wrapper approach utilizes ML algorithms to obtain the most optimal feature subset (Gupta, 2021).

The filter methods used in this study are explained in detail below.

1- Correlation-based Feature Selection is a method for feature selection in Weka that uses a filter approach. It assesses the correlation between features and class labels, as well as the intercorrelation between features. The goal of Correlation-based Feature Selection is to identify feature subsets that are highly correlated with the class labels while minimizing redundancy among them (Hall and Smith, 1999). The algorithm of this method is as follows:

- Calculate the correlation between each feature and the class labels. This can be done using statistical measures like Pearson correlation coefficient or mutual information.
- For each subset of features, compute the average correlation between the features in the subset and the class labels. This will measure how well the subset as a whole is correlated with the class labels.
- Calculate the average pairwise correlation between features in the subset to measure redundancy among the selected features.
- Combine the subset-class correlation and the subset-feature correlation to compute a merit score for each subset. The merit score reflects the trade-off between the subset's correlation with the class labels and the redundancy among its features.
- Choose the subset with the highest merit score as the selected feature subset.

The mathematical formulas of this method are as follows (Hall and Smith, 1999):

The merit score ($\text{Merit}(S)$) for a feature subset S is calculated as:

$$\text{Merit}(S) = (k * \text{Corr}(S, \text{Class})) / (\sqrt{(k + (k * (k - 1) * \text{Corr}(S, \text{Feat}_i)^2))}) \quad (3.8)$$

Where:

k is the number of features in subset S .

$\text{Corr}(S, \text{Class})$ is the correlation between subset S and the class labels.

$\text{Corr}(S, \text{Feat}_i)$ is the correlation between feature i and the other features in subset S .

2- ClassifierSubsetEval (CSE) is another filter method used for feature selection. It assesses feature subsets based on their performance with a specific classifier. This evaluation helps to determine the relevance and importance of features for classification tasks. CSE measures how well each subset of features can discriminate between different classes in the dataset. By comparing the classification performance of various feature subsets, CSE identifies the most informative features for the given classification task (Hall et al., 2009).

The algorithm of this method is as follows (Guyon and Elisseeff, 2003):

- Subset Selection: Begin a loop to create all possible feature combinations, and train a classifier with each subset.
- Evaluation of Classification Performance: Use specific performance metrics like accuracy, precision, recall, or F1 score to measure the classification performance for each subset.
- Calculation of Average Performance Metrics: Once you've measured the classification performance for each subset, calculate the average performance metrics for all subsets to evaluate the overall classification ability of each.
- Selection of the Best Subset: Choose the feature subset with the highest performance metrics as the selected subset.

The mathematical formulas of this method are as follows:

The average performance metric (AvgPerf) for subsets is calculated as:

$$\text{AvgPerf} = (1/N) * \sum \text{Performance}(\text{Subset}_i) \quad (3.9)$$

Where:

N is the total number of feature subsets.

Performance (Subset_i) represents the classification performance of the i-th feature subset.

3- WrapperSubsetEval is a mechanism used to evaluate feature subsets, although it is not exactly a filter method. Unlike filter methods that assess features independently of any learning algorithm, wrapper methods evaluate subsets based on their performance with a specific learning algorithm. WrapperSubsetEval selects the most relevant features by testing different subsets with a classifier and selecting the subset that yields the best performance. It employs a learning algorithm, such as a classifier, to assess the effectiveness of feature subsets. WrapperSubsetEval iteratively evaluates subsets by training and testing them with the classifier, considering their performance metrics such as accuracy, precision, or F1 score. This iterative process allows WrapperSubsetEval to explore the space of feature subsets and select the one that maximizes the chosen performance metric (Kohavi and John, 1997).

The algorithm of this method is as follows (Kohavi and John, 1997):

- One approach to feature selection is Subset Generation, which involves creating various subsets of features for evaluation. This can be done by either iterating through all possible combinations of features or using a heuristic search algorithm to explore the feature space.
- Once feature subsets are generated, the next step is Evaluation of Subset Performance. This involves training and testing a specified learning algorithm with the subset of features. First, the dataset is divided into training and testing sets. Then, the algorithm is trained using the selected subset of features on the training set. Finally, its performance is evaluated on the testing set.
- A Performance Metric Calculation is then performed for each feature subset. This metric provides a quantifiable measure of how well the learning algorithm performs when using the selected subset of features. Examples of performance metrics include accuracy, precision, recall, and F1 score.
- Finally, Subset Selection is used to identify the feature subset that yields the best performance according to the chosen performance metric. This subset is then considered as the final selected subset of features.

While there are no specific mathematical formulas associated with WrapperSubsetEval, mathematical formulas may be used to calculate the chosen performance metric, such as accuracy.

The wrapper methods are explained in detail below.

1- The Best First method is utilized in Weka for feature selection. It implements a greedy search approach to traverse the feature subset space. At each step, the learning algorithm's performance is evaluated on different feature subsets, and the one that maximizes a specific evaluation criterion is selected (Hall and Holmes, 2003).

The "Best First" algorithm in Weka doesn't have a mathematical formula. This algorithm acts as a search strategy in a search space, where feature subsets are evaluated using a performance metric. However, the performance metrics used for feature selection are usually classification or regression performance criteria. For example, in classification scenarios, metrics such as Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC) are used to measure

the classification performance of a model. The Best First algorithm selects feature subsets based on these performance metrics, with the goal of selecting the features that provide the best performance. Despite this, the Best First algorithm doesn't rely on a formal mathematical formula. Instead, it operates as a search strategy that evaluates feature performance using performance metrics. Therefore, there is no mathematical formula at the heart of this algorithm. The "Best First" feature selection method in Weka is a powerful tool that selects features by evaluating different feature subsets and choosing the subset that yields the best performance according to a specified evaluation metric. Although the specific implementation of "Best First" in Weka is not explicitly detailed in the literature, the general process follows several steps:

- **Starting Point:** The process begins with either a blank or pre-existing set of features.
- **Generating Candidates:** At each stage, the process creates new feature subsets by incorporating one or more features not already present in the subset. The features come from a pool of available options.
- **Choosing the Best:** The process evaluates each candidate subset based on a performance metric derived from a classification or regression algorithm. The best-performing subset, whether it has the highest metric or shows the most improvement, is chosen.
- **Updating the State:** The selected subset becomes the new starting point, and the process repeats.
- **Stopping Point:** The process continues until a stopping criterion is met, such as reaching a specified time limit, achieving a certain number of features, or encountering insufficient performance improvements.

2- GeneticSearch is another wrapper method in Weka, inspired by genetic algorithms. It evolves potential feature subsets iteratively using genetic operators such as crossover and mutation. It maintains a population of feature subsets and selects the strongest subsets based on their performance (Goldberg, 1989).

The GeneticSearch feature selection method utilizes a genetic algorithm approach to select features. Here's an overview of how it typically works:

- **Initialization:** GeneticSearch starts by creating a population of feature subsets, with each subset representing a potential solution.

- **Evaluation:** The algorithm evaluates each subset's performance using a specific evaluation metric. This metric could be based on the performance of a classification or regression algorithm.
- **Selection:** GeneticSearch selects subsets based on their fitness, which is determined by their evaluation scores. Subsets with better performance have a higher chance of being selected for reproduction.
- **Crossover:** Selected subsets undergo crossover, where features from two parent subsets are combined to create new offspring subsets. This process introduces diversity into the population and explores different feature combinations.
- **Mutation:** Occasionally, subsets in the population undergo mutation, where random changes are made to their features. Mutation helps explore new regions of the search space and prevents premature convergence.
- **Replacement:** Offspring subsets replace less fit subsets in the population, keeping the population size constant.
- **Termination:** The algorithm continues iterating through selection, crossover, mutation, and replacement steps until a termination criterion is met. This criterion could be a maximum number of iterations, reaching a desired level of performance, or a specified time limit.

Overall, GeneticSearch in Weka utilizes genetic algorithms to gradually improve the performance of feature subsets over multiple generations. This iterative approach allows for effective feature selection by evolving towards better solutions.

3- GreedyStepwise is a wrapper method that starts with an empty feature subset and gradually adds or removes features based on their individual performance. It explores the feature subset space in a stepwise manner, greedily selecting features that optimize a specific evaluation metric (Kohavi and John, 1997).

The Weka GreedyStepwise feature selection method performs feature selection by gradually adding or removing features based on their individual performance. The process generally follows these steps (Kohavi and John, 1997):

- Initialization: GreedyStepwise starts with an initial feature subset, which may be empty or contain a subset of features already.
- Forward Selection: The algorithm then evaluates the performance of adding each feature that is not already in the subset. It selects the feature that improves performance the most and adds it to the subset. This step continues until adding additional features no longer improves performance.
- Backward Elimination: After forward selection, the algorithm evaluates the performance of removing each feature from the subset. It removes the feature that results in the least decrease in performance when removed. This step continues until removing additional features no longer improves performance.
- Stopping Criteria: The process stops when a predefined stopping criterion is met, such as reaching a maximum number of features, no further improvement in performance, or exceeding a specified time limit.
- Final Subset Selection: Finally, the algorithm selects the feature subset that yields the best performance according to a specified evaluation metric.

Techniques that combine the advantages of filters and wrappers have been introduced as hybrid approaches. In the first step, a filter approach is used to reduce the feature space dimension, which can result in multiple subsets that are suitable. These subsets can then be further analyzed using a wrapper approach to identify the best candidate subset. By using hybrid techniques, it is possible to achieve both high wrapper accuracy and great filter efficiency (Hall and Holmes, 2003).

Prior to estimating the scores for the 125 countries in the GII 2012-2022 period, it was necessary to identify the parameters that had the greatest impact on the GII. To accomplish this, the Weka program version 3.8.5 was utilized for feature selection. Weka is among the many programs designed to derive meaningful insights from patterns within data files, and its popularity can be attributed to its open-source software and user-friendly interface (Zupan and Demsar, 2008). Feature selection is a crucial process in ML that involves eliminating redundant and irrelevant features to improve the efficiency and accuracy of the learning algorithm. It reduces the feature space, which not only improves performance but also reduces computational complexity and the risk of overfitting.

The performance of classification techniques is improved by the process of feature selection. Here is a breakdown of the procedure involved:

- Employing search algorithms to generate subsets of potential features from the original set.
- Evaluating the relevance of each candidate subset to the classification task using metrics such as distance, dependence, information, consistency, and classifier error rate.
- The optimal feature subset or relevant subset is identified based on the termination condition.
- Validation is conducted to confirm that the selected feature subset is accurate (Hall and Smith, 1999).

WEKA's attribute selection methods consist of two components: a search method including best-first, forward selection, random, exhaustive, genetic algorithm, and ranking; and an evaluation method such as correlation-based, wrapper, information gain, or chi-squared. The attribute selection mechanism in WEKA is highly flexible, allowing for nearly any combination of these two methods. In this study, feature selection was performed in a hybrid way by combining these two methods. Attribute evaluators and search methods appropriate to the data set were performed to search for all possible attribute combinations in the data and find out which subset of attributes gave the best results for prediction. The evaluator determines which method to use to assign a value to each subset of attributes. The search method determines which search style will be performed (Aksenova, 2004). Attribute Evaluators suitable for the data set using in the study are CfsSubsetEval, ClassifierSubsetEval and WrapperSubsetEval, while search methods are best first, genetic search and greedy stepwise.

In addition to selecting the appropriate attribute evaluator and search method, the attribute selection mode must be carefully chosen. It is imperative to consider the available options for feature selection mode, which include:

- Use the full training set: The value of the attribute subset is determined by the full set of training data.
- Cross-validation: The value of a subset of attributes is determined by a process of cross-validation. The 'Fold' and 'Seed' fields set the number of folds to use and the random seed used when shuffling the data (Aksenova, 2004).

The practical importance of the cross-validation method in model selection application is emphasized. Cross-validation techniques contribute to the reliability of model evaluation (James, Witten, Hastie, and Tibshirani, 2013). The utilization of cross-validation as a mechanism to assess the generalization performance of predictive models has been subject to extensive investigation. It is well-established that cross-validation presents an effective solution to issues such as overfitting, and offers a more realistic estimate of a model's efficacy when presented with unseen data (Hastie et al., 2009). The present study advocates for the use of the full training set over cross-validation for feature selection. Cross validation is widely considered more effective for evaluating a model's ability to generalize and minimizing bias in performance evaluation. However, cross-validation offers benefits particularly when working with small dataset sizes. In this study, it was utilized full training set because of the large size of the dataset. In line with the operational principles of the aforementioned feature selection, the approach employed for feature selection in this study can be outlined as followed Figure 3.3.

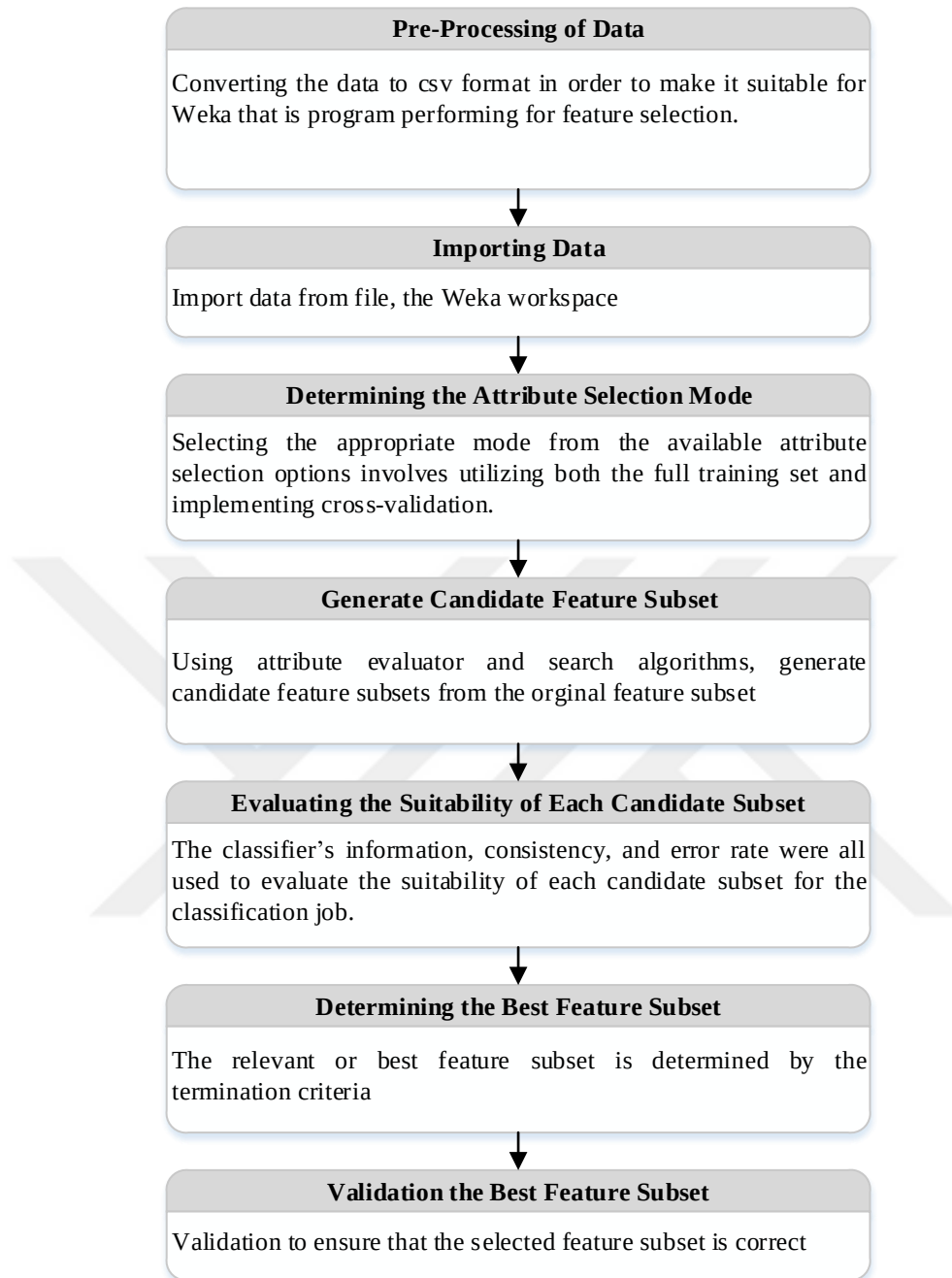


Figure 3.3. Feature selection methodology applied in this study

This study primarily concentrates on ANN and Logistic Regression, both of which are supervised ML techniques. Consequently, comprehensive details about these methods are presented in the subsequent subsections.

3.2.3. Artificial Neural Network Method

ANNs are computer systems that are modeled after the structure and function of the human brain's neurons. ANNs comprise a series of nodes or neurons arranged in layers, exhibiting

global behaviors determined by the connections between processing elements and relevant parameters. The connections between neurons in successive layers have weights that determine the strength of their connection across different layers. The neural network architecture comprises an input layer, one or more hidden layers, and an output layer. The selection of the optimal number of neurons and hidden layers is contingent upon the complexity of the underlying system. The general structure of an ANN with two hidden layers is presented in the Figure 3.4 below (Green et al., 2006).

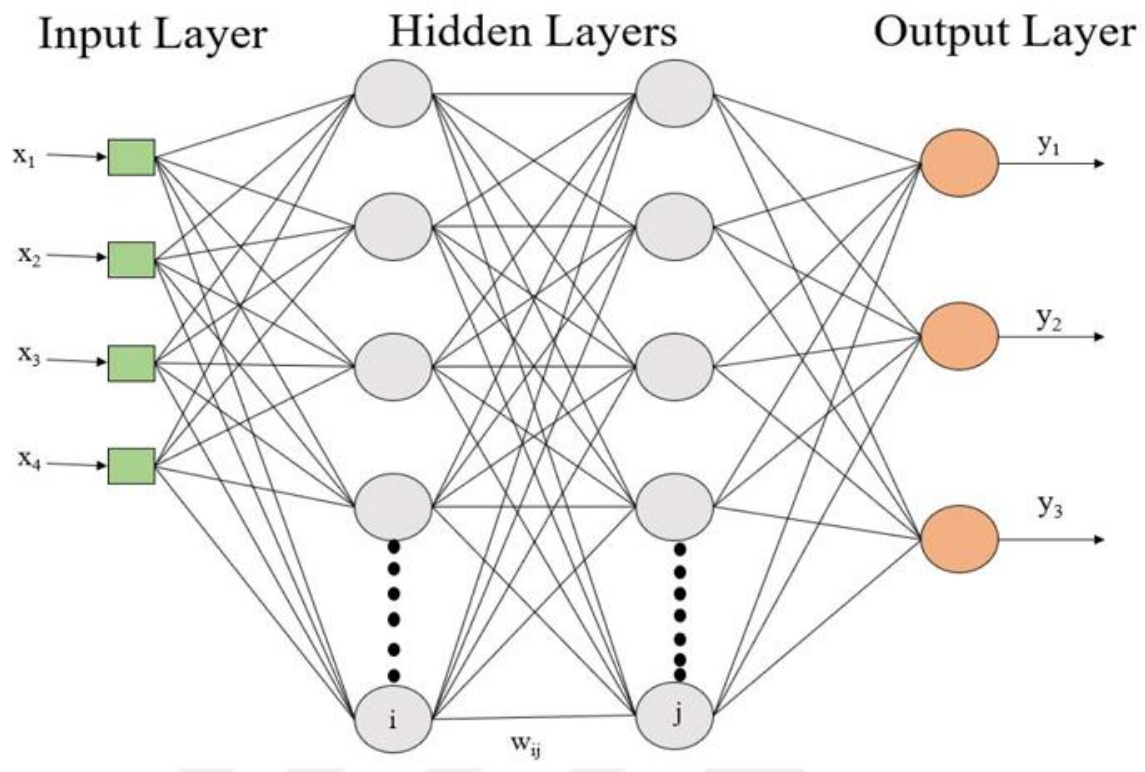


Figure 3.4. General structure of an ANN with two hidden layers

The input layer neurons of an ANN receive data, which is subsequently transmitted to the neurons in the first hidden layer via weighted connections. Through mathematical processing at each layer, the data is then forwarded to the succeeding layer. The subsequent steps delineate the processing of input data (x_i) by the j th node in the subsequent layer.

The explanation below outlines the basic functioning of a neuron:

- First, the weighted sum is computed, followed by the addition of a bias term—a constant value incorporated into a neuron's input before it undergoes activation function processing.

- The net input undergoes transformation using a mathematical 'transfer function.' Activation functions play a pivotal role in ANNs as they dictate how a neuron responds to a given input.
- Various activation functions, such as the sigmoid function, rectified linear unit function, hyperbolic tangent (tanh) function, and leaky rectified linear unit function, exist. These functions are commonly utilized in neural networks, each with distinct characteristics and purposes.
- Ultimately, the resulting output is transmitted to the neuron in the subsequent layer.

The output layer signifies the conclusive outcome generated by the network, usually presented as a predicted value or class label. It consolidates the information processed by preceding layers in the network, factoring in the weights of the connections between neurons, and then formulates a prediction based on this assimilated information.

Once a neural network has been created for an application, the next step involves training it using randomly assigned initial weights. These models gain popularity due to their adaptive characteristics, learning by identifying patterns within the input data. The training process can be approached in two manners: supervised training and unsupervised training. In supervised training, the desired output must be supplied along with the inputs. This helps optimize the network weights to discover the most effective set of weights that minimize errors in the output. Various neural network architectures exist, including feedforward neural networks, convolutional neural networks, recurrent neural networks, and long short-term memory networks. For this study, feedforward neural networks were employed. These networks represent the most straightforward neural network type, comprising an input layer, one or more hidden layers, and an output layer. Information progresses unidirectional from the input layer to the output layer, and the inter-neuron connections do not form cycles.

The explanation of the Levenberg-Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient (SCG) algorithms used for prediction in MATLAB's nftool are as follows:

- The Levenberg-Marquardt algorithm: One popular optimization technique for resolving non-linear least squares issues is the Levenberg-Marquardt algorithm. Because it can effectively decrease the error between the expected and real outputs, it is very useful for

training neural networks. The Levenberg-Marquardt algorithm combines the Gauss-Newton optimization techniques with the steepest descent. Operation: The weights are initialized at the beginning. The difference between the expected and actual outputs is then calculated. It updates the weights by combining the Gauss-Newton and steepest descent techniques. Until the convergence requirements are satisfied or the maximum number of iterations is achieved, this procedure is repeated (Levenberg, 1944). It can be represented by the following formula:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - (\mathbf{J}_k^T \mathbf{J}_k + \lambda \mathbf{I})^{-1} \mathbf{J}_k^T \mathbf{e}_k \quad (3.10)$$

Where:

- $\mathbf{w}_k$ represents the weight vector,
 - $\mathbf{J}_k$ represents the Jacobian matrix,
 - λ represents the Levenberg-Marquardt parameter,
 - $\mathbf{I}$ represents the identity matrix,
 - $\mathbf{e}_k$ represents the error, and
 - k represents the iteration number.
- The Bayesian regularization algorithm: To regularize neural network training, the Bayesian regularization technique employs Bayesian principles. To successfully balance the complexity of the model with data fitting, it adds a prior distribution across the network weights and changes this distribution during training depending on the observed data. Normalization the Bayesian regularization technique makes use of Bayesian concepts to regularize the training of neural networks. The way it works is as follows: It provides a previous distribution across the weights in the network. Using the previous and observed data, it initializes the posterior distribution. Using Bayesian inference techniques like variational inference and Markov Chain Monte Carlo (MCMC) sampling, it updates the posterior distribution. Until convergence or a certain number of iterations is attained, this procedure is repeated (Neal, 2012). The following formula may be used to express Bayesian regularization:

$$P(w|D) = \frac{P(D|w)P(w)}{P(D)} \quad (3.11)$$

Where:

- $P(w|D)$ represents the weight distribution given the data,
 - $P(D|w)$ represents the likelihood of the data under the weights,
 - $P(w)$ represents the prior distribution of the weights, and
 - $P(D)$ represents the probability of the data.
- The SCG algorithm: Neural network training algorithms employ the SCG algorithm as an optimization method. In comparison to conventional gradient descent methods, it achieves faster convergence by fusing the simplicity of scaled gradient methods with the efficiency of conjugate gradient methods. An optimization method based on conjugate gradient techniques is called Scaled Conjugate Gradient, or SCG. The way it works is as follows: The weights are initialized initially. It calculates the error function's gradient in relation to the weights. Using the scaled gradient data, it calculates the conjugate gradient direction. Along the conjugate gradient direction, it calculates the ideal step size. The conjugate gradient direction and ideal step size are used to update the weights. Until the maximum number of iterations is achieved or the convergence criteria are satisfied, this process is repeated (Møller, 1993). The SCG algorithm can be represented by the following formula:

$$w_{k+1} = w_k + a_k p_k \quad (3.12)$$

Where:

- w_{k+1} represents the updated weights after the k -th iteration in the optimization process
- a_k represents the step size, and
- p_k represents the conjugate gradient direction.

In line with the operational principles of the aforementioned ANN algorithm, the approach employed for forecasting in this study can be outlined as followed Figure 3.5.

In the initial step illustrated in the above figure, the data underwent normalization using the min-max technique. The min-max normalization is a widely employed data transformation method designed to maintain the sensitivity of attributes in an individual-focused dataset (Saranya and Manikandan, 2013). This technique involves normalizing the original dataset values through a min-max normalization function, considering the minimum and maximum values within the dataset. The min-max normalization technique performs a linear transformation on the original dataset, proving particularly beneficial for classification purposes. It has found applications in various domains, including AI, clustering, nearest-neighbor classification, neural networks, and more (Mendes and Vilela, 2017). Min-max normalization performs a linear transformation on the original data. This technique gets all the scaled data in the range (0, 1). The formulation of the min-max normalization technique is as follows:

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3.13)$$

x' indicates the normalized data, x_i indicates the input value, x_{min} indicates the smallest number in the input set, x_{max} indicates the largest number in the input set.

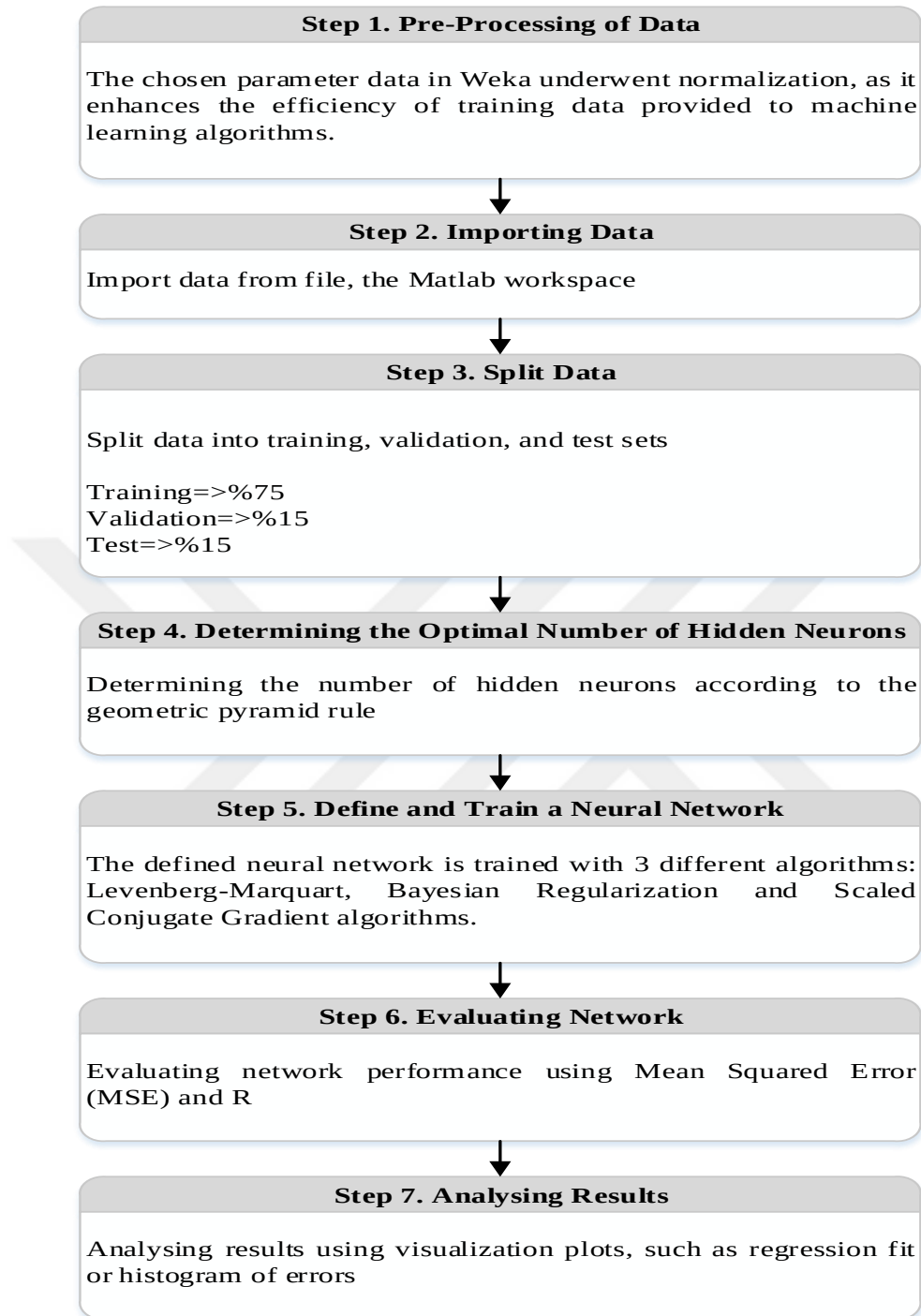


Figure 3.5. ANN methodology applied in the study

As part of the second step in the prediction-oriented ANN methodology, the MATLAB program R2022b version was employed. During this stage, the data was imported into the MATLAB workspace and then selected neural net fitting tool.

As part of the third step in the field of ANN methodology, it is a widely adopted practice to allocate data into 75% for training, 15% for validation, and 15% for testing, as observed in the literature. This segmentation is designed to facilitate the model's learning process on the training set, optimize its performance on the validation set, and assess its overall effectiveness on the testing set. However, it is crucial to acknowledge that these ratios may vary across studies due to differences in datasets and problem characteristics. In this study, the data partition aligns with our dataset, with a distribution of 75% for training, 15% for validation, and 15% for testing.

In the fourth step, the presence of hidden layer(s) in the network contributes to its ability to generalize. Despite its significance, there is no definitive formula for determining the optimal number of hidden neurons, leading researchers to rely on experimentation. Nevertheless, certain heuristic guidelines have been suggested, and one such approximation is offered by the geometric pyramid rule proposed by Masters. According to the geometric pyramid rule, the number of neurons should decrease from the input layer to the output layer. The count of hidden neurons should not surpass twice the number of input neurons. Simultaneously, this count should not fall below the square root of the product of the number of input and output neurons. In this investigation, the number of hidden neurons was determined based on adherence to this rule.

In the fifth stage, the network underwent training using three distinct algorithms provided by Matlab for neural networks: Levenberg-Marquardt, Bayesian Regularization, and SCG. Furthermore, the hidden neuron quantities, as dictated by the geometric pyramid rule, were employed in all three algorithms.

In step 6 of this study, conventional statistical metrics—specifically, Mean Square Error (MSE) and Correlation Coefficient (R)—were utilized to assess and validate the performance of ML models. Essentially, MSE represents the mean squared difference between estimated and actual values. Lower MSE values indicate a higher predictive capability of the models. Additionally, R is employed to assess the correlation between predicted and actual values of soil shear strength. R values range between -1 and +1, where closer absolute values to 1 indicate superior predictive ability. The computation of these metrics can be carried out using the following formulas (Nguyen et al., 2019):

$$R = \sqrt{1 - \frac{\sum((y_i - x_i))^2}{\sum((y_i - y_{ort}))^2}} \quad (3.14)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \quad (3.15)$$

Here, i represents the number of data, y represents the actual value, and x represents the predicted value.

Ultimately, the assessment results were analyzed through the utilization of visualization plots.

3.2.4. Regression Methods

Regression analysis is a statistical method for modeling the connection between a dependent variable and one or more independent variables. It is utilized to comprehend the nature of this relationship and to forecast the value of the dependent variable using the values of the independent variables. The dependent variable is typically referred to as the outcome variable because its value is influenced by other variables. The independent variables are predictor variables because their values are used to predict the value of the dependent variable. The regression model can be linear or nonlinear. Linear regression attempts to model the dependent variable's linear relationship with the independent variables, while nonlinear regression can handle this relationship in a more complex manner (Montgomery et al., 2021).

Types of regression models are Linear Regression and Nonlinear Regression. Linear Regression predicts the dependent variable through a linear function using the first-degree terms of the independent variables. It is employed to assess each independent variable's influence on the dependent variable as well as the degree to which the dependent variable can be explained by the independent variables. On the other hand, Nonlinear Regression uses a more complex function to predict the dependent variable since the relationship cannot be expressed through a linear function. It is employed in situations where slopes or effects are not linear. While linear regression and its types fall under linear regression, regression trees, SVMs, kernel approximation models, ensembles of regression trees, and neural network regression, along with their types, belong to non-linear regression models.

The Regression Learner App provides a user-friendly environment compatible with MATLAB R2022b and subsequent versions and also is designed to train regression models for predicting data. Within this tool, users have the flexibility to choose from various regression methods to automatically train and validate regression models. Following the training of multiple models, a comparative analysis can be conducted to identify the most suitable one. The application streamlines processes like data importation, feature selection, model training, and result evaluation. It incorporates widely used regression techniques such as linear regression models, regression trees, SVMs, kernel approximation models, ensembles of regression trees, and neural network regression models as followed Table 3.3.

Table 3.3. Regression Models

Regression Models	Types of Regression Models
Linear Regressions	Linear Regression Model
	Interactions Regression Model
	Robust Regression Model
	Stepwise Linear Regression Model
Regression Trees	Fine Tree
	Medium Tree
	Coarse Tree
SVMs	Linear SVM
	Quadratic SVM
	Cubic SVM
	Fine Gaussian SVM
	Medium Gaussian SVM
	Coarse Gaussian SVM
Kernel Approximation Models	SVM Kernel
	Least Squared Regression Kernel
Ensembles of Regression Trees	Boosted Trees
	Bagged Trees
Neural Network Regression	Narrow Neural Network
	Medium Neural Network
	Wide Neural Network
	Bilayered Neural Network
	Trilayered Neural Network

3.2.4.1. Linear Regressions

Linear regression analysis is a statistical technique that use a linear function to forecast the dependent variable. The relationship between the dependent and independent variables is stated

using first-degree terms. This method aids in determining how much of the dependent variable is explained by the independent factors while also evaluating the influence of each independent variable on the dependent variable. The simple linear regression model is expressed as follows (Johnson and Bhattacharyya, 2019):

$$y = \beta_0 + \beta_1 x + \epsilon \quad (3.16)$$

Here:

- y represents the dependent variable,
- x represents the independent variable,
- β_0 and β_1 are the intercept and slope coefficients respectively, and
- ϵ represents the error term.

Types of linear regression models are Interactions Regression Model, Robust Regression Model, and Stepwise Linear Regression Model. These types are explained below:

- **Interactions Regression Model:** The interactions regression model aims to capture the interaction effects between independent variables in predicting the dependent variable. This model extends the basic linear regression model by including interaction terms, which are the products of two or more independent variables. For example, if we have two independent variables x_1 and x_2 , an interaction term would be $x_1 \times x_2$. By including these interaction terms in the regression model, we can account for instances in which the effect of one independent variable on the dependent variable is determined by the value of another independent variable. This model is extensively used in various fields such as social sciences, economics, and psychology (Sparkman Jr, 1979).

The mathematical formula for the Interactions Regression Model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon \quad (3.17)$$

Where:

- y is the dependent variable (outcome variable),
- x_1 and x_2 are the independent variables (predictor variables),
- $\beta_0, \beta_1, \beta_2$, and β_{12} are the model parameters (coefficients),

- x_1x_2 is the interaction term between x_1 and x_2 ,
- ϵ is the error term.

In this formula, the term $\beta_{12}x_1x_2$ represents the interaction between x_1 and x_2 . This term is added to the model to capture the effect resulting from the interaction between x_1 and x_2 . It allows the model to account for the effect arising from the interaction of the independent variables in addition to their individual effects.

- **Robust Regression Model:** The robust regression model is designed to mitigate the influence of outliers or influential data points on the estimation of model parameters. Unlike ordinary least squares regression, which is sensitive to outliers, robust regression techniques use alternative estimation methods that are less affected by extreme observations. These methods include M-estimation, iteratively reweighted least squares, and Huber regression. Robust regression models provide more reliable parameter estimates when the assumption of normally distributed errors is violated or when the data contain influential observations (Weisberg, 2005). The formula for the robust regression model closely resembles that of simple linear regression, as both models aim to predict the dependent variable (Y) based on one or more independent variables ($x_1, x_2, \dots, x_k$). However, the difference lies in the method used to estimate the coefficients ($\beta_0, \beta_1, \beta_2, \dots, \beta_k$). In simple linear regression, coefficients are estimated with the ordinary least squares method, which minimizes the sum of squared residuals. Nevertheless, this method can be sensitive to outliers or influential observations, meaning that extreme data points can disproportionately influence the estimated coefficients. Robust regression model, on the other hand, uses alternative estimation techniques to mitigate the impact of outliers or influential observations during the coefficient estimation process. These robust estimation methods, such as M-estimation, iteratively reweighted least squares, or Huber regression, adjust the estimation process to down weight the influence of extreme observations, leading to more reliable parameter estimates in the presence of outliers. Therefore, while simple linear regression and robust regression share the same basic formula structure, the latter incorporates robust estimation techniques to enhance the reliability of the coefficient estimates, especially when outliers or influential data points are present.

- **Stepwise Linear Regression Model:** Stepwise linear regression is a variable selection technique used to identify the most relevant independent variables for predicting the dependent variable. This method iteratively adds or removes independent variables from the model based on statistical criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or p-values. The stepwise procedure typically involves two main steps: forward selection and backward elimination. While backward elimination involves removing variables from the model if they do not significantly contribute to the prediction, forward selection involves adding independent variables to the model one at a time. Stepwise linear regression is useful for identifying the subset of predictors that best explain the variation in the dependent variable (Montgomery et al., 2021).

The Stepwise Linear Regression Model's mathematical formula isn't as straightforward as simple linear regression or robust regression's formulas. This is because stepwise regression entails a variable selection process rather than a single formula. Nevertheless, it outlines the general procedure of stepwise linear regression:

Step 1: Begin with an empty model or a model that includes all potential predictor variables.

$$y = \beta_0 + \epsilon \quad (3.18)$$

Step 2: Iteratively add or remove predictor variables from the model based on certain criteria (e.g., p-values, Akaike Information Criterion (AIC), (BIC), etc.) using forward selection or backward elimination.

$$y = \beta_0 + \beta_1 x_1 + \epsilon \quad (3.19)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (3.20)$$

⋮
⋮

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_k x_k + \epsilon \quad (3.21)$$

Step 3: Continue adding or removing variables until the model's fit is no longer improved by the addition or removal of additional variables according to the specified criteria.

As a result, the stepwise linear regression model doesn't have a single mathematical formula. Instead, it involves a series of models with different sets of predictor variables based on the stepwise selection process.

3.2.4.2. Regression Trees

Regression trees are a type of decision tree used for regression problems and belong to the category of non-linear regression models. They use a tree structure to direct an observation, given a set of independent variables, to the predicted value of a dependent variable. This structure is created by recursively partitioning the independent variables with a series of decision nodes, where each node makes a specific prediction (Hastie et al., 2009). Regression trees are typically expressed as a tree structure because there is no simple mathematical formulation that can reflect the complexity of variables and relationships in the dataset. When making predictions, regression trees follow a specific process.

- Each observation starts at the root node and progresses through the tree based on specific rules.
- At each decision node, the tree examines a particular feature of the observation and makes a decision on which direction to split based on a threshold value for that feature.
- Based on this decision, the observation is directed to the next node or leaf node. Finally, at a leaf node, the predicted value for the observation is obtained.
- This prediction could typically be the mean value of the dependent variable for observations in that leaf node.

In this manner, each observation obtains a prediction by following the path in the tree. These predictions are derived from the training dataset used to construct the tree, and can be used to make predictions for new data points.

Types of regression trees models are Fine Tree, Medium Tree, and Coarse Tree. These types are explained below:

Regression trees are a popular choice for predictive modeling, with Fine Tree, Medium Tree, and Coarse Tree being the most commonly used variations. Let's take a closer look at each type to see how they differ from one another:

- **Fine Tree:** Fine trees are highly detailed and complex, with many decision nodes. They aim to capture even the most subtle nuances in the data, resulting in detailed predictions. However, this level of granularity can sometimes lead to overfitting or modeling noise if not properly regulated.
- **Medium Tree:** Medium trees strike a balance between detail and interpretability, with a moderate level of complexity. They aim to capture the main trends in the data without sacrificing too much simplicity. This makes them a great choice for those looking for a compromise between fine and coarse granularity.
- **Coarse Tree:** Coarse trees have a simplified and generalized structure, with fewer decision nodes than the other two types. They focus on capturing broad patterns in the data, sacrificing detail for simplicity and interpretability. This makes them a great choice for those looking for a high-level understanding of the data.

In conclusion, each type of regression tree offers a unique trade-off between detail and interpretability, catering to different modeling needs and preferences.

3.2.4.3. Support Vector Machine

SVM is a type of regression model, specifically falling under the category of non-linear regression models. SVM is commonly used for both classification and regression tasks (Hastie et al., 2009).

The basic mathematical formulation of SVM for regression, also known as Support Vector Regression (SVR), is as follows:

Given a training dataset consisting of input-output pairs (x_i, y_i) , where x_i represents the input features and y_i represents the corresponding output, SVR aims to find a function $f(x)$ that predicts the output y for any given input x . The function $f(x)$ is represented as:

$$f(x) = \langle w, x \rangle + b \tag{3.22}$$

where:

- w represents the weight vector,
- x represents the input feature vector,
- b represents the bias term.
-

Mathematically, if $w = (w_1, w_2, \dots, w_n)$ and $x = (x_1, x_2, \dots, x_n)$ are two vectors of the same dimension n , then the dot product $\langle w, x \rangle$ is computed as:

$$\langle w, x \rangle = w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (3.23)$$

The goal of SVR is to find the optimal values of w and b that minimize the error between the predicted output $f(x)$ and the true output y_i for all training samples, while simultaneously maximizing the margin between the predicted outputs and a margin of tolerance ϵ .

SVR also incorporates the concept of kernels to handle non-linear relationships between input features and outputs. The kernel function $K(x_i, x_j)$ allows SVR to implicitly map the input features into a higher-dimensional space where a linear relationship may exist, even if the relationship is non-linear in the original feature space.

The SVR algorithm is designed to find the best hyperplane that can predict the output for a given input while allowing some margin of tolerance. Here's a brief overview of how the SVR algorithm works:

- **Input Data:** The SVR algorithm requires a training dataset that consists of input-output pairs (x_i, y_i) , where x_i represents the input features, and y_i represents the corresponding output.
- **Kernel Function:** To map the input features into a higher-dimensional space where a linear relationship may exist, SVR often uses a kernel function. Common kernel functions include linear, polynomial, radial basis function, and sigmoid kernels.
- **Model Training:** SVR aims to find a function $f(x)$ that can predict the output y for any given input x . This function is represented as: $f(x) = \langle w, x \rangle + b$. To achieve this, the weight vector w and the bias term b are optimized during the training process to minimize the error between the predicted output $f(x)$ and the true output y_i for all training samples, while also maximizing the

margin between the predicted outputs and a margin of tolerance ϵ . This optimization problem is usually formulated as a constrained quadratic programming problem, where the objective function minimizes the error subject to certain constraints, including the margin of tolerance ϵ .

- Prediction: Once the SVR model is trained, it can be used to make predictions for new input data points by calculating the dot product between the weight vector w and the input feature vector x and adding the bias term b . The predicted output $f(x)$ represents the regression estimate for the given input x .
- Evaluation: SVR models are evaluated using various performance metrics, such as MSE, mean absolute error (MAE), or coefficient of determination (R-squared), to assess the accuracy and generalization ability of the model.

Overall, SVR is a powerful regression technique that can capture complex relationships between input features and output values, especially in high-dimensional and non-linear datasets.

Types of SVM models are Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, and Coarse Gaussian SVM. These types are explained below:

SVM models are commonly used for regression tasks. There are six different variants of SVM models, each with its own unique features and strengths. These include Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, and Coarse Gaussian SVM (Cristianini and Shawe-Taylor, 2000).

A brief overview of each type is as follows:

- Linear SVM: This type uses a linear kernel function to separate input space with a linear decision boundary. It's best suited for datasets with linearly separable classes or when the relationship between input features and output is approximately linear.
- Quadratic SVM: Quadratic SVM captures non-linear relationships between input features and output by using a quadratic kernel function. This technique enables more complex decision boundaries compared to linear SVM. It's beneficial when the relationship between input features and output exhibits quadratic patterns or non-linearities.
- Cubic SVM: This variant utilizes a cubic kernel function to capture even more complex non-linear relationships between input features and output. It can model

higher-order interactions between features and is appropriate for datasets with highly non-linear relationships that cannot be adequately captured by linear or quadratic models.

- **Fine Gaussian SVM, Medium Gaussian SVM, and Coarse Gaussian SVM:** These SVM variants use Gaussian (or Radial Basis Function) kernel functions with different bandwidths. Fine Gaussian SVM has a smaller bandwidth, offering a more detailed and flexible model. Medium Gaussian SVM has a moderate bandwidth, providing a balance between detail and generalization. Coarse Gaussian SVM has a larger bandwidth, resulting in a smoother and more generalized model. Gaussian SVMs are effective for capturing complex non-linear relationships in the data.

The main distinction between these SVM models lies in the type of kernel function they use and their ability to capture non-linear relationships. Linear, Quadratic, and Cubic SVMs are explicitly designed to capture specific types of non-linear relationships using polynomial kernel functions, while Gaussian SVMs offer a more flexible and generalizable approach with Gaussian kernel functions. Compared to other SVM models, Gaussian SVMs typically provide better performance in capturing complex non-linear patterns but may require more computational resources and tuning.

3.2.4.4. Kernel Approximation Models

Kernel Approximation Models are commonly employed to decrease the computational complexity of kernel methods or to enhance their effectiveness on large datasets. These models approximate the kernel function by transforming the data into a lower-dimensional feature space instead of operating directly on the original high-dimensional feature space. This strategy reduces the computational cost while still providing satisfactory performance (Rahimi and Recht, 2007).

The algorithm for Kernel Approximation Models typically involves the following steps:

- **Data Transformation:** To transform a dataset with high-dimensional input features, the first step is to use a kernel function $k(x, x')$ to convert the data into a lower-dimensional feature space. This kernel function defines the similarity between data points x and x' in the original feature space.

- **Approximation of Kernel Function:** To avoid directly computing the kernel function for each pair of data points in a high-dimensional feature space, an approximation technique is used to estimate the kernel function in a lower-dimensional feature space. This approximation can be achieved through various methods such as Random Fourier Features, Nystrom Approximation, or other techniques.
- **Training Model:** Once the kernel function is approximated, a learning algorithm such as SVM or Kernel Ridge Regression (KRR) is trained using the transformed data in the lower-dimensional feature space. This step involves optimizing the model parameters to minimize the prediction error on the training data.
- **Prediction:** After training the model, it can be used to make predictions for new data points. The input features of the new data are first transformed into the lower-dimensional feature space using the same kernel approximation technique. Then, the trained model is applied to these transformed features to predict the corresponding output.
- **Evaluation:** The Kernel Approximation Model's performance is evaluated in this step using metrics such as MSE or MAE. The purpose is to assess how well the model performs when presented with new data and to gain insights into its predictive accuracy.

Overall, the algorithm for Kernel Approximation Models involves transforming the data into a lower-dimensional space, approximating the kernel function in this space, training a learning algorithm using the transformed data, making predictions, and evaluating the model's performance.

Types of kernel approximation models, such as SVM kernels and least squares regression kernels, are used to approximate kernel functions to reduce the computational complexity of kernel methods or improve their scalability to large data sets.

Below is an explanation of each type along with their differences from other kernel approximation models:

- **SVM kernel:**

- SVM kernel approximation models aim to approximate functions kernel used in SVM algorithms.

- These models often use techniques such as random Fourier features or Nystrom approximation to estimate the kernel function.
- The approximate kernel function is then used during SVM training and prediction, allowing faster computation and better scalability for large data sets.
- Kernel SVM approximate models are especially useful for SVM-based classification or regression tasks.

• **Least squares regression kernel:**

- Least squares regression kernel models are used to estimate the kernel function in the KRR algorithm.
- Similar to SVM kernel approximation models, these models use techniques such as random Fourier features or Nystrom approximation to approximate the kernel function.
- The approximate kernel function is then used in Kernel Ridge Regression for regression tasks, which makes computation faster and improves scalability.
- Kernel least squares regression models are especially useful for regression problems in which kernel methods are applied.

Differences from other kernel approximation models:

- SVM and least squares regression kernel models are specific types of kernel approximation models suitable for SVM algorithms respectively and KRR.
- While other kernel approximation models may target different learning algorithms or have different goals, SVM and least squares kernel regression kernel models focus on Kernel function approximation for SVM and KRR respectively.
- These models have similarities with other kernel approximation techniques in terms of kernel function approximation, but their application areas and target algorithms set them apart.

3.2.4.5. Ensembles of Regression Trees

Regression tree ensembles are a type of non-linear regression model that are frequently useful for complicated and huge datasets. The prediction aggregation from several decision trees is the foundation of these models. A regression model called an ensemble of regression trees is created by combining many decision trees. Using the input characteristics, each decision tree generates a forecast. The final prediction is calculated by averaging, or weighting, the

predictions made by each tree. Mathematical formulation for Ensembles of Regression Trees is as follows (Breiman, 2001):

Let M be the total number of decision trees in the ensemble, and $f_i(x)$ be the prediction of the i _th decision tree for input features x . Then, the total prediction $\hat{y}$ can be expressed as:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^M f_i(x) \quad (3.24)$$

Here:

- $\hat{y}$ represents the total prediction.
- M is the total number of decision trees.
- $f_i(x)$ represents the prediction of the i _th decision tree for input features x .

According to this formulation, the ensemble's overall prediction is calculated by averaging the forecasts from each decision tree. The ultimate forecast is balanced by the ensemble prediction, which is the product of the equal contributions from each decision tree.

The algorithm for Ensembles of Regression Trees Models typically involves the following steps:

- Bootstrap sampling: For each tree, a bootstrap sample was taken from the original data set. This involves random sampling by replacement from the data set, allowing the same data point to be selected multiple times.
- Building a decision tree: The decision tree is trained on each sample. However, during this process, a different subset of features will be selected for each decision tree. This ensures that each tree is independent of the others and improves its generalization ability.
- Prediction: Each decision tree predicts the test example. These predictions are generally continuous values obtained from the regression tree.
- Combined predictions: Predictions from all decision trees combined. This is typically done by averaging each tree's predictions or combining them in a weighted manner. The final prediction is the mean or weighted average of these combined predictions.

The algorithm for ensembles of regression trees produces a model that can make stronger and more general predictions by aggregating these trees. This allows each decision tree to compensate for its errors, resulting in an overall prediction that balances the weaknesses of each tree, thereby creating a stronger overall prediction.

Types of Ensembles of Regression Trees models are Boosted Trees and Bagged Trees. These types are explained below:

- **Booster Tree:**

Booster Tree, also known as Gradient Booster Machine (GBM), is an ensemble learning technique that combines multiple weak learners (usually decision trees) to create strong learners. In Boosted Trees, each tree is built sequentially, with each new tree attempting to correct the errors of previous trees. Predictions are made by aggregating predictions from all trees, typically using weighted sum or gradient descent methods.

Boosted trees differ from kernel approximation models in the following ways:

- Model type: Boosted trees are based on a set of decision trees, while kernel approximation models kernel based on approximate kernel functions.
- Learning method: Boosted trees use sequential learning where each tree is trained to correct the errors of previous trees, while kernel approximation models typically involve direct approximation kernel function without sequential training.
- Problem Area: Boosted trees are often used for supervised learning tasks such as regression and classification, while kernel approximation models are often used for classification and regression tasks non-linearity, especially when dealing with high-dimensional data.

- **Bagging Tree:**

Bagging Tree, short for Bootstrap Aggregate Tree, is another ensemble learning technique that builds multiple decision trees using bootstrap samples from the original data set. Each tree is trained independently and predictions are made by average or majority voting across all trees.

Bagging trees help reduce overfitting and improve model stability and accuracy. Bagging trees differ from kernel approximation models in the following ways:

- Model type: bagging trees are based on a set of decision trees, while kernel bagging trees are based on approximate kernel function.

- Training method: Bagging trees involves training multiple decision trees independently, while kernel approximation models involve directly approximating the kernel function.
- Prediction Strategy: Bagging trees make predictions by aggregating predictions from multiple decision trees, while kernel approximation models make predictions based on the resulting transformed feature space through nuclear approximation.

These explanations highlight the differences between boosted tree, bagged tree, and kernel approximation models in terms of their fundamentals, training methods, and prediction strategies.

3.2.4.6. *Neural Network Regression*

Neural network regression is a type of nonlinear regression model that uses ANN to estimate complex relationships between input features and target variables. These models can capture complex patterns in data and handle linear and non-linear relationships. Neural network regression involves training a neural network model to predict continuous target variables based on input features (Bishop, 1995).

Neural networks consist of multiple interconnected layers of neurons, including an input layer, a hidden layer, and an output layer. Each neuron applies a weighted sum of inputs followed by an activation function to produce an output. Mathematical formulation for Neural Network Regression is as follows:

Let X denote the input feature matrix with m samples and n features, and y represent the target variable vector. The neural network model can be represented mathematically as follows:

$$\hat{y} = f(x; \theta) \quad (3.25)$$

Where:

- $\hat{y}$ is the predicted target variable vector.
- $f(\cdot)$ represents the mapping function defined by the neural network.
- θ denotes the parameters of the neural network, including weights and biases.
-

During training, the parameters θ are optimized to minimize a loss function $L(y, \hat{y})$, which measures the discrepancy between the predicted and actual target values. This is typically done using gradient-based optimization algorithms such as stochastic gradient descent (SGD) or its variants.

The algorithm for Neural Network Regression Models typically involves the following steps:

- Working algorithm: Initialization: Initialize the weights and biases of the neural network with random values.
- Forward Propagation: Computes the prediction outputs of a neural network by propagating input features across the layers of the network.
- Loss Calculation: Calculate the loss between the planned and actual target value using the appropriate loss function.
- Backpropagation: Compute the gradient of the loss function with respect to network parameters using backpropagation.
- Update parameters: Update network parameters (weights and biases) using gradient descent or its variations to minimize loss.
- Iteration: Repeat steps 2 to 5 until convergence or for a predefined number of epochs.

Types of Neural Network Regression Models are Narrow Neural Network, Medium Neural Network, Wide Neural Network, Bilayered Neural Network and Trilayered Neural Network. These types are explained below:

- **Narrow neural networks:**
 - Narrow neural networks usually have a smaller number of neurons in the hidden layers.
 - The ability to capture complex patterns in the data may be limited due to its smaller size.
 - This architecture requires less computation and can be suitable for simpler regression tasks with fewer input features.
- **Medium neural network:**
 - A medium neural network has a moderate number of neurons in its hidden layers.
 - It balances model complexity and computational efficiency.
 - This architecture can handle a broader range of regression tasks with moderate complexity compared to narrow networks.
- **Wide Neural Network:**

- Extended Neural Network has a larger number of neurons in its hidden layers.
- It is better able to capture complex relationships in data due to its larger size.
- This architecture is effective for regression tasks with large numbers of input features or complex nonlinear relationships.
- **Bilayered Neural Network :**
 - Bilayered neural network consists of two hidden layers between the input layer and the output layer.
 - This architecture can capture hierarchical features in data, allowing for more complex representations than shallow networks.
 - It can be suitable for medium to high complexity regression tasks.
- **Trilayered Neural Network :**
 - Trilayered neural network has three hidden layers between the input layer and the output layer.
 - It provides the ability to learn even greater complex models and representations than two-layer networks.
 - This architecture excels in regression tasks with high-dimensional input data or highly non-linear relationships.

Differences from other neural network regression models: Each type of neural network regression model differs in its architecture and ability to capture complex patterns in the data: Narrow neural networks vs medium/wide neural networks: Narrow networks have fewer neurons and lower capacity than medium and large networks. Medium and wide networks can capture more complex relationships due to their larger size. Trilayered network provides an even greater ability to learn complex models than the two-layer network. In short, the difference lies in the scale and depth of the neural network architecture, which affects the ability to model complex relationships in data.

The Figure 3.6 below shows the working methodology workflow of regression models applied using the regression learner app.

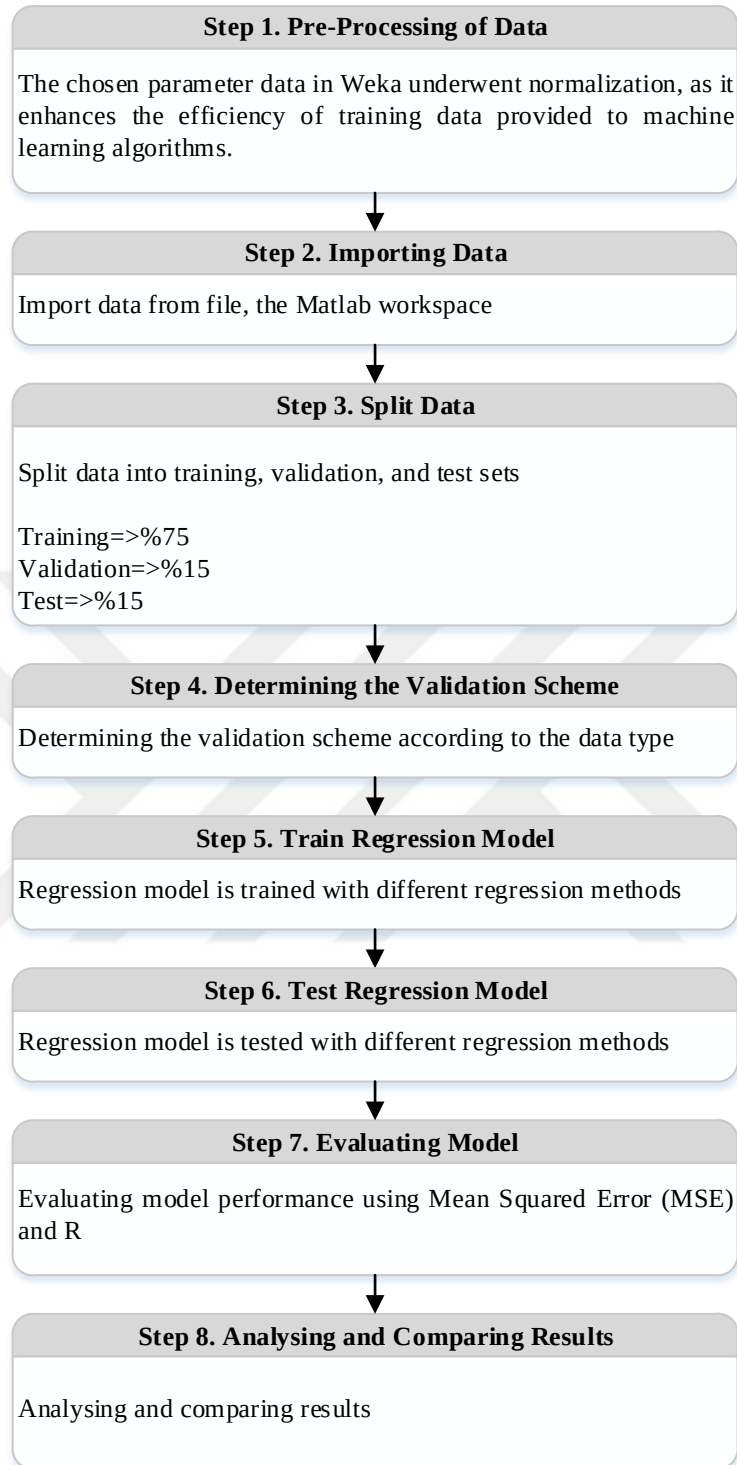


Figure 3.6. LR methodology applied in the study

All regression models mentioned above were used in this study, and the application was made with the help of MALTAB's regression learner app. In the first stage of the regression learner method, normalized data was used in the first step of the ANN regression method. After the data is imported into Matlab, was divided into 75% for training, 15% for validation and 15%

for testing, as in ANN. The regression learner tool presents users with three distinct validation methods: cross-validation, holdout validation, and resubstitution. Cross-validation protects against overfitting. For data not set aside for testing, the app partitions the data into folds and estimates the accuracy on each fold. Holdout validation is recommended for large data sets. The app selects a percentage of the data to use as a validation set. Resubstitution validation provides no protection against overfitting. The app uses the same data for both training and validation. The holdout validation scheme was selected due to the substantial size of the dataset utilized in this study, facilitating a more straightforward comparison between the results obtained from the ANN and regression learner methods. Following this step, the model undergoes initial training and testing using various regression models. The outcomes are then examined and compared based on MSE and R values. As elaborated in ANN, the model demonstrating an MSE value approaching zero and an R value approaching one is considered the most optimal.

3.2.5. Decision Support System Method

In order to obtain precise GII score predictions with the most efficient predictive model derived from the current study, a web-based DSS application was developed using the .net framework and programmed in C# in Visual Studio program. This prototype DSS is composed of three interfaces: the main screen, data entry screen, and result screen, each designed to perform specific functions. The main screen provides users with an overview of the system and its capabilities, while the data entry screen allows users to input relevant data for the prediction process. Once the data is entered, the system utilizes the predictive model to generate a GII score, which is then displayed on the result screen. A flow diagram of the DSS application is also included in Figure 3.7, which illustrates the DSS's functionality in a clear and concise manner.

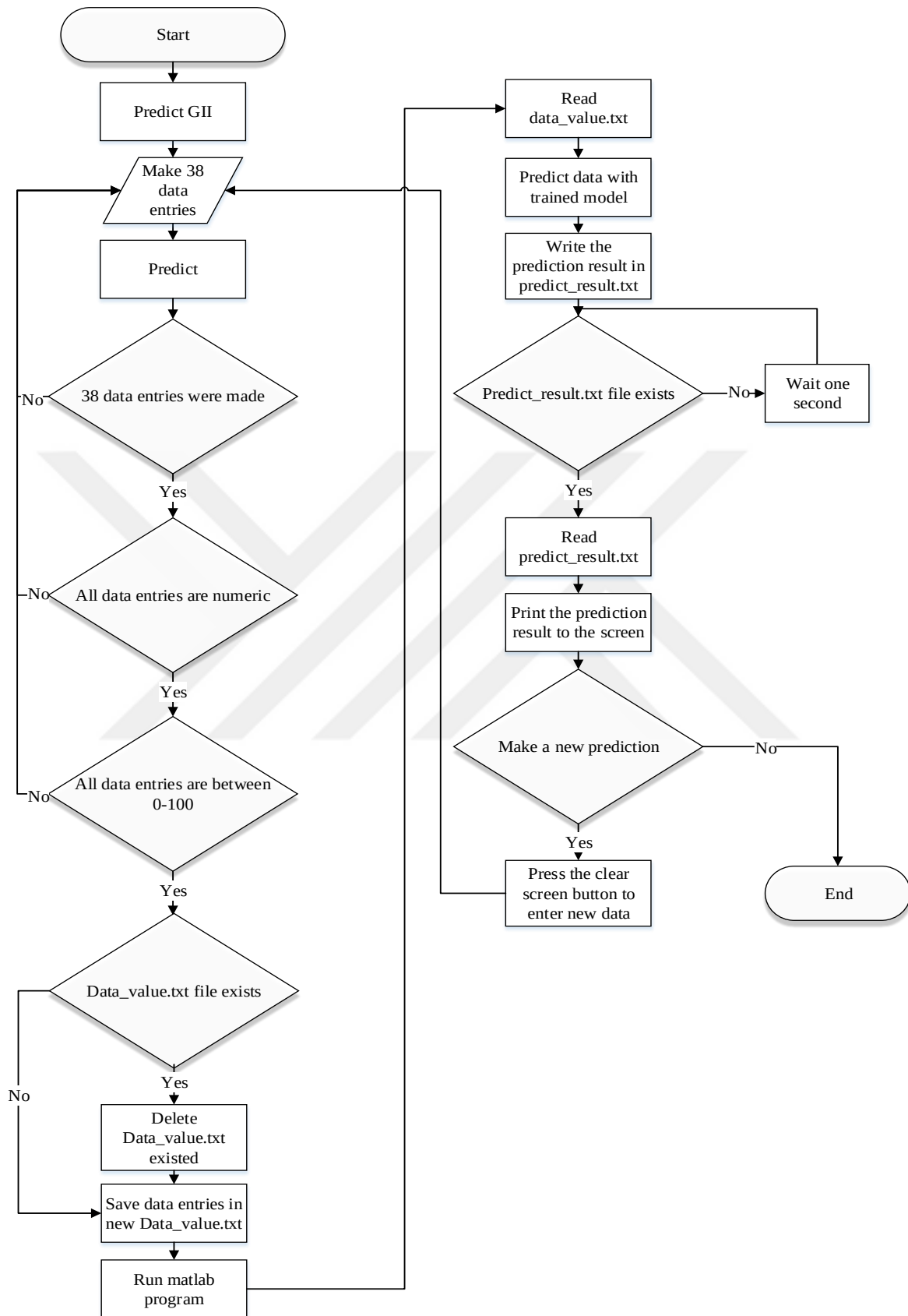


Figure 3.7. A flow diagram of the DSS application

The designed main menu includes a predict GII button, which, when pressed, opens the data entry screen. This screen enables the user to input 38 pieces of data, which are then checked for correctness with a series of queries. The first query checks whether all 38 data entries have been made completely. If positive, the next query checks whether all data entries are numeric. If all data entries are numeric, the program proceeds to check whether all entries are between the number values of 0 and 100. If all these queries are met, the entered data is written to the data_value.txt file. If a previously printed data file exists, it is deleted and a new one is printed.

After this initial process, the Matlab 2023b program runs with the code written in the Visual Studio program. The Matlab program reads data from the data_value.txt file, and predictions are made using a previously trained model. The prediction result is then printed in the predict_result.txt file, and the Matlab program terminates.

The code written in the Visual Studio program checks whether the predict_result.txt file exists or not. If it does not exist, the program waits for one second and then checks again, looping until the file exists. After reading the file, the prediction result is printed on the prediction result screen. Finally, the program asks the user whether they would like to make a new prediction via a button. If the user selects yes, they are directed to the 38 data entry screen with a clear screen button for new data entry. If there is no new data entry, the system terminates. This complex process ensures that the user inputs accurate and correct data and receives accurate predictions.

4. FINDINGS AND DISCUSSION

4.1. Findings

This section includes the classification results, followed by the feature selection results, repeated classification results, ANN method results, Regression method results, and results comparing ANN and Regression method performances.

4.1.1. Classification Findings

Table 4.1 below shows the classification results for a dataset that contains 79 parameter scores of 125 countries. The models' correlation coefficients are listed from highest to lowest. The Random Forest dataset performed the best with a correlation coefficient rate of 98.78%. Furthermore, the error percentages for RAE (14.9602%) and RRSE (15.9146%) were the lowest for Random Forest.

Table 4.1. Classification Results

Classification Techniques	Classifiers/ Algorithms	Correlation Coefficient	RAE	RRSE
Trees	RandomForest	0,9878	14.9602 %	15.9146 %
Meta	RandomCommittee	0,9869	15.0647 %	16.2693 %
Functions	MultilayerPerceptron	0,9839	15.3115 %	18.0537 %
Trees	M5P	0,9839	16.4964 %	17.8528 %
Functions	Linear Regression	0,9838	16.7968 %	17.9128 %
Rules	M5Rules	0,9836	16.6308 %	18.0349 %
Functions	SMOreg	0,9831	16.5421 %	18.4058 %
Lazy	IBk	0,983	15.4831 %	18.4418 %
Meta	RandomSubSpace	0,9829	18.0097 %	18.6865 %
Meta	Bagging	0,982	18.2072 %	18.987 %
Lazy	KStar	0,9791	16.7009 %	20.4291 %
Trees	REPTree	0,9609	25.4608 %	27.7778 %
Trees	RandomTree	0,9579	25.1822 %	29.0344 %
Meta	RegressionByDiscretization	0,9555	27.2808 %	29.719 %
Rules	DecisionTable	0,9362	31.7655 %	35.1463 %
Meta	RandomizableFilteredClassifier	0,9361	27.7777 %	35.8442 %
Meta	Additive Regression	0,9182	38.4723 %	39.8187 %
Functions	Gaussian Processes	0,8818	33.9075 %	47.7603 %
Functions	Simple Linear regression	0,8731	47.391 %	48.7371 %
Lazy	LWL	0,8366	52.5302 %	54.79 %
Trees	DecisionStump	0,765	62.2545 %	64.3874 %

As a result, classification will be repeated to measure the performance of feature selection, which is the next step of the study. Random Forest Classifier, which gives the best classification result, will be used to compare the performance of the feature selection method before and after feature selection.

4.1.2. Feature Selection Findings

The parameters that determine the GII score were selected with the feature selection methods explained in detail in the method section. The feature selection results of 79 parameters that determine the GII score of 125 countries are shown in Table 4.2 below.

Table 4.2. Feature Selection Results

Attribute Evaluator	Search Method	Number of Selected Feature
CfsSubsetEval	Best First	38
ClassifierSubsetEval	Best First	23
WrapperSubsetEval	Best First	23
CfsSubsetEval	Genetic Search	36
ClassifierSubsetEval	Genetic Search	42
WrapperSubsetEval	Genetic Search	50
CfsSubsetEval	Greedy Stepwise	29
ClassifierSubsetEval	Greedy Stepwise	13
WrapperSubsetEval	Greedy Stepwise	16

A total of 38 parameters were selected using a hybrid of CfsSubsetEval and Best First feature selection methods. Additionally, 23 parameters were selected via a hybrid approach using ClassifierSubsetEval and Best First, along with 23 through a combination of WrapperSubsetEval and Best First. Further, 36 parameters were chosen using a hybrid of CfsSubsetEval and Genetic Search, while 42 were selected using a hybrid of ClassifierSubsetEval and Genetic Search, and 50 were chosen using a hybrid of WrapperSubsetEval and Genetic Search. Lastly, a total of 29 parameters were selected using a combination of CfsSubsetEval and Greedy Stepwise, along with 13 using a hybrid approach of ClassifierSubsetEval and Greedy Stepwise, and 16 using a combination of WrapperSubsetEval and Greedy Stepwise.

4.1.3. Repeated Classification Results Based on Feature Selection Findings

The results of repeated classification based on the feature selection results of the GII score data of 79 parameters are shown in Table 4.3 below.

Table 4.3. Repeated Classification Results

Attribute Evaluator	Search Method	Correlation Coefficient (Before Feature Selection)	Correlation Coefficient (After Feature Selection)
CfsSubsetEval	Best First	0,9878	0,9883
ClassifierSubsetEval	Best First	0,9878	0,9882
WrapperSubsetEval	Best First	0,9878	0,9879
CfsSubsetEval	Genetic Search	0,9878	0,9874
ClassifierSubsetEval	Genetic Search	0,9878	0,9874
WrapperSubsetEval	Genetic Search	0,9878	0,9881
CfsSubsetEval	Greedy Stepwise	0,9878	0,987
ClassifierSubsetEval	Greedy Stepwise	0,9878	0,9862
WrapperSubsetEval	Greedy Stepwise	0,9878	0,9876

The classification results obtained as a result of a hybrid of CfsSubsetEval and Best First, a hybrid of ClassifierSubsetEval and Best First, a hybrid of WrapperSubsetEval and Best First and a hybrid of WrapperSubsetEval and Genetic Search indicate an increase in success rates compared to the results obtained before the feature selection. The hybrid of CfsSubsetEval and Best First method achieved the highest success rate with an R value of 0.9883.

On the contrary, there was a decrease in the classification success rate as a result of a hybrid of CfsSubsetEval and Genetic Search, a hybrid of ClassifierSubsetEval and Genetic Search, a hybrid of CfsSubsetEval and Greedy Stepwise, a hybrid of ClassifierSubsetEval and Greedy Stepwise and a hybrid of WrapperSubsetEval and Greedy Stepwise methods. This shows that these feature selection methods do not select correctly.

In order to decide the feature selection method that provides the highest success rate, it is also necessary to check whether the error values of the hybrid of CfsSubsetEval and Best First method, which has the highest correlation coefficient value, decrease as a result of post-selection classification. Table 4.4 below shows the error values resulting from the classification of The hybrid of CfsSubsetEval and Best First method before and after the feature selection.

Tablo 4.4. The error values of the hybrid of CfsSubsetEval and Best First method

Attribute Evaluator	Search Method	RAE (Before Feature Selection)	Root Relative Squared Error (Before Feature Selection)	RAE (After Feature Selection)	Root Relative Squared Error (After Feature Selection)
CfsSubsetEval	Best First	14.9602 %	15.9146 %	14.7233 %	15.5014 %

As can be seen from the table, while the RAE and Root Relative Squared Error values of the hybrid CfsSubsetEval and Best First method before feature selection are 14.9602% and 15.9146%, respectively, the classification result after feature selection is 14.7233% and 15.5014% respectively. After feature selection, both error values decreased.

Considering both the increase in the correlation coefficient and the decrease in error values, the hybrid CfsSubsetEval and Best First was determined as the best feature selection method. Thus, the estimation part of this study continued with the 38 parameters selected by this hybrid method. These 38 selected parameters are shown in Table 4.5 below.

Tablo 4.5. Selected Parameters

Selected Parameters	
C1.1.1. Political and operational stability	C5.2.1. University-industry R&D collaboration
C1.2.1. Regulatory quality	C5.2.4. Joint venture/strategic alliance deals/bn PPP\$ GDP
C1.2.2. Rule of law	C5.2.5. Patent families/bn PPP\$ GDP
C1.3.1. Ease of starting a business	C5.3.3. ICT services imports, % total trade
C1.3.2. Ease of resolving insolvency	C6.1.1. Patents by origin/bn PPP\$ GDP
C2.1.4. PISA scales in reading, maths and science	C6.2.1. Labor productivity growth, %
C2.2.1. Tertiary enrolment, % gross	C6.2.4. ISO 9001 quality certificates/bn PPP\$ GDP
C2.3.1. Researchers, FTE/mn pop.	C6.3.1. Intellectual property receipts, % total trade
C2.3.3. Global corporate R&D investors, top 3, mn US\$	C6.3.2. High-tech exports
C3.1.1. ICT access	C6.3.3. ICT services exports
C3.2.1. Electricity output, GWh/mn pop.	C6.3.4. Foreign direct investment net outflows
C3.2.2. Logistics performance	C7.1.1. Trademarks by origin/bn PPP\$ GDP
C3.3.2. Environmental performance	C7.1.3. ICTs and organizational model creation
C4.1.1. Ease of getting credit	C7.2.2. National feature films/mn pop.
C4.1.2. Domestic credit to private sector	C7.2.3. Entertainment and media market/th pop.
C4.2.1. Ease of protecting minority investors	C7.3.1. Generic top-level domains (TLDs)/th pop.
C4.2.2. Market capitalization, % GDP	C7.3.2. Country-code TLDs/th pop.
C4.3.2. Intensity of local competition	C7.3.3. Wikipedia edits/mn pop.
C5.1.3. GERD performed by business, % GDP	C7.3.4. 'Mobile app creation/bn PPP\$ GDP

4.1.4. Artificial Neural Network Findings

In the previous finding, the best 38 parameters that determine the GII score were selected among 79 parameters that determine the GII score of 125 countries. 38 parameter data was used in the ANN method to find out how accurately it predicted the GII score by making predictions with these 38 parameters. Matlab 2022b was used to estimate the GII score for 125 countries. As stated in the first step of the ANN methodology, the study continued on the normalized data.

After the data is loaded into neural network fitting tool in Matlab, the input data and output data are set as follows: The selected 38 parameters on the 2012-2022 GII reports were determined as input data. As the target data, the GII scores of the years 2012-2022 was used. Data set was set as 70% training, 15% validation, and 15% testing. The ANN network of the study was as followed in Figure 4.1.

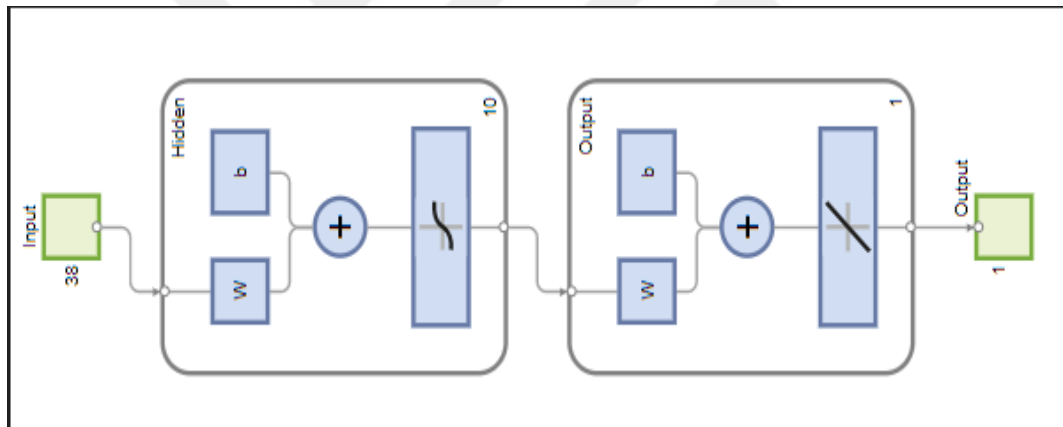


Figure 4.1. ANN Network

The number of hidden neurons was determined before training the network. The number of hidden neurons was determined according to the geometric pyramid rule. According to this rule; Since the number of hidden neurons will not exceed twice the number of input neurons, the number of hidden neurons should not exceed 76, which is twice the number of 38 input parameters. At the same time, since this number should not fall below the square root of the product of the number of input and output neurons, the number of hidden neurons should not exceed 6.16, which is the square root of the product of 38 input neurons and 1 output neuron. As a result, the number of hidden neurons should be between 6, 16 and 76. However, when the number of hidden neurons is determined by numbers with commas or by incrementing them by one, it will cause a lot of processing complexity. Therefore, for a more accurate prediction,

hidden neuron numbers in multiples of 10 between 6.16 and 76 will give better results. In other words, while training the network, the numbers of 10, 20, 30, 40, 50, 60 and 70 hidden neurons were tried separately and the prediction results were obtained and noted for comparison. Table 4.6 below shows the performance data for training, validation and testing of the data put into nftool for prediction.

Table 4.6. ANN Prediction Performance Results

Algorithm	Hidden Neuron Size	Steps	MSE	R	Epoch Size
Levenberg-Marquart	10	Training	0.0007	0.9925	14/1000 EPOCH
		Validation	0.0016	0.9797	
		Test	0.0017	0.9769	
Levenberg-Marquart	20	Training	0.0002	0.997	12/1000 EPOCH
		Validation	0.0012	0.9936	
		Test	0.0010	0.9899	
Levenberg-Marquart	30	Training	0.0006	0.9936	10/1000 EPOCH
		Validation	0.0011	0.9843	
		Test	0.0014	0.9808	
Levenberg-Marquart	40	Training	0.0002	0.998	11/1000 EPOCH
		Validation	0.0019	0.9751	
		Test	0.0013	0.984	
Levenberg-Marquart	50	Training	0.0004	0.997	10/1000 EPOCH
		Validation	0.0016	0.9814	
		Test	0.0015	0.9815	
Levenberg-Marquart	60	Training	0.0003	0.997	10/1000 EPOCH
		Validation	0.0018	0.9762	
		Test	0.0023	0.9695	
Levenberg-Marquart	70	Training	0.0001	0.9984	10/1000 EPOCH
		Validation	0.0024	0.9666	
		Test	0.0043	0.9444	
Bayesian Regularization	10	Training	0.0002	0.9968	407/1000 EPOCH
		Validation	NaN	NaN	
		Test	0.0011	0.985	
Bayesian Regularization	20	Training	0	0.9994	1000/1000 EPOCH
		Validation	NaN	NaN	
		Test	0.0027	0.9673	
Bayesian Regularization	30	Training	0	1	437/1000 EPOCH
		Validation	NaN	NaN	
		Test	0.0042	0.9462	
Scaled Conjugate Gradient	10	Training	0.0019	0.9728	41/1000 EPOCH
		Validation	0.0018	0.9778	
		Test	0.0019	0.9753	
	20	Training	0.0018	0.9752	61/1000 EPOCH
		Validation	0.0021	0.9664	

Scaled Conjugate Gradient		Test	0.0024	0.9688	
Scaled Conjugate Gradient	30	Training	0.0009	0.9873	67/1000 EPOCH
		Validation	0.0015	0.9785	
		Test	0.0014	0.9793	
Scaled Conjugate Gradient	40	Training	0.0006	0.9922	138/1000 EPOCH
		Validation	0.001	0.9839	
		Test	0.0022	0.9691	
Scaled Conjugate Gradient	50	Training	0.001	0.987	103/1000 EPOCH
		Validation	0.0022	0.9701	
		Test	0.0021	0.9696	
Scaled Conjugate Gradient	60	Training	0.0008	0.9883	106/1000 EPOCH
		Validation	0.0014	0.9785	
		Test	0.0012	0.9865	
Scaled Conjugate Gradient	70	Training	0.0016	0.9791	73/1000 EPOCH
		Validation	0.0033	0.9539	
		Test	0.0028	0.9586	

As seen in the table, unlike the other 2 methods, the Bayesian Regularization method was not trained on the numbers of 40, 50, 60 and 70 hidden neurons. The reason for this is that the R value reached 1 when training the model with the number of 30 hidden neurons. This means that the model is overfitting. Overfitting occurs when a statistical model captures noise in the data rather than the underlying relationship. This can occur when the model is too complex relative to the amount of training data, leading to the model fitting the training data too closely and performing poorly on new, unseen data (Goodfellow, Bengio, and Courville, 2016). Since the model learned too much, the model could not continue to be trained with the number of hidden neurons of 40, 50, 60 and 70.

R value and MSE value are examined to evaluate the performance of the trained data. The R value closest to 1 and the MSE value closest to 0 indicate the best prediction model. In addition, the fact that the prediction model reaches the prediction value at the nearest epoch is an indication that the model works well. Furthermore, after the model is trained, the model is tested. It is the R, MSE and epoch values of the model test results that determine the performance of the model. When the test performance of the models is compared in the table, the model with the lowest MSE value of 0.0010 and the highest R value of 0.9899 is Levenberg-Marquart with 20 hidden neurons. Additionally, this model achieved the prediction result in the 12th epoch of 1000 epochs. This result shows that the method giving the best results when

estimating a new GII data is the Levenberg-Marquart model with 20 neurons of the ANN method.

The slope and direction of the fit line indicate how it fits with the data. If the line is like a linear extension of the data points, the model fits the real data well. However, if the line is sloping or uneven, it may indicate that the model is making misleading predictions in certain conditions or that the data is trending in a particular way. Based on this interpretation, the fit line is like a linear extension of the data points, as seen in the regression plot of the Levenberg model below in Figure 4.2. This means that the model fits the real data well. This shows that the model makes correct predictions.

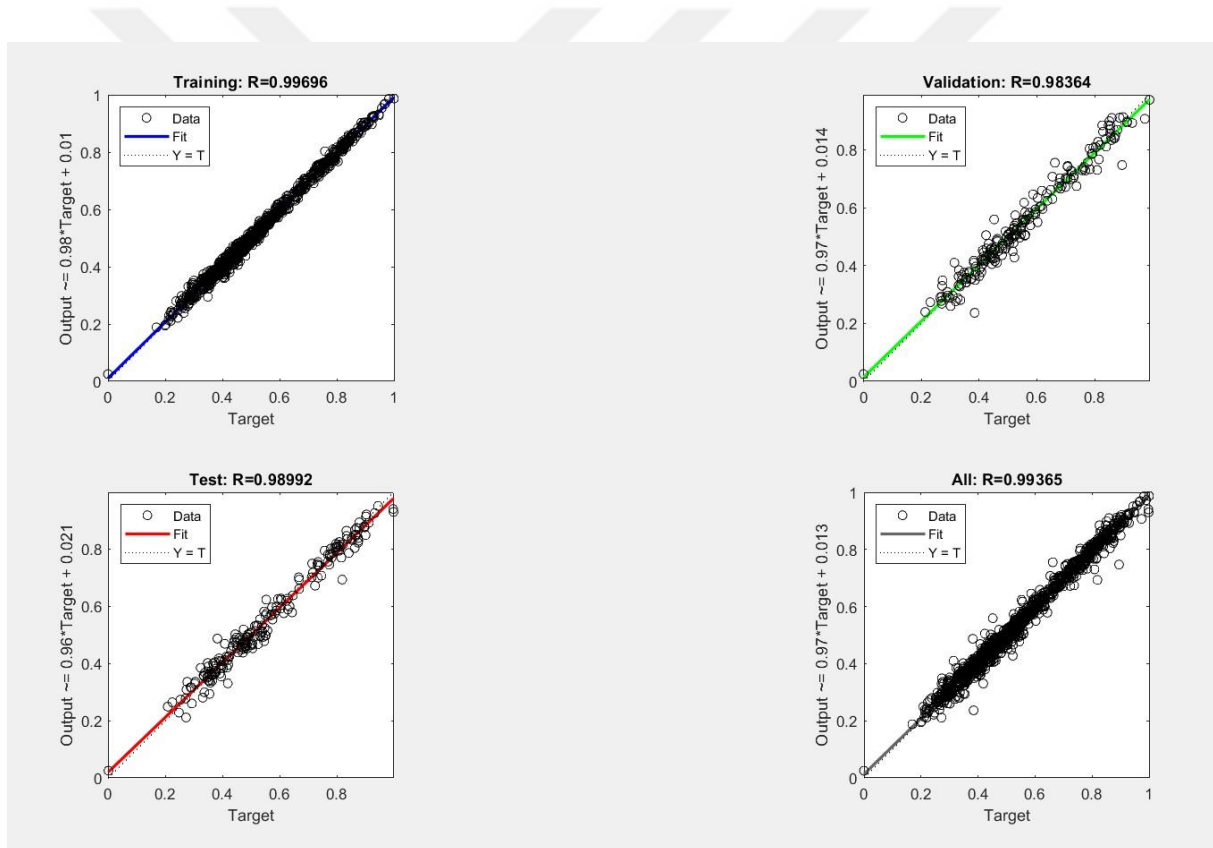


Figure 4.2. The Regression Plot of Levenberg-Marquardt Model

4.1.5. Regression Findings

The best 38 parameters used in the previous section and selected in feature selection were used in the prediction study with regression. The regression learner tool in Matlab 2022b was used to predict the GII score of 125 countries. As stated in the first step of the ANN methodology,

work continued on normalized data. After the data was loaded into the regression learner tool, the input data and output data were set as follows: 38 parameters selected in the 2012-2022 GII reports were determined as input data. GII scores from 2012-2022 were used as target data.

Validation scheme must be selected before the data is trained. Since large data set was used in this study, holdout validation scheme was chosen. In addition, as in the ANN method, the data set is divided into 70% training, 15% validation and 15% test data. This way, comparison of Regression and ANN results is facilitated. Table 4.7 below shows the performance data for validation and testing of the data put into regression learner tool for prediction.

Table 4.7. The Results of Regression Models

Regression Models	Types of Regression Models	MSE Validation	R Validation	MSE Test	R Test
Linear Regression	Linear Regression Model	0.0012	0.9820	0.0017	0.9780
Linear Regression	Interactions Regression Model	0.0083	0.8636	0.0036	0.9537
Linear Regression	Robust Regression Model	0.0012	0.9822	0.0017	0.9779
Regression Trees	Fine Tree	0.0027	0.9573	0.0023	0.9702
Regression Trees	Medium Tree	0.0028	0.9561	0.0022	0.9716
Regression Trees	Coarse Tree	0.0041	0.9353	0.0030	0.9610
SVM	Linear SVM	0.0012	0.9820	0.0017	0.9778
SVM	Quadratic SVM	0.0007	0.9900	0.0013	0.9840
SVM	Cubic SVM	0.0007	0.9897	0.0014	0.9819
SVM	Fine Gaussian SVM	0.0208	0.6050	0.0218	0.6729
SVM	Medium Gaussian SVM	0.0009	0.9865	0.0010	0.9879
SVM	Coarse Gaussian SVM	0.0011	0.9832	0.0015	0.9815
Ensembles of Regression Trees	Boosted Trees	0.0018	0.9718	0.0021	0.9728
Ensembles of Regression Trees	Bagged Trees	0.0012	0.9816	0.0012	0.9846
Neural Network Regression	Narrow Neural Network	0.0010	0.9840	0.0012	0.9844
Neural Network Regression	Medium Neural Network	0.0009	0.9859	0.0011	0.9865
Neural Network Regression	Wide Neural Network	0.0012	0.9810	0.0016	0.9802
Neural Network Regression	Bilayered Neural Network	0.0008	0.9882	0.0012	0.9844
Neural Network Regression	Trilayered Neural Network	0.0009	0.9858	0.0010	0.9872
Kernel Approximation Models	SVM Kernel	0.0015	0.9769	0.0018	0.9767
Kernel Approximation Models	Least Squared Regression Kernel	0.0018	0.9726	0.0018	0.9771

The data was trained with a total of 22 regression models. The Stepwise Linear Regression Model timed out while being trained and could not make predictions. The results of the other 21 models are as shown in the table. Looking at the test performances of the models, Medium

Gaussian SVM has the highest R value of 0.9879 and the lowest MSE value of 0.0010. This model is the best predictive model among the regression models.

The predicted response of the model is plotted against the actual response below. The predicted response of a perfect regression model is equal to the actual response, so all points lie on a diagonal line. The vertical distance of the line to any point is the error of the estimate made for that point. A good model has small errors, which means the predictions are distributed close to the line. Based on this interpretation, in the Medium Gaussian SVM regression model in Figure 4.3 below, all points are generally on the diagonal line and, except for some points, the predictions are distributed close to the line. The points slightly away from the line are the model's prediction error of 0.0010. As a result, the Medium Gaussian SVM regression model is determined to be a better predictive model by showing better performance than other regression models.

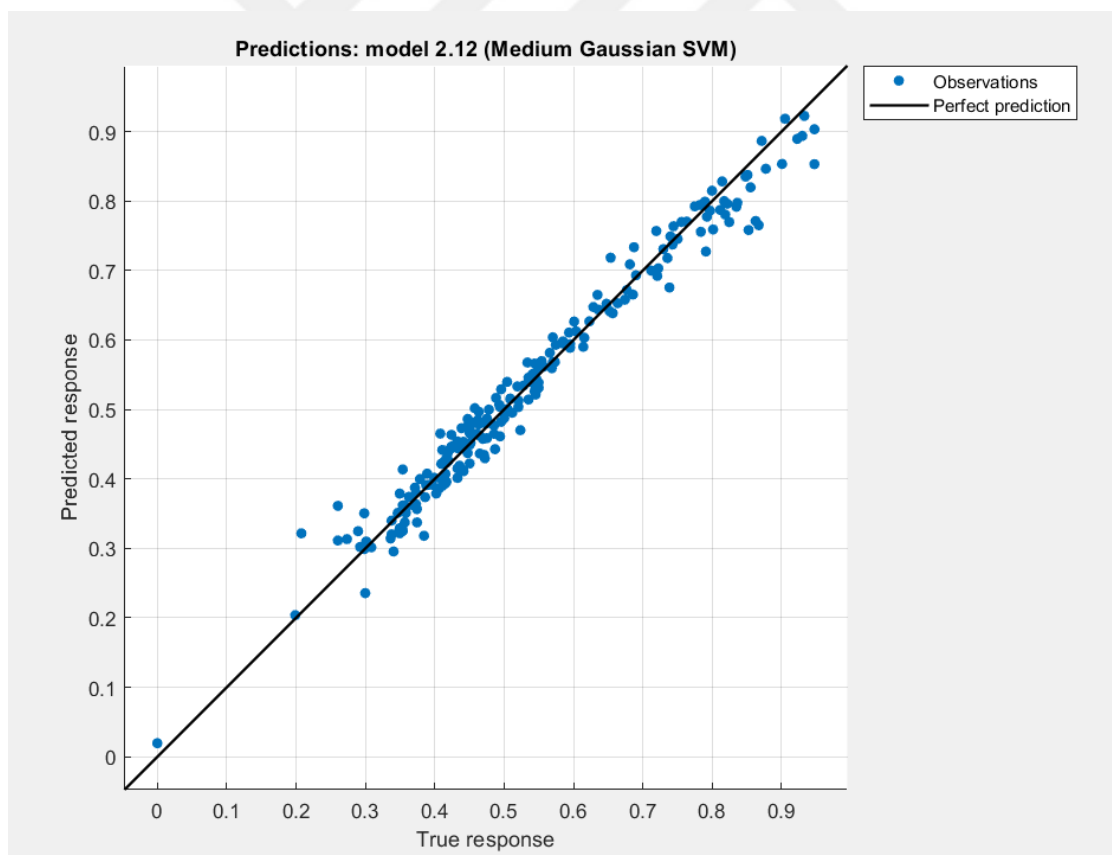


Figure 4.3. Predicted and True Response Plot

4.1.6. The Comparison Findings of the Artificial Neural Network and Regression Findings

In the previous sections, prediction studies were carried out with 3 different algorithms of the ANN method and different numbers of neurons, as well as 22 different models of the Regression method. As a result of the prediction study with the ANN method, the Levenberg-Marquart algorithm with a number of 20 neurons gave the best prediction performance, while the Medium Gaussian SVM regression model gave the best prediction performance as a result of the prediction study with the Regression method. Table 4.8 containing the prediction performance results of these two models is as follows.

Table 4.8. The Performance Results of the Best Two Methods

Model	Data Splitting	MSE-Test Data	R-Test Data
Levenberg-Marquart with 20 Hidden Neuron	%70 training, %15 validation and %15 test	0.001	0.9899
Medium Gaussian SVM Regression	%70 training, %15 validation and %15 test	0.001	0.9879

The results shown in the table are the R and MSE values of the test data of Levenberg-Marquart with 20 Hidden Neuron which was determined as the best prediction model of the ANN methods, and Medium Gaussian SVM Regression which was determined as the best prediction model of the Regression methods. When the performance results of both models are compared, it is seen that the MSE values are the same with 0.0010. In this case, the R value is looked at to determine which model is better. Of these two models, the model with the R value closest to 1 is the Levenberg-Marquart with 20 Hidden Neuron model of the ANN method with an R value of 0.9899.

As a result of the prediction study, the ANN method best predicts the GII score of 125 countries with 38 parameters. In addition, the Levenberg-Marquart model with 20 hidden neurons of the ANN method showed better prediction performance than other ANN models and Regression models, and also revealed that 38 parameters were sufficient to predict the GII score with its prediction performance.

4.1.7. Decision Support System Application Findings

Based on the findings of the prediction study, it was concluded that the Levenberg-Marquart ANN method with 20 hidden neurons, which was identified as the most effective prediction model among the ANN methods, providing the most accurate prediction of the GII score for 125 countries using 38 parameters. To facilitate this process, a prototype DSS has been designed with user-friendly interface screens for inputting the 38 parameters. The system employs the best prediction model to generate results that are displayed on the screen.

In the designed prototype DSS, Figure 4.4 illustrates the hierarchy of interface screens. Upon clicking the Predict GII button on the main screen, the data entry screen, as the second interface, appears. Upon completion of data entry, clicking the predict button yields the result screen, displaying the prediction outcome. Additionally, for users seeking to input new data on the result screen, a clear screen for new data button is included.

Decision Support System Based on Machine Learning to Predict the Global Innovation Index (GII)



Global Innovation Index (GII)

The GI functions as a vital tool for assessing innovation within the intricate context of global economic and geopolitical uncertainties. Through its analysis of innovation performance across approximately 132 economies, it identifies the foremost innovative nations while illuminating their specific strengths and weaknesses in innovation domains. https://www.wipo.int/global_innovation_index/en/

Predict GI

Data Entry Screen for Predicting GI

Political and operational stability:	Enter number between 0 and 100	Regulatory quality:	Enter number between 0 and 100
Rule of law:	Enter number between 0 and 100	Ease of starting a business:	Enter number between 0 and 100
Ease of resolving insolvency:	Enter number between 0 and 100	PISA scales in reading, maths and science:	Enter number between 0 and 100
Tertiary enrolment, % gross:	Enter number between 0 and 100	Researchers, FTE/mn pop.:	Enter number between 0 and 100
Global corporate R&D investors, top 3, mn US\$:	Enter number between 0 and 100	ICT access:	Enter number between 0 and 100
Electricity output, GWh/mn pop.:	Enter number between 0 and 100	Logistics performance:	Enter number between 0 and 100
Environmental performance:	Enter number between 0 and 100	Ease of getting credit:	Enter number between 0 and 100
Domestic credit to private sector:	Enter number between 0 and 100	Ease of protecting minority investors:	Enter number between 0 and 100
Market capitalization, % GDP:	Enter number between 0 and 100	Intensity of local competition:	Enter number between 0 and 100
GERD performed by business, % GDP:	Enter number between 0 and 100	University-industry R&D collaboration:	Enter number between 0 and 100
Joint venture/strategic alliance deals/bn PPP\$ GDP:	Enter number between 0 and 100	Patent families/bn PPP\$ GDP:	Enter number between 0 and 100
ICT services imports, % total trade:	Enter number between 0 and 100	Patents by origin/bn PPP\$ GDP:	Enter number between 0 and 100
Labor productivity growth, %:	Enter number between 0 and 100	ISO 9001 quality certificates/bn PPP\$ GDP:	Enter number between 0 and 100
Intellectual property receipts, % total trade:	Enter number between 0 and 100	High-tech exports:	Enter number between 0 and 100
ICT services exports:	Enter number between 0 and 100	Foreign direct investment net outflows:	Enter number between 0 and 100
Trademarks by origin/bn PPP\$ GDP:	Enter number between 0 and 100	ICTs and organizational model creation:	Enter number between 0 and 100
National feature films/mn pop.:	Enter number between 0 and 100	Entertainment and media market/th pop.:	Enter number between 0 and 100
Generic top-level domains (TLDs)/th pop.:	Enter number between 0 and 100	Country-code TLDs/th pop.:	Enter number between 0 and 100
Wikipedia edits/mn pop.:	Enter number between 0 and 100	Mobile app creation/bn PPP\$ GDP:	Enter number between 0 and 100

Predict

Prediction Result

Clear Screen for New Data Entry

Figure 4.4. Interface Screens Hierarchy

The main screen prototype of DSS is displayed in Figure 4.5. It provides a concise overview of the GI along with a button to predict the GI score.

Decision Support System Based on Machine Learning to Predict the Global Innovation Index (GII)



Figure 4.5. Main screen to predict the GII score

The screen for entering data is displayed in Figure 4.6. There are 38 parameter names that have been identified as the top predictors for the GII score. Next to each parameter name, there is a field where values between 0-100 can be entered.

Data Entry Screen for Predicting GII

Political and operational stability: Enter number between 0 and 100	Regulatory quality: Enter number between 0 and 100
Rule of law: Enter number between 0 and 100	Ease of starting a business: Enter number between 0 and 100
Ease of resolving insolvency: Enter number between 0 and 100	PISA scales in reading, maths and science: Enter number between 0 and 100
Tertiary enrolment, % gross: Enter number between 0 and 100	Researchers, FTE/mn pop.: Enter number between 0 and 100
Global corporate R&D investors, top 3, mn US\$: Enter number between 0 and 100	ICT access: Enter number between 0 and 100
Electricity output, GWh/mn pop.: Enter number between 0 and 100	Logistics performance: Enter number between 0 and 100
Environmental performance: Enter number between 0 and 100	Ease of getting credit: Enter number between 0 and 100
Domestic credit to private sector: Enter number between 0 and 100	Ease of protecting minority investors: Enter number between 0 and 100
Market capitalization, % GDP: Enter number between 0 and 100	Intensity of local competition: Enter number between 0 and 100
GERD performed by business, % GDP: Enter number between 0 and 100	University-industry R&D collaboration: Enter number between 0 and 100
Joint venture/strategic alliance deals/bn PPP\$ GDP: Enter number between 0 and 100	Patent families/bn PPP\$ GDP: Enter number between 0 and 100
ICT services imports, % total trade: Enter number between 0 and 100	Patents by origin/bn PPP\$ GDP: Enter number between 0 and 100
Labor productivity growth, %: Enter number between 0 and 100	ISO 9001 quality certificates/bn PPP\$ GDP: Enter number between 0 and 100
Intellectual property receipts, % total trade: Enter number between 0 and 100	High-tech exports: Enter number between 0 and 100
ICT services exports: Enter number between 0 and 100	Foreign direct investment net outflows: Enter number between 0 and 100
Trademarks by origin/bn PPP\$ GDP: Enter number between 0 and 100	ICTs and organizational model creation: Enter number between 0 and 100
National feature films/mn pop.: Enter number between 0 and 100	Entertainment and media market/th pop.: Enter number between 0 and 100
Generic top-level domains (TLDs)/th pop.: Enter number between 0 and 100	Country-code TLDs/th pop.: Enter number between 0 and 100
Wikipedia edits/mn pop.: Enter number between 0 and 100	Mobile app creation/bn PPP\$ GDP: Enter number between 0 and 100

[Predict](#)

Figure 4.6. Data entry screen

The error displayed on the data entry screen is depicted in Figure 4.7. It occurs when a value is entered which falls outside the range of 0-100.

Oops...
enterNumbers between 0 and 100

Figure 4.7. The error displayed on the data entry screen

The result of the prediction and the clear screen button for entering new data can be seen in Figure 4.8.

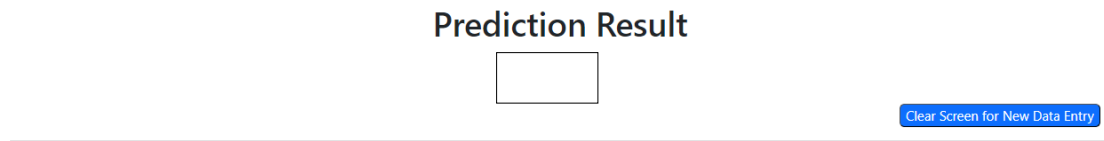


Figure 4.8. The prediction result screen

In order to evaluate the efficacy of the newly designed prototype DSS, the dataset of 2020-2023 GII scores was utilized. It's worth noting that the 2023 dataset was not part of the previously applied ML methods.

The designed prototype DSS was tested using a specific data set consisted of the 2020, 2021, 2022 and 2023 GII datas, which were used to evaluate whether the system worked correctly. For this purpose, the first three countries in the years between 2020-2023 GII ranking were selected as the representative data points. The designed prototype DSS predicted GII scores of the three countries for the years of 2020-2023. These GII scores were entered into the system, and after pressing the prediction button, the system generated predictions for these countries. A chart was generated to compare the predicted GII scores with the real GII scores of these countries. This chart in Figure 4.9 provides a comparison of the predicted and real GII scores for the years 2020-2023.

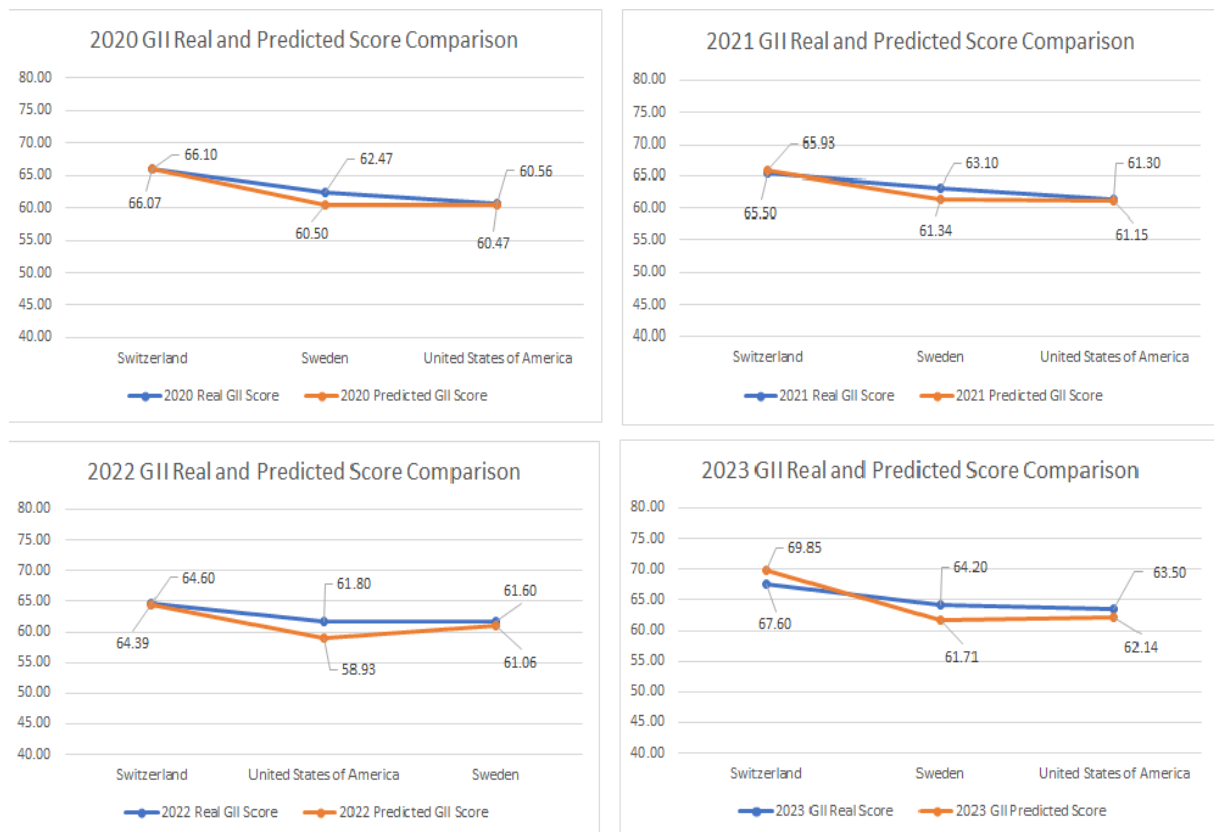


Figure 4.9. 2023 GII Real and Predicted Score Comparison

Switzerland, the top-ranking country in GII score in 2020, had an actual GII score of 66.10 and an estimated GII score of 66.07. Sweden, ranked second, had a predicted GII score of 60.50 and an actual score of 62.47. The United States of America, ranked third, had an actual GII score of 60.56 and a predicted score of 60.47 for 2020. For 2021, the estimated GII scores for the top 3 countries were compared to the actual scores. Switzerland had an actual GII score of 65.50 and a predicted score of 65.93. Sweden had an actual GII score of 63.10 and a predicted score of 61.34. The United States had an actual GII score of 61.30 and a predicted score of 61.15. The comparison continued for 2022 and 2023. In 2022, Switzerland's actual GII score was 64.60, and the predicted score was 64.39. The United States had an actual GII score of 61.80 and a predicted score of 58.93. Sweden's actual GII score was 61.60, and the predicted score was 61.06. In 2023, Switzerland's actual GII score was 67.60, and the predicted score was 69.85. Sweden's actual GII score was 64.20, and the predicted score was 61.71. The United States had an actual GII score of 63.50 and a predicted score of 62.14. The close correlation between predicted and actual scores demonstrates the accuracy of the DSS prototype designed and tested using the 2020-2023 data. These results indicate the effective functionality of the system.

4.2. Discussion

Table 4.9 below shows a total of 21 sub-indexes, 7 of which determine the input and output that determine the GII score. In addition, a total of 79 parameters that determine these sub-indexes and whether these parameters are selected or not are shown with the feature selection.

Table 4.9. Selected and Not Selected Parameters

Inputs and outputs that determine the GII score	Sub_indexes	Parameters/ Parameters that determine sub indexes	Selected parameter or not?
C1 Institutions	C1.1. Political environment	C1.1.1. Political and operational stability	Selected
		C1.1.2. Government effectiveness	Not Selected
	C1.2. Regulatory environment	C1.2.1. Regulatory quality	Selected
		C1.2.2. Rule of law	Selected
		C1.2.3. Cost of redundancy dismissal	Not Selected
	C1.3. Business environment	C1.3.1. Ease of starting a business	Selected
		C1.3.2. Ease of resolving insolvency	Selected
C2 Human capital & research	C2.1. Education	C2.1.1. Expenditure on education, % GDP	Not Selected
		C2.1.2. Government funding/pupil, secondary, % GDP/cap	Not Selected
		C2.1.3. School life expectancy, years	Not Selected
		C2.1.4. PISA scales in reading, maths and science	Selected
		C2.1.5. Pupil-teacher ratio, secondary	Not Selected
	C2.2. 'Tertiary education'	C2.2.1. Tertiary enrolment, % gross	Selected
		C2.2.2. Graduates in science and engineering, %	Not Selected
		C2.2.3. Tertiary inbound mobility, %	Not Selected
	C2.3. 'Research and development (R&D)'	C2.3.1. Researchers, FTE/mn pop.	Selected
		C2.3.2. Gross expenditure on R&D, % GDP	Not Selected
		C2.3.3. Global corporate R&D investors, top 3, mn US\$	Selected
		C2.3.4. QS university ranking, top 3	Not Selected
C3 Infrastructure	C3.1. Information and communication technologies (ICTs)	C3.1.1. ICT access	Selected
		C3.1.2. ICT use	Not Selected

		C3.1.3. Government's online service	Not Selected
		C3.1.4. E-participation	Not Selected
	C3.2. General infrastructure	C3.2.1. Electricity output, GWh/mn pop.	Selected
		C3.2.2. Logistics performance	Selected
		C3.2.3. Gross capital formation, % GDP	Not Selected
	C3.3. Ecological sustainability	C3.3.1. GDP/unit of energy use	Not Selected
		C3.3.2. Environmental performance	Selected
		C3.3.3. ISO 14001 environmental certificates/bn PPP\$ GDP	Not Selected
C4 Market sophistication	C4.1. Credit	C4.1.1. Ease of getting credit	Selected
		C4.1.2. Domestic credit to private sector	Selected
		C4.1.3. Microfinance gross loans, % GDP	Not Selected
	C4.2. Investment	C4.2.1. Ease of protecting minority investors	Selected
		C4.2.2. Market capitalization, % GDP	Selected
		C4.2.3. Venture capital recipients, deals/bn PPP\$ GDP	Not Selected
	C4.3. Trade, diversification, and market scale	C4.3.1. Applied tariff rate, weighted avg., %	Not Selected
		C4.3.2. Intensity of local competition	Selected
		C4.3.3. Domestic market scale, bn PPP\$	Not Selected
C5 Business sophistication	C5.1. Knowledge workers	C5.1.1. Knowledge-intensive employment, %	Not Selected
		C5.1.2. Firms offering formal training, %	Not Selected
		C5.1.3. GERD performed by business, % GDP	Selected
		C5.1.4. GERD financed by business, %	Not Selected
		C5.1.5. Females employed w/advanced degrees, %	Not Selected
	C5.2. Innovation linkages	C5.2.1. University-industry R&D collaboration	Selected
		C5.2.2. State of cluster development and depth	Not Selected
		C5.2.3. GERD financed by abroad, % GDP	Not Selected
		C5.2.4. Joint venture/strategic alliance deals/bn PPP\$ GDP	Selected

	C5.3. Knowledge absorption	C5.2.5. Patent families/bn PPP\$ GDP	Selected
		C5.3.1. Intellectual property payments, % total trade	Not Selected
		C5.3.2. High-tech imports, % total trade	Not Selected
		C5.3.3. ICT services imports, % total trade	Selected
		C5.3.4. FDI net inflows, % GDP	Not Selected
		C5.3.5. Research talent, % in businesses	Not Selected
C6 Knowledge and technology outputs	C6.1. Knowledge creation	C6.1.1. Patents by origin/bn PPP\$ GDP	Selected
		C6.1.2. PCT patents by origin/bn PPP\$ GDP	Not Selected
		C6.1.3. Utility models by origin/bn PPP\$ GDP	Not Selected
		C6.1.4. Scientific and technical articles/bn PPP\$ GDP	Not Selected
		C6.1.5. Citable documents H index	Not Selected
	C6.2. Knowledge impact	C6.2.1. Labor productivity growth, %	Selected
		C6.2.2. New businesses/th pop.	Not Selected
		C6.2.3. Software spending, % GDP	Not Selected
		C6.2.4. ISO 9001 quality certificates/bn PPP\$ GDP	Selected
		C6.2.5. High-tech manufacturing, %	Not Selected
	C6.3. Knowledge diffusion	C6.3.1. Intellectual property receipts, % total trade	Selected
		C6.3.2. High-tech exports	Selected
		C6.3.3. ICT services exports	Selected
		C6.3.4. Foreign direct investment net outflows	Selected
C7 Creative outputs	C7.1. Intangible assets	C7.1.1. Trademarks by origin/bn PPP\$ GDP	Selected
		C7.1.2. Industrial designs by origin/bn PPP\$ GDP	Not Selected
		C7.1.3. ICTs and organizational model creation	Selected
	C7.2. Creative goods and services	C7.2.1. Cultural and creative services exports, % total trade	Not Selected
		C7.2.2. National feature films/mn pop.	Selected

		C7.2.3. Entertainment and media market/th pop.	Selected
		C7.2.4. Printing and other media, % manufacturing	Not Selected
		C7.2.5. Creative goods exports, % total trade	Not Selected
	C7.3. Online creativity	C7.3.1. Generic top-level domains (TLDs)/th pop.	Selected
		C7.3.2. Country-code TLDs/th pop.	Selected
		C7.3.3. Wikipedia edits/mn pop.	Selected
		C7.3.4. Mobile app creation/bn PPP\$ GDP	Selected

It is understood from the table that the 38 parameters selected with the feature selection method selected at least one from each sub-index and input/output to which it was connected. This shows that every input, output and sub-index that determines the GII score is very important. This study shows that it is not possible to determine the GII score without them.

Table 4.10 below shows what percentage of the 7 input and output index parameters that determine the GII score are selected in feature selection.

Table 4.10. Percentage of Selected Parameters

Indexes	Percentage of Selected Parameters
C1 Institutions	71.43%
C2 Human capital & research	33.33%
C3 Infrastructure	40.00%
C4 Market sophistication	55.56%
C5 Business sophistication	33.33%
C6 Knowledge and technology outputs	50.00%
C7 Creative outputs	66.67%

It is shown in the table that the parameters that determine the C1 Institutions index are mostly selected. Then, from most to least, C7 Creative outputs, C4 Market sophistication, C6 Knowledge and technology outputs, C3 Infrastructure, C5 Business sophistication and C2 Human capital & research are selected. This result shows that the most important indexes determining the GII score are C1 Institutions. Studies emphasizing the importance of Institutions indexes on GII performance are supported by the literature below.

Institutions can be categorized as either formal or informal (Minto-Coy and McNaughton, 2016). Formal institutions include property rights, contracts, policies, regulations, laws, and constitutions. On the other hand, informal institutions refer to the culture and societal norms that dictate the operations of businesses (Berman, 2013; Minto-Coy and McNaughton, 2016). A country with weak institutional factors or poorly implemented institutional factors hinders innovation by eroding the confidence of investors, customers, and industries (Jovovic, Draskovic, Delibasic, and Jovovic, 2017; Szalacha-Jarmużek and Pietrowicz, 2018). The political environment of a country also affects the efficiency of innovation. Therefore, increasing the investment attractiveness of locally-owned firms with the assistance of local authorities creates good opportunities for innovation (Yachmeneva, 2014). According to Oluwatobi, Ola-David, Olurinola, Alege, and Ogundipe (2016), the quality of institutions, including government effectiveness, political stability, the rule of law, and regulatory quality, significantly affects innovation output in Sub-Saharan Africa. Similarly, a study by Okrah and Hajduk-Stelmachowicz (2020) concludes that a favorable institutional environment fosters innovativeness. A conducive political climate encourages small businesses and entrepreneurs to explore various opportunities, and the people's confidence in their political system stimulates innovativeness. Additionally, a sound legal and regulatory structure promotes the growth and innovativeness of small businesses (Nyarku and Oduro, 2018). Wang, Kafouros, Yi, Hong, and Ganotakis (2020) further assert that Chinese firms with high government affiliations are more innovative than others. They also argue that the government's influence on innovativeness is primarily related to the intensity of the protection of intellectual property by legal institutions (C. Wang et al., 2020; Wang, Luo, Sari, and Shao, 2020). Institutional pillar variables such as business environment, legal environment and political environment are part of the set of variables in the best innovation prediction model and are also supported by previous results (Jovovic et al., 2017; Szalacha-Jarmużek and Pietrowicz, 2018; Udimal, Jincai, Ibn Musah, and Hua, 2019).

The GII score, which is utilized to evaluate a country's innovation performance, is determined by 38 meticulously selected parameters that are supported by various literature studies. These studies have emphasized the significance of these parameters in accurately determining the GII score of countries. To provide a more comprehensive understanding of these parameters, the studies have explained their importance in detail and summarized them in a table at the end of

these explanations. Through analyzing these parameters, policy makers and researchers can acquire valuable information about the innovation performance of countries and develop effective strategies to address this issue.

Achieving economic development is a crucial factor for political stability. This is because high levels of economic growth improve living standards, which eventually leads to political and democratic change (Van de Walle, 1998). Political and operational stability are crucial for economic growth. A stable political environment encourages investment, innovation, and sustainable growth. On the other hand, operational stability ensures smooth production, timely delivery, and customer satisfaction. Both political and operational stability play a pivotal role in achieving organizational goals and economic prosperity (Acar, 2019; Li, Niu, Chen, and Zhou, 2020; Laura T. Schelhas et al., 2019; Singha and Singh, 2022; Sui, Chang, and Chu, 2021).

An indicator of regulatory quality assesses the government's ability to establish and execute effective policies and regulations, which encourage private sector growth. The quality of regulations is a critical factor for economic advancement and is a top priority for the World Bank and the OECD. According to their international network, countries are evaluated based on the quality of their rule-making process and the efficiency of their regulatory environment (Bank, 2001; Jacobzone, Choi, and Miguet, 2007).

Shaikh and Malik (2020) conducted an extensive study on the theoretical concepts and critical analysis of the historical understanding and development of the rule of law. They emphasized the importance of the rule of law for nations, as having stable legal frameworks can lead to economic progress and development. According to a report by Iglesias and Ordenes (2019), the media plays a crucial role in shaping public opinion and influencing the rule of law. The rule of law is necessary to ensure fairness and protection in emerging digital labor markets. In summary, the rule of law is not only a legal concept, but it is also a fundamental element of societal well-being and economic progress. Its impact extends beyond legal systems and shapes governance, human rights, and innovation. These studies highlight the critical role of the rule of law in promoting stability, justice, and prosperity (Postema, 2023; Sellers, 2014).

Bétila (2021) examines how Ease of Doing Business impacts economic growth in 44 African countries. The study finds a significant and positive correlation, indicating that regulatory reforms contribute to sustainable economic growth. Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2002) have emphasized that by simplifying business startup procedures, the number of entrepreneurs can increase and thus foster economic growth. Klapper and Love (2011) have demonstrated that by easing the business startup process, financial development can be promoted and businesses can have enhanced access to finance. According to a study conducted by Ayyagari, Demirgüç-Kunt, and Maksimovic (2011), reducing the bureaucratic burden associated with starting a business leads to an increase in the rate of firm entry, which in turn contributes positively to economic growth. In addition, Haltiwanger, Jarmin, and Miranda (2013) conducted research showing that facilitating business startups is associated with higher levels of job creation and employment growth. All these studies provide further evidence of the positive impact of simplifying business startup procedures on economic growth, firm entry rates, and job creation.

Porta, Lopez-de-Silanes, and Shleifer (2008) state that a well-functioning system for resolving insolvencies leads to better economic efficiency by allowing unproductive companies to exit the market quickly and reallocating resources to more productive uses. According to a study by World-Bank (2016), efficient insolvency systems help to increase access to credit by reducing the risk for creditors, which in turn stimulates investment and economic growth. Claessens and Klapper (2005) found that properly designed insolvency procedures help promote entrepreneurship by providing a safety net for entrepreneurs to take risks, knowing that they can exit ventures without facing undue hardship. Berkowitz, Pistor, and Richard (2003) suggest that effective insolvency resolution mechanisms have a positive impact on firm dynamics by facilitating the exit of inefficient firms and promoting competition in the market. Research by Beck, Demirgüç-Kunt, and Martinez Peria (2008) demonstrates that strong insolvency frameworks boost investor confidence, encouraging greater capital flows and greater economic resilience. Collectively, these studies highlight the crucial role played by efficient insolvency resolution mechanisms in promoting economic efficiency, facilitating credit access, encouraging entrepreneurship, promoting sound firm dynamics, and enhancing investor confidence.

Hanushek and Woessmann (2023) research demonstrated that a strong correlation exists between the PISA scores of a country in reading, mathematics, and science, and its long-term economic growth. Better educational outcomes, indicated by higher PISA scores, play a crucial role in economic development. In a study conducted by Chetty, Hendren, Kline, Saez, and Turner (2014), it was found that social mobility is closely linked to PISA scores, and countries with higher scores generally experience greater intergenerational income mobility. According to Heckman and LaFontaine (2010), PISA scores are vital in assessing the formation of skills, and higher PISA scores are related to better cognitive and non-cognitive skill development among students. Helliwell and Huang (2014) research demonstrated that PISA scores are positively correlated with a country's global competitiveness, as measured by factors such as innovation, productivity, and economic resilience. These studies collectively underscore the significant impact of PISA scores on economic growth, social mobility, skill formation, economic inequality, and global competitiveness, highlighting the importance of educational quality in shaping a nation's future.

Becker (2009) contended that the rise in tertiary enrollment rates contribute towards economical development by generating a proficient workforce, encouraging innovation, and endorsing technological progression. Oreopoulos, Page, and Stevens (2006) discovered that individuals with tertiary education attain higher earnings and have a lesser chance of unemployment, indicating that an increase in tertiary enrollment rates can enhance labor market outcomes and decrease income inequality. According to Blanden and Machin (2004), tertiary education acts as a means of social mobility, enabling people from disadvantaged backgrounds to access better-paying jobs and improve their socioeconomic status. A study by Bils (2000) implied that a rise in tertiary enrollment rates leads to more innovation and technological growth, which are vital drivers of long-term economic development. Lutz and Kc (2010) emphasized the importance of increasing tertiary enrollment rates among women for achieving gender equality and empowering women in society. According to Psacharopoulos (1994), higher tertiary enrollment rates are associated with greater economic resilience, as a well-educated workforce can adapt more effectively to changing economic conditions. The studies collectively highlight the diverse benefits of increasing tertiary enrollment rates, including economic development, labor market outcomes, social mobility, innovation, gender equality, and economic resilience.

Stephan (2012) conducted a study that demonstrated a correlation between higher densities of researchers per million population and increased innovation levels, as measured by indicators such as scientific publications and patent filings. Adams (1990) discovered a positive link between the density of researchers and economic growth, indicating that investments in research capacity contribute to long-term economic development. Jaffe (1989) highlighted the crucial role of researcher density in assisting knowledge spillovers and technology transfer, which are essential drivers of innovation and economic progress. Bloom, Jones, Van Reenen, and Webb (2020) found that regions with higher densities of researchers per population tend to have more innovative firms and higher levels of productivity. Phelan (1994) discovered a robust and affirmative relationship between researcher density and scientific output, implying that increasing the number of researchers per population leads to a greater quality and quantity of research. De la Fuente and Ciccone (2003) found that regions with higher densities of researchers experience greater social progress, as determined by indicators such as quality of life, health outcomes, and education attainment. These studies collectively emphasize the critical role of researcher density in driving innovation, economic growth, knowledge spillovers, firm performance, scientific output, and social progress.

Archibugi and Coco (2004) conducted a research that demonstrated the contribution of top corporate R&D investors to innovation and technological development worldwide, which eventually leads to economic growth and competitiveness. Hall and Ziedonis (2001) carried out a study to investigate the effect of leading corporate R&D investments on industry competitiveness. They found that the industries with higher levels of R&D investments from top corporations tend to be more innovative and competitive. Griliches (1998) conducted a research on top corporate R&D investors and their impact on economic development, and suggested that increased R&D investments by leading corporations promote productivity growth and enhance national competitiveness. Hall and Lerner (2010) carried out a study to investigate the relationship between corporate R&D investment and firm performance. They found that companies among the top R&D investors tend to achieve higher levels of productivity, profitability, and market value. Chesbrough and Appleyard (2007) explored the strategies of top corporate R&D investors in adapting to and shaping global technological trends, emphasizing their critical role in driving technological change and innovation. These studies collectively highlight the significant impact of top global corporate R&D investors on

innovation, industry competitiveness, technological leadership, economic development, firm performance, and global technological trends.

Governments worldwide are increasingly adopting Information and Communication Technology (ICT) to provide online services, also known as eGovernment. According to West (2005), 19% of all governmental organizations worldwide currently offer online services. The implementation of eGovernment services can result in several benefits, including improved efficiency, better access to services, greater accountability, transparency, and empowerment of citizens (Lam, 2005; Tung and Rieck, 2005). It can also lead to reduced costs and time for services (Bhatnagar, 2000; Gilbert, Balestrini, and Littleboy, 2004), better decision-making, improved knowledge sharing and organizational learning, and better interactions with citizens, businesses, and other government organizations, leveraging market forces for better public-private sector relations, and greater capacity to implement organizational change management (Grant and Chau, 2005; Tung and Rieck, 2005; Zhang, Dawes, and Sarkis, 2005).

Sovacool (2014) analyzed the environmental consequences of electricity output per capita. It highlighted the significance of transitioning to cleaner energy sources to mitigate environmental degradation and combat climate change. Dogan and Deger (2016) investigated the role of electricity output per capita in promoting industrialization. Their research showed that higher levels of electricity production support the growth of manufacturing and other energy-intensive sectors, driving economic diversification and structural transformation. Sovacool and Mukherjee (2011) examined the connection between electricity output per capita and energy security. They emphasized the importance of ensuring reliable and affordable electricity supply to enhance energy independence and resilience to supply disruptions. These studies collectively highlight the multifaceted importance of electricity output measured in GWh per million population, spanning economic development, quality of life, environmental sustainability, industrialization, energy security, and urbanization.

Jaramillo et al. (2018) conducted that efficient logistics performance has a significant impact on a country's economic competitiveness. It reduces transportation costs, enhances supply chain efficiency and market access, thereby encouraging trade and economic growth. A study by Arvis, Alina Mustra, Panzer, Ojala, and Naula (2007) found that there is a positive correlation between logistics performance and firm productivity. Effective logistics operations help firms

streamline processes, reduce lead times, and enhance customer satisfaction, ultimately leading to improved productivity and competitiveness. Djankov, Freund, and Pham (2010) investigated the relationship between logistics performance and trade facilitation policies. They concluded that efficient logistics systems, supported by trade facilitation measures such as customs efficiency and infrastructure quality, are essential for reducing trade costs and promoting international trade. Limao and Venables (2001) conducted a study on the impact of logistics performance on export growth and found that countries with better logistics infrastructure and services have higher export growth rates. This highlights the crucial role of logistics in facilitating trade expansion. Peck and Christopher (2004) emphasized the crucial role of logistics performance in enhancing supply chain resilience. Effective logistics management enables firms to respond swiftly to disruptions, minimize risks, and maintain business continuity, particularly in the face of unforeseen events. Mentzer, Stank, and Esper (2008) investigated the relationship between logistics performance and customer satisfaction. They concluded that superior logistics services, including timely delivery and order accuracy, contribute significantly to customer satisfaction and loyalty. These studies collectively highlight the critical role of logistics performance in driving economic competitiveness, firm productivity, trade facilitation, export growth, supply chain resilience, and customer satisfaction.

Stern (2004) conducted research which illustrates the significant impact that environmental performance can have on economic growth. Efficient environmental policies and practices can promote long-term economic prosperity through sustainable development and technological innovation. Ambec and Lanoie (2008) explored the correlation between corporate profitability and environmental performance. They observed that firms with superior environmental performance tend to benefit from cost savings, enhanced brand reputation, and improved financial performance over time. Horbach (2008) studied the relationship between environmental performance and regulatory compliance. Their findings suggest that companies with strong environmental performance are better prepared to comply with environmental regulations, which reduces the risk of legal liabilities and penalties. Prüss-Üstün, Wolf, Corvalán, Bos, and Neira (2016) conducted research into the connection between environmental performance and public health outcomes. They found that pollution control and resource conservation lead to significant health benefits, including reduced morbidity and mortality rates. Martinez-Alier (2003) emphasized the importance of environmental

performance in promoting social equity. Environmental degradation often affects vulnerable populations disproportionately, increasing social inequalities. Improving environmental performance can help address these disparities and promote environmental justice. Metz, Davidson, Bosch, Dave, and Meyer (2014) assessed the role of environmental performance in climate change mitigation efforts. Effective environmental policies, such as the adoption of renewable energy and carbon pricing, are crucial for reducing greenhouse gas emissions and limiting global warming. These studies collectively underline the critical importance of environmental performance in driving economic growth, corporate profitability, regulatory compliance, public health, social equity, and climate change mitigation efforts.

Bayraktar (2015) states that credit indicators play a crucial role in attracting FDI inflows. According to him, countries with better credit indicators tend to receive more FDI inflows. Piwonski's comprehensive research also confirms that getting credit can enhance FDI flows (Piwonski, 2010). Bayraktar (2015); MogesEbero and Begum (2016); Morris and Aziz (2011); Singh (2015) have also observed that better business behavior can attract more FDI.

The significance of private sector domestic credit cannot be overstated when it comes to economic development. Its impact is far-reaching, contributing significantly to increased investment, employment, productivity, and a reduction in poverty levels (Hamida Begum and Aziz, 2019). Various studies have explored the link between private sector credit and economic growth, revealing a definitive long-term relationship (Ahmed, 2008; Akpansung and Babalola, 2011; Aliero, Abdullahi, and Adamu, 2013; Majeed and Ahmad, 2009; Pradhan, 2009).

Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998) conducted a significant study that emphasized the significance of safeguarding minority investors for investor confidence and market development. They discovered that countries with strong legal protections for minority shareholders tend to have higher levels of stock market participation and investment. Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2008) studied the influence of minority investor protection on economic development. They concluded that countries with more robust legal frameworks for safeguarding minority investors are more likely to attract foreign investment, facilitate entrepreneurship, and promote economic growth. La Porta, Lopez-de-Silanes, Shleifer, and Vishny (2002) conducted research that explored the correlation between minority investor protection and corporate governance practices. They found that countries with strong

legal protections for minority shareholders tend to have better corporate governance standards, including greater transparency, accountability, and board independence. Doidge, Karolyi, and Stulz (2007) investigated the effect of minority investor protection on market liquidity. They concluded that countries with strong legal protections for minority investors tend to have more liquid and efficient stock markets, as investors are more confident in the fairness and transparency of trading. Claessens, Djankov, and Lang (2000) analyzed the correlation between minority investor protection and FDI inflows. They found that countries with strong legal safeguards for minority shareholders attract higher levels of FDI, as investors perceive lower risks and greater confidence in the legal framework. La Porta, Lopez-de-Silanes, and Shleifer (1999) studied the influence of minority investor protection on firm performance. They found that countries with strong legal protections for minority shareholders tend to have higher firm valuations, better investment opportunities, and improved long-term performance. Collectively, these studies highlight the critical importance of protecting minority investors for investor confidence, economic development, corporate governance, market liquidity, foreign investment, and firm performance.

Levine and Zervos (1998) conducted research that explores the correlation between economic development and market capitalization. Their findings indicate that a higher market capitalization relative to GDP can lead to increased financial intermediation, investment efficiency, and economic growth. This suggests that well-developed capital markets play a significant role in fostering economic development. Merton (1992) carried out a study that examines how market capitalization as a percentage of GDP affects financial stability. The study suggests that a large enough capital market, reflected in significant market capitalization relative to GDP, can help enhance financial resilience by providing diverse sources of financing and risk-sharing mechanisms. Bekaert, Harvey, and Lundblad (2005) analyzed the relationship between market capitalization and investor confidence. Their research shows that a larger market capitalization relative to GDP is associated with greater investor confidence. This is because it reflects deeper and more liquid capital markets that foster trust and participation from domestic and international investors. Beck, Demirgüç-Kunt, and Levine (2000) conducted research that investigates the role of market capitalization relative to GDP in capital formation. The study found that countries with higher market capitalization levels tend to attract more investment capital. This, in turn, stimulates entrepreneurship and facilitates capital formation, contributing to overall economic growth and development. Demirguc-Kunt (2008) carried out

a study that explores the correlation between market capitalization as a percentage of GDP and financial inclusion. They argue that deeper and more developed capital markets, as reflected in higher market capitalization levels, are associated with increased financial access and inclusion. This promotes broad-based economic growth and poverty reduction. Claessens and Yurtoglu (2013) conducted research that examines the relationship between market capitalization relative to GDP and wealth distribution. The study found that countries with larger capital markets, as reflected in higher market capitalization levels, tend to have more equitable wealth distribution. This suggests that capital market development can contribute to reducing income inequality. These studies collectively highlight the critical role of market capitalization as a percentage of GDP in areas such as economic development, financial stability, investor confidence, capital formation, financial inclusion, and wealth distribution.

Porter (2011) research highlights that local competition has a positive effect on firm innovation. Competing with local rivals encourages firms to invest in R&D, adopt new technologies, and enhance product quality to gain an advantage, ultimately leading to innovation and economic growth. Berry, Levinsohn, and Pakes (1995) study examines how local competition affects consumer benefits. They found that intense local competition results in lower prices, improved product variety, and better service quality, which ultimately benefits consumers. Caves (1998) study investigates how local competition affects firm efficiency. Competitive pressure from local rivals forces firms to streamline operations, reduce costs, and increase productivity to maintain their market share, leading to overall efficiency gains and economic performance. (Audretsch and Feldman, 1996) research explores the relationship between local competition and entrepreneurship. They found that competitive local markets offer opportunities for new entrants to challenge incumbents, introduce innovative business models, and spur economic dynamism. (Tirole, 1988) study analyzes the relationship between local competition and market structure. Intense local competition often results in fragmented market structures with numerous small firms competing vigorously, fostering innovation, flexibility, and responsiveness to changing consumer preferences. These studies collectively emphasize the critical importance of the intensity of local competition in driving firm innovation, consumer benefits, firm efficiency, entrepreneurship, market structure, and firm strategy.

Studies in econometrics have revealed that the Gross Domestic Expenditure on Research and Development (GERD) in comparison to the GDP is a significant driver of economic growth,

measured through GDP per capita (Coccia, 2008). One of the primary indicators of science and technology policy, GERD is presented as a percentage of GDP, as stated by Bernal (1939). This indicator includes public investment in research, which usually leads to basic research that generates long-term knowledge advantageous to the country. For instance, in the pharmaceutical industry, private companies and foundations invest more in basic research than governments do (Cockburn and Henderson, 1998).

It is possible to evaluate the experience of university-industry collaboration experiments, which have been ongoing for almost 20 years, and make predictions about future collaborations based on their actual interests. The first phase of policy research in this area examined the structural coordination required for successful cooperation between universities and industry, with studies on companies that have already collaborated successfully (Blumenthal, Gluck, Louis, Stoto, and Wise, 1986; Campbell and Slaughter, 1995; Cohen, Florida, and Goe, 1994; Lee, 1995; Matkin, 1994; Rahm, 1995). Some studies have identified a list of factors and expectations that are considered "personally significant" to university faculty members when initiating collaborative research with industry, based on the literature (Lee, 1996; Roessner, Lee, Shapira, and Bozeman, 1996). The reasons for collaboration are to gain knowledge in the field of research, supplement scientific research funding, and test research theories and their practical application.

Ge and Chen (2008) found that joint venture deals have a considerable impact on economic growth. According to the study, joint ventures help with technology transfer, knowledge sharing, and market access, which ultimately contribute to productivity enhancement and overall economic development. In their study, Ko, Kim, Lee, and Song (2020) examined the relationship between strategic alliance deals and firm performance. The research concluded that firms engaged in strategic alliances experience improved innovation capabilities, expanded market reach, and enhanced financial performance, which leads to a sustained competitive advantage and value creation. Girma, Gong, and Görg (2008) investigated the role of joint ventures in facilitating technology transfer. The study suggests that joint ventures provide platforms for knowledge exchange, skill development, and technology diffusion between partner firms, leading to improved productivity and competitiveness. Mohr, Fisher, and Nevin (1996) analyzed the impact of strategic alliances on market expansion. The study found that firms entering into strategic alliances can leverage partner networks, resources, and expertise

to access new markets, penetrate existing markets more effectively, and achieve sustainable growth. Hagedoorn (1993) delved into the role of joint ventures in risk sharing. The study suggests that joint ventures enable firms to pool resources, share investment costs, and mitigate risks associated with market uncertainty, enhancing their ability to pursue high-risk, high-return opportunities. Kale, Singh, and Perlmutter (2000) explored the relationship between strategic alliances and competitive advantage. The study concluded that firms engaged in strategic alliances can access complementary resources, capabilities, and markets, leading to enhanced competitiveness and improved financial performance. All the studies mentioned above highlight the crucial role of joint ventures and strategic alliance deals in promoting economic growth, enhancing firm performance, facilitating technology transfer, expanding market reach, sharing risks, and building competitive advantage.

Hall, Jaffe, and Trajtenberg (2000) conducted research that highlights the importance of patent families for economic growth and innovation. According to their study, countries with more patent families relative to GDP tend to experience greater technological advancements, productivity improvements, and long-term economic prosperity. In another study by Harhoff, Narin, Scherer, and Vopel (1999), the relationship between patent families and technological development was examined. The researchers found that patent families serve as indicators of knowledge diffusion across borders, collaboration, and technology transfer, which contribute to global innovation networks and drive technological progress. Criscuolo, Haskel, and Slaughter (2010) explored the correlation between patent families and firm competitiveness. Their study suggests that companies with more patent families relative to GDP are more likely to innovate, differentiate their products, and gain a competitive edge in the global marketplace, driving economic growth and industry dynamism. Jaffe, Trajtenberg, and Henderson (1993) studied the role of patent families in facilitating knowledge spillovers and technology diffusion. Their research suggests that patent families serve as channels for disseminating technological know-how, fostering collaboration, and stimulating innovation ecosystems that drive economic development and prosperity. Dechezleprêtre, Martín, and Mohnen (2014) explored the relationship between patent families and entrepreneurship. Their findings suggest that countries with more patent families relative to GDP exhibit a vibrant entrepreneurial ecosystem, characterized by high levels of innovation, startup activity, and economic dynamism. Collectively, these studies emphasize the critical importance of patent families as a percentage

of GDP in driving innovation, technological development, firm competitiveness, economic resilience, knowledge spillovers, and entrepreneurship.

Niebel (2018) conducted research that highlights the significant influence of ICT services imports on economic growth. The study indicates that countries that import a higher proportion of ICT services relative to total trade tend to experience enhanced productivity, innovation, and competitiveness, leading to sustained economic growth. (Vu, 2011) investigated the effect of ICT services imports on export competitiveness. The study suggests that countries importing a higher proportion of ICT services relative to total trade have a better chance of improving export competitiveness by enhancing production efficiency, market access, and product quality. A study by (Bénassy-Quéré, Fontagné, and Lahrière-Révil, 2005) explored the link between ICT services imports and industrial structure. They found that countries with a higher proportion of ICT services imports in total trade tend to undergo structural transformation towards knowledge-based industries, promoting innovation and economic diversification. Collectively, these studies highlight the critical importance of ICT services imports as a percentage of total trade in driving economic growth, trade balance improvements, export competitiveness, industrial transformation, productivity growth, and technological adoption.

Eaton and Kortum (1994) conducted research that emphasizes how domestic patents have a significant impact on economic development. The study suggests that countries that have higher levels of domestic patents in relation to GDP tend to experience greater innovation, technological progress, and long-term economic growth. Keller (2004) conducted a study that examined the role of foreign patents in technological transfer. They found that foreign patents contribute significantly to technological diffusion, innovation, and productivity enhancement when adopted and adapted domestically, thereby driving economic development. Aghion, Blundell, Griffith, Howitt, and Prantl (2004) conducted research that explores the relationship between domestic patents and industrial competitiveness. The study suggests that countries with a higher number of domestic patents relative to GDP are better positioned to compete globally since patents serve as indicators of technological prowess, innovation capacity, and industrial dynamism. Branstetter, Fisman, and Foley (2006) conducted a study that investigated the impact of foreign patents on economic integration. They found that access to foreign patents through trade, licensing, and technology transfer agreements facilitates economic integration, knowledge spillovers, and innovation diffusion, promoting economic convergence and growth.

Mohnen and Hoareau (2003) analyzed the relationship between domestic patents and human capital development. They found that countries with a higher number of domestic patents relative to GDP tend to invest more in education, research, and skill development, fostering human capital accumulation and innovation capacity. Grossman and Helpman (1993) conducted a study that examined the role of foreign patents in economic catch-up. They found that access to foreign patents through trade and technology transfer agreements enables latecomer countries to leapfrog technological barriers, accelerate industrialization, and catch up with advanced economies. These studies collectively underscore the critical importance of patents by origin as a percentage of GDP in driving economic development, technological transfer, industrial competitiveness, economic integration, human capital development, and economic catch-up.

Oulton (2012) highlights labor productivity growth's critical role in economic performance. According to the study, consistent increases in labor productivity lead to overall economic growth, improved living standards, and enhanced competitiveness by allowing firms to produce more output with fewer resources. Van Ark, O'Mahony, and Timmer (2008) examines the correlation between labor productivity growth and firm competitiveness. Their research shows that companies with higher labor productivity growth rates tend to outperform their competitors, achieve cost efficiencies, and gain market share, leading to sustained profitability and business success. Jorgenson and Timmer (2011) emphasizes the connection between labor productivity growth and innovation. The study suggests that technological advancements, organizational improvements, and investment in human capital are responsible for driving labor productivity growth, which in turn fosters innovation-led economic growth and structural transformation. Gordon (2012) investigates the relationship between labor productivity growth and wage growth. The findings suggest that sustained increases in labor productivity lead to higher wages, improved living standards, and reduced income inequality, as workers' contributions to economic output increase over time. Timmer, Dietzenbacher, Los, Stehrer, and De Vries (2015) research explores the role of labor productivity growth in economic resilience. The study suggests that economies with higher labor productivity growth rates are more resilient to external shocks, as productivity gains enhance adaptive capacity, resource allocation efficiency, and structural flexibility. Greenwood, Hercowitz, and Krusell (1997) examines the correlation between labor productivity growth and technological change. The findings suggest that technological innovation drives long-term labor productivity growth by enabling firms to

adopt more efficient production methods, introduce new products, and expand market opportunities. In summary, these studies collectively underscore the critical importance of labor productivity growth in driving economic performance, firm competitiveness, innovation, wage growth, economic resilience, and technological change.

Research conducted by Corbett, Montes-Sancho, and Kirsch (2005) showcases the substantial effect of ISO 9001 certificates on economic progress. According to the study, firms that have been certified with ISO 9001 tend to have higher levels of productivity, efficiency, and customer satisfaction, which leads to improved competitiveness and economic growth. In a study carried out by Han, Chen, and Ebrahimpour (2007), the correlation between ISO 9001 certificates and export competitiveness was analyzed. The study found that businesses certified with ISO 9001 are more likely to comply with international quality standards, gain access to global markets, and enhance export competitiveness, which leads to an increase in trade flows and economic benefits. Samson and Terziovski (1999) explores the influence of ISO 9001 certificates on firm innovation. They found that ISO 9001 certification encourages firms to implement quality management practices, adopt continuous improvement initiatives, and invest in innovation, leading to enhanced product quality, process efficiency, and market competitiveness. Chen, Wu, and Zhai (2019) examines the significance of ISO 9001 certificates in attracting investment. They discovered that ISO 9001 certification indicates a firm's commitment to quality, reliability, and customer satisfaction, which enhances its reputation, credibility, and attractiveness to investors, stimulating economic growth and development. Melnyk, Sroufe, and Calantone (2003) conducted research in 2003 that investigates the influence of ISO 9001 certificates on small business growth. They found that ISO 9001 certification helps small businesses improve operational efficiency, standardize processes, and enhance customer satisfaction, paving the way for sustainable growth and economic contribution. Heras, Casadesús, and Dick (2002) explores the relationship between ISO 9001 certificates and public sector performance. They found that ISO 9001 certification enhances service quality, efficiency, and accountability in public sector organizations, leading to enhanced citizen satisfaction, trust, and economic development. These studies collectively emphasize the critical importance of ISO 9001 quality certificates as a percentage of GDP in driving economic performance, export competitiveness, firm innovation, investment attraction, small business growth, and public sector performance.

Ginarte and Park (1997) research highlights the considerable influence of intellectual property receipts on economic growth. The study indicates that countries with greater intellectual property receipts relative to overall trade tend to experience more innovation, technology transfer, and economic development. Maskus and Penubarti (1995) study examines the relationship between intellectual property receipts and export competitiveness. They conclude that countries with higher intellectual property receipts as a percentage of total trade are better positioned to compete globally because they own valuable intellectual assets that boost product differentiation and market access. Keller (2004) research emphasizes the connection between intellectual property receipts and technological innovation. The study suggests that countries receiving higher intellectual property receipts relative to total trade are more likely to innovate, invest in R&D, and adopt advanced technologies, contributing to economic growth and competitiveness. Yang and Maskus (2001) investigation explores the role of intellectual property receipts in attracting FDI. They find that countries with higher intellectual property receipts as a share of total trade are more attractive destinations for FDI, as intellectual property protection signals a conducive environment for innovation and technology-driven industries. Park and Ginarte (1997) research investigates the relationship between intellectual property receipts and economic diversification. They find that countries with higher intellectual property receipts as a percentage of total trade tend to diversify their economies into high-value-added sectors, such as technology, pharmaceuticals, and creative industries. Overall, these studies collectively underscore the essential importance of intellectual property receipts as a percentage of total trade in driving economic growth, export competitiveness, technological innovation, FDI attraction, economic diversification, and economic resilience.

Important factors for the growth of high-tech exports in EU countries are the size of public and private research expenditures, the number of human resources engaged in knowledge-intensive activities, and the inclination to participate in international commerce. According to the results of the econometric analysis, there is a causal relationship between the independent variable and the level of high-tech exports to the EU. The study also reveals that there is a positive correlation between high-tech exports and total R&D expenditure, although the relationship varies by country. Private R&D spending has a stronger impact on high-tech exports than government R&D spending, according to the findings of Sandu and Ciocanel (2014). Even though R&D results can benefit high-tech exports, viable institutions are required to mediate and facilitate

the application of research results in high-tech production, as concluded by Braunerhjelm and Thulin (2008).

Erumban and Das (2016) study investigates the correlation between ICT services exports and trade balance, concluding that countries with a higher proportion of ICT services exports within their total trade tend to have more favorable trade balances. This is because ICT exports contribute significantly to export revenue and economic stability. Ramasamy, Yeung, and Laforet (2012) explore how exporting ICT services can impact a firm's competitiveness. By leveraging ICT capabilities to enhance productivity and market reach, firms engaged in exporting ICT services tend to be more innovative, efficient, and competitive. Forman and Goldfarb (2020) examines the relationship between ICT services exports and FDI, with countries that have a strong ICT services export sector attracting more FDI. This is because multinational corporations seek access to skilled labor, advanced technologies, and global market opportunities. De Mel and Jayaratne (2012) study investigates the contributions of ICT services exports to economic development. Countries leveraging ICT services exports as a strategic growth driver experience diversified economic activities, increased employment opportunities, and enhanced living standards. These studies collectively underscore the critical importance of ICT services exports in driving economic growth, trade balance improvements, firm competitiveness, technological innovation, FDI attraction, and economic development.

Lipsey (2004) demonstrates the significant effect of FDI net outflows on domestic investment. The study suggests that countries experiencing FDI net outflows might redirect capital towards domestic investment opportunities, which could stimulate economic growth and development. Cheng (2019) examines the link between FDI net outflows and technology transfer. They suggest that countries engaging in FDI net outflows may gain advanced technologies and managerial expertise from foreign subsidiaries, which could enhance domestic innovation capabilities and industrial competitiveness. Alfaro, Chanda, Kalemli-Ozcan, and Sayek (2006) investigates the role of FDI net outflows in portfolio diversification. The study argues that emerging market multinational corporations engaging in FDI net outflows might diversify their investment portfolios, reducing risk exposure and enhancing long-term financial stability. Lall and Narula (2004) analyses the role of FDI net outflows in economic diplomacy. They argue that countries engaging in FDI net outflows might use outward investment as a tool for enhancing political influence, fostering strategic alliances, and securing access to critical

resources and markets. Gereffi, Humphrey, and Sturgeon (2005) examines the relationship between FDI net outflows and global value chains. The study suggests that countries engaging in FDI net outflows might participate in global value chains by establishing overseas production facilities, sourcing inputs, and accessing new markets, thereby enhancing economic integration and competitiveness. Hirsch and Nov (2004) explores the implications of FDI net outflows for economic development strategy. They suggest that countries engaging in FDI net outflows might pursue strategic objectives such as market-seeking, resource-seeking, or efficiency-seeking investments, aligning outward investment flows with national development priorities. Together, these studies highlight the many dimensions of FDI net outflows and their potential implications for domestic investment, technology transfer, portfolio diversification, economic diplomacy, global value chains, and economic development strategy.

Keller and Lehmann (2006) emphasizes the impact of the origin of trademarks on the value of a brand, suggesting that trademarks from reputable and established markets are associated with higher brand premiums due to consumer perceptions of quality, reliability, and prestige. Research by Blonigen and Haynes (1999) explores the connection between trademark origin and export performance. They discovered that trademarks originating from countries with strong brand reputations are more likely to facilitate export market entry, expansion, and penetration, driving export growth and economic success. Head and Mayer (2004) study investigates the role of trademark origin in attracting FDI. They found that trademarks originating from countries with strong brand equity and market presence may enhance FDI attractiveness by signaling a favorable business environment and reducing information asymmetry. Luo and Bhattacharya (2006) explores the role of trademark origin in guiding market expansion strategies. They found that trademarks originating from countries with favorable brand perceptions may facilitate market entry and acceptance, enabling firms to penetrate new markets and achieve sustainable growth. Overall, these studies highlight the crucial significance of trademarks by origin in influencing brand value, market competitiveness, export performance, FDI attraction, consumer preferences, and market expansion strategies as a percentage of GDP.

Brynjolfsson and Hitt (2000) demonstrates the significant influence of ICTs on the development of organizational models. The study suggests that ICTs allow organizations to create fresh business models, simplify procedures, and improve productivity through digitization,

automation, and data-based decision-making. Sambamurthy, Bharadwaj, and Grover (2003) examines the role of ICTs in promoting agile organizational models. They observe that ICTs enable firms to adapt rapidly to changing market conditions, customer preferences, and competitive dynamics by fostering teamwork, flexibility, and innovation. Orlikowski and Iacono (2001) explore the process of ICT-enabled organizational transformation. They establish that ICTs play a fundamental role in transforming organizational structures, processes, and cultures, allowing for continuous adaptation, learning, and innovation. Powell, Koput, and Smith-Doerr (1996) investigates the correlation between ICTs and networked organizational models. They observe that ICTs facilitate the emergence of networked organizations characterized by decentralized decision-making, virtual collaboration, and agile resource allocation, empowering firms to effectively utilize external expertise and resources. Malone, Yates, and Benjamin (1989) explores the idea of ICT-enabled virtual organizational models. They find that ICTs allow for the establishment of virtual organizations that transcend geographical boundaries, bringing together dispersed individuals and resources to collaborate on projects and initiatives effectively. Chesbrough (2006) examines the role of ICTs in fostering open innovation models. They find that ICTs enable firms to access external sources of knowledge, ideas, and expertise, encouraging collaboration, co-creation, and value capture in innovation ecosystems. These studies collectively emphasize the critical importance of ICTs in shaping organizational model creation by enabling innovation, agility, transformation, networked collaboration, virtualization, and open innovation.

Numerous national feature films are produced across various cultural industries. The production of such films involves certain constraints due to the need for economic supply of relatively standardized lifetime products in a volatile market that seeks innovation. Additionally, group composition presents limitations to creativity, as differing artistic values, personal characteristics, and expectations of team members can hinder the ability to generate innovative ideas (Caves, 2000). To enhance innovation, Perretti and Negro (2007) proposed two hypotheses, suggesting that the inclusion of newcomers to teams or a combination of new and old members could have a positive impact. This argument is based on the idea that genre innovation is a key measure of innovation in the feature film industry in the United States.

Throsby (2010) examines the diversity of the entertainment and media market and its cultural impact, finding that a diverse entertainment and media market catering to different tastes,

interests, and demographics within the population contributes to cultural enrichment, social cohesion, and creative expression. Florida (2002) explores the relationship between the entertainment and media market and urbanization, concluding that urban centers with a vibrant entertainment and media sector in proportion to the population size attract creative talent, cultural amenities, and investment, thereby driving urban growth and development. Hesmondhalgh and Baker (2013) conducted research on the correlation between cultural diversity and the size of the entertainment and media market. The study aimed to explore how a wide variety of media offerings, targeting different demographic segments in society, contributes to cultural enrichment and pluralism. Norris (2000) carried out a study that examined the impact of the entertainment and media market on political discourse and public opinion. The research investigated how media ownership concentration, along with market size, affects the range of perspectives and quality of democratic deliberation in society. Bennett (2000) conducted research on the role of the entertainment and media market in promoting social cohesion. The study analyzed how shared cultural experiences and media consumption patterns create a sense of community, identity, and belonging among diverse population groups. Together, these studies emphasize the critical importance of the entertainment and media market as a percentage of the population in driving economic growth, technological innovation, consumer behavior, cultural impact, social development, and urbanization.

One of the types of TLDs that are maintained by the Internet Designated Numbers Authority for use on the Internet is known as a generic TLD. Generic TLDs can be either unlimited (such as .com, .info, .net, and .org) or restricted, which means that they can only be used if certain eligibility criteria are met (such as .biz, .name, and .pro). Country-code TLDs are two-letter domain names that are specifically designated for use by a particular economy, country, or autonomous territory. The use of domain names has played a critical role in the development and popularization of the Internet. According to Duchesne (2012), website names have become a part of everyday discourse, evolving from their original purpose of directing computers to their proper destinations and acquiring brand identifiers. Domain names have made the Internet accessible to everyone, especially in a socioeconomic context where they represent a commodity of significant monetary value (Komaitis, 2010).

Kittur, Chi, and Suh (2009) conducted a study to examine the factors that motivate Wikipedia edits. The investigation identified self-interest, altruism, and a desire for recognition as key

drivers that influence editing behavior on the platform. Halfaker, Geiger, Morgan, and Riedl (2013) carried out research to evaluate the effectiveness of quality control mechanisms on Wikipedia. The study analyzed the impact of human editors and automated bots in maintaining the accuracy and reliability of information within the platform. Yasseri, Sumi, Rung, Kornai, and Kertész (2012) conducted research to explore temporal patterns in Wikipedia editing activity. The study analyzed editing behavior over time to identify variations in contribution rates, topical interests, and collaborative dynamics within the community. Graham, Hogan, Straumann, and Medhat (2014) conducted a study to examine the geographic patterns in Wikipedia editing. The research investigated the distribution of editors across various regions and explored how local context and cultural factors influence editing behavior and content creation. Kittur, Suh, Pendleton, and Chi (2007) conducted a study to investigate the collaborative dynamics in Wikipedia editing. The research analyzed coordination patterns, conflict resolution, and knowledge integration among contributors to understand how collective intelligence emerges within the editing community. These studies provide valuable insights into different aspects of Wikipedia editing behavior, contributing to our understanding of collaboration, quality control, bias, and temporal dynamics within the platform.

The impact of mobile app development on the economy was explored by Gao and Bai (2014). They investigated how the proliferation of mobile apps contributes to job creation, innovation, and revenue generation within the digital economy. The relationship between mobile app development and user engagement was studied by Lee, Kozar, and Larsen (2003). The research explored strategies for enhancing user experience, retention, and monetization through effective app design and development practices. Qiang, Kuek, Dymond, and Esselaar (2012) investigated the contribution of mobile app development to economic growth. The study analyzed how mobile apps stimulate productivity, efficiency, and competitiveness across various sectors, driving overall economic development. West and Mace (2010) investigated the role of mobile app development in fostering innovation ecosystems. The research explored how mobile platforms enable developers to create and distribute innovative apps, driving ecosystem growth and dynamism. These studies provide valuable insights into the economic, entrepreneurial, user-centric, inclusive, and innovation-driven aspects of mobile app development.

In Table 4.11 below, studies in the literature regarding the selected parameters that emphasize the importance of these parameters are summarized.

Table 4.11. The Related Studies of the Selected Parameters

Selected Parameters	Related Studies	Summary
C1.1.1. Political and operational stability	(Van de Walle, 1998) (Acar, 2019; Li et al., 2020; Laura T. Schelhas et al., 2019; Laura T Schelhas et al., 2019; Singha and Singh, 2022; Sui et al., 2021)	These studies highlight the crucial role of political and operational stability in driving economic growth and organizational success.
C1.2.1. Regulatory quality	(Bank, 2001; Jacobzone et al., 2007; Policy, 2011)	These studies emphasize the importance of regulatory quality in fostering private sector growth and economic advancement.
C1.2.2. Rule of law	(Iglesias and Ordenes, 2019; Postema, 2023; Sellers, 2014; Shaikh and Malik, 2020)	These studies highlight the critical role of the rule of law in promoting stability, justice, and economic progress.
C1.3.1. Ease of starting a business	(Ayyagari et al., 2011; Bétila, 2021; Djankov et al., 2002; Haltiwanger et al., 2013; Klapper and Love, 2011)	These studies provide evidence that simplifying business startup procedures positively impacts economic growth, firm entry rates, and job creation.
C1.3.2. Ease of resolving insolvency	(Beck et al., 2008; Berkowitz et al., 2003; Claessens and Klapper, 2005; Porta et al., 2008; World-Bank, 2016)	These studies emphasize the crucial role of efficient insolvency resolution mechanisms in enhancing economic efficiency, credit access, entrepreneurship, firm dynamics, and investor confidence.
C2.1.4. PISA scales in reading, maths and science	(Chetty et al., 2014; Hanushek and Woessmann, 2023; Hanushek and Woessmann, 2008; Heckman and LaFontaine, 2010; Helliwell and Huang, 2014)	These studies underscore the significant impact of PISA scores on economic growth, social mobility, skill formation, and global competitiveness, highlighting the importance of educational quality in shaping a nation's future.
C2.2.1. Tertiary enrolment, % gross	(Becker, 2009; Bils, 2000; Blanden and Machin, 2004; Lutz and Kc, 2010; Oreopoulos et al., 2006; Psacharopoulos, 1994)	These studies highlight the diverse benefits of increasing tertiary enrollment rates, including economic development, improved labor market outcomes, social mobility, innovation, gender equality, and economic resilience.
C2.3.1. Researchers, FTE/mn pop.	(Adams, 1990; Bloom et al., 2020; De la Fuente and Ciccone, 2003; Jaffe, 1989; Phelan, 1994; Stephan, 2012)	These studies collectively emphasize the critical role of researcher density in driving innovation, economic growth, knowledge spillovers, firm performance, scientific output, and social progress.
C2.3.3. Global corporate R&D investors, top 3, mn US\$	(Archibugi and Coco, 2004; Chesbrough and Appleyard, 2007; Griliches, 1998; Hall and Lerner, 2010; Hall and Ziedonis, 2001)	These studies collectively highlight the significant impact of top global corporate R&D investors on innovation, industry competitiveness, technological leadership, economic development, firm performance, and global technological trends.
C3.1.1. ICT access	(Bhatnagar, 2000; Gilbert et al., 2004; Grant and Chau, 2005; Lam, 2005; Tung and Rieck, 2005; West, 2005; Zhang et al., 2005)	Governments worldwide are adopting ICT for eGovernment, offering benefits like efficiency, accessibility, accountability, transparency, cost savings, and improved interactions with citizens and businesses.
C3.2.1. Electricity output, GWh/mn pop.	(Dogan and Deger, 2016; Sovacool, 2014; Sovacool and Mukherjee, 2011)	These studies collectively highlight the multifaceted importance of electricity output per capita, spanning economic

		development, environmental sustainability, industrialization, energy security, and urbanization.
C3.2.2. Logistics performance	(Arvis et al., 2007; Djankov et al., 2010; Jaramillo et al., 2018; Limao and Venables, 2001; Mentzer et al., 2008; Peck and Christopher, 2004)	These studies collectively underscore the critical role of logistics performance in driving economic competitiveness, firm productivity, trade facilitation, export growth, supply chain resilience, and customer satisfaction.
C3.3.2. Environmental performance	(Ambec and Lanoie, 2008; Horbach, 2008; Martinez-Alier, 2003; Metz et al., 2014; Prüss-Üstün et al., 2016; Stern, 2004)	These studies collectively underline the critical importance of environmental performance in driving economic growth, corporate profitability, regulatory compliance, public health, social equity, and climate change mitigation efforts.
C4.1.1. Ease of getting credit	(Bayraktar, 2015; MogesEbero and Begum, 2016; Morris and Aziz, 2011; Piwonski, 2010; Singh, 2015)	Credit indicators and improved business behavior are crucial factors that can attract higher levels of foreign direct investment (FDI).
C4.1.2. Domestic credit to private sector	(Ahmed, 2008; Akpansung and Babalola, 2011; Aliero et al., 2013; Hamida Begum and Aziz, 2019; Majeed and Ahmad, 2009; Pradhan, 2009)	Private sector domestic credit plays a pivotal role in economic development, contributing to increased investment, employment, productivity, and poverty reduction.
C4.2.1. Ease of protecting minority investors	(Claessens et al., 2000; Djankov et al., 2008; Doidge et al., 2007; La Porta et al., 1999; La Porta et al., 2002; Porta et al., 1998)	These studies collectively highlight the critical importance of protecting minority investors for investor confidence, economic development, corporate governance, market liquidity, foreign investment, and firm performance.
C4.2.2. Market capitalization, % GDP	(Beck et al., 2000; Bekaert et al., 2005; Claessens and Yurtoglu, 2013; Demirguc-Kunt, 2008; Levine and Zervos, 1998; Merton, 1992)	These studies collectively highlight the critical role of market capitalization as a percentage of GDP in fostering economic development, financial stability, investor confidence, capital formation, financial inclusion, and equitable wealth distribution.
C4.3.2. Intensity of local competition	(Audretsch and Feldman, 1996; Berry et al., 1995; Caves, 1998; Porter, 2011; Tirole, 1988)	These studies collectively emphasize the critical importance of local competition intensity in driving firm innovation, consumer benefits, firm efficiency, entrepreneurship, market structure, and firm strategy.
C5.1.3. GERD performed by business, % GDP	(Bernal, 1939; Coccia, 2008; Cockburn and Henderson, 1998)	GERD as a percentage of GDP is a significant driver of economic growth, particularly through its support for basic research and innovation.
C5.2.1. University-industry R&D collaboration	(Blumenthal et al., 1986; Campbell and Slaughter, 1995; Cohen et al., 1994; Lee, 1995; Lee, 1996; Matkin, 1994; Rahm, 1995; Roessner et al., 1996)	Research on university-industry collaboration over nearly two decades has highlighted the structural coordination needed for successful partnerships and identified key factors that are personally significant to university faculty initiating these collaborations.
C5.2.4. Joint venture/strategic alliance deals/bn PPP\$ GDP	(Ge and Chen, 2008; Girma et al., 2008; Hagedoorn, 1993; Kale et al., 2000; Ko et al., 2020; Mohr et al., 1996)	The studies collectively highlight the significant role of joint ventures and strategic alliances in promoting economic growth, enhancing firm performance,

		facilitating technology transfer, expanding market reach, sharing risks, and building competitive advantage.
C5.2.5. Patent families/bn PPP\$ GDP	(Criscuolo et al., 2010; Dechezleprêtre et al., 2014; Hall et al., 2000; Harhoff et al., 1999; Jaffe et al., 1993)	The studies collectively underscore the critical importance of patent families as a percentage of GDP in driving innovation, technological development, firm competitiveness, economic resilience, knowledge spillovers, and entrepreneurship.
C5.3.3. ICT services imports, % total trade	(Bénassy-Quéré et al., 2005; Niebel, 2018; Vu, 2011)	The studies collectively highlight the critical importance of ICT services imports as a percentage of total trade in driving economic growth, trade balance improvements, export competitiveness, industrial transformation, productivity growth, and technological adoption.
C6.1.1. Patents by origin/bn PPP\$ GDP	(Aghion et al., 2004; Branstetter et al., 2006; Eaton and Kortum, 1994; Grossman and Helpman, 1993; Keller, 2004; Mohnen and Hoareau, 2003)	The studies collectively emphasize how patents, both domestic and foreign, play crucial roles in driving economic development through innovation, technological progress, and industrial competitiveness.
C6.2.1. Labor productivity growth, %	(Gordon, 2012; Greenwood et al., 1997; Jorgenson and Timmer, 2011; Oulton, 2012; Timmer et al., 2015; Van Ark et al., 2008)	These studies collectively underscore the critical importance of labor productivity growth in driving economic performance, firm competitiveness, innovation, wage growth, economic resilience, and technological change.
C6.2.4. ISO 9001 quality certificates/bn PPP\$ GDP	(Chen et al., 2019; Corbett et al., 2005; Han et al., 2007; Heras et al., 2002; Melnyk et al., 2003; Samson and Terziovski, 1999)	These studies collectively emphasize the critical importance of ISO 9001 quality certificates in driving economic performance, export competitiveness, firm innovation, investment attraction, small business growth, and public sector performance.
C6.3.1. Intellectual property receipts, % total trade	(Ginarte and Park, 1997; Keller, 2004; Maskus and Penubarti, 1995; Park and Ginarte, 1997; Yang and Maskus, 2001)	These studies collectively underscore the essential importance of intellectual property receipts as a percentage of total trade in driving economic growth, export competitiveness, technological innovation, FDI attraction, economic diversification, and economic resilience.
C6.3.2. High-tech exports	(Braunerhjelm and Thulin, 2008; Sandu and Ciocanel, 2014)	Factors crucial for boosting high-tech exports in EU countries include substantial public and private research expenditures, a robust workforce engaged in knowledge-intensive activities, and active participation in international trade, supported by evidence linking these factors to increased high-tech export levels.
C6.3.3. ICT services exports	(De Mel and Jayaratne, 2012; Erumban and Das, 2016; Forman and Goldfarb, 2020; Ramasamy et al., 2012)	ICT services exports play a crucial role in enhancing trade balances, firm competitiveness, technological innovation, FDI attraction, and overall economic development, as evidenced by recent studies in the field.

C6.3.4. Foreign direct investment net outflows	(Alfaro et al., 2006; Cheng, 2019; Gereffi et al., 2005; Hirsch and Nov, 2004; Lall and Narula, 2004; Lipsey, 2004)	FDI net outflows can influence domestic investment, technology transfer, portfolio diversification, economic diplomacy, participation in global value chains, and economic development strategy, as indicated by recent research.
C7.1.1. Trademarks by origin/bn PPP\$ GDP	(Blonigen and Haynes, 1999; Head and Mayer, 2004; Keller and Lehmann, 2006; Luo and Bhattacharya, 2006)	The impact of trademark origin on brand value, export performance, FDI attraction, and market expansion strategies underscores the importance of consumer perceptions regarding quality, reliability, and prestige associated with trademarks from reputed markets.
C7.1.3. ICTs and organizational model creation	(Brynjolfsson and Hitt, 2000; Chesbrough, 2006; Malone et al., 1989; Orlikowski and Iacono, 2001; Powell et al., 1996; Sambamurthy et al., 2003)	These studies collectively underscore the critical role of ICTs in transforming organizational models by enabling innovation, agility, adaptation to market changes, networked collaboration, virtualization across geographical boundaries, and fostering open innovation ecosystems.
C7.2.2. National feature films/mn pop.	(Caves, 2000; Perretti and Negro, 2007)	National feature film production in cultural industries encounters economic constraints and group composition challenges that hinder creativity, with strategies like integrating newcomers or mixing experienced and new team members suggested to foster innovation.
C7.2.3. Entertainment and media market/th pop.	(Bennett, 2000; Florida, 2002; Hesmondhalgh and Baker, 2013; Norris, 2000; Throsby, 2010)	The entertainment and media market, catering to diverse tastes and demographics, plays a crucial role in cultural enrichment, social cohesion, urban growth, and fostering democratic discourse.
C7.3.1. Generic top-level domains (TLDs)/th pop.	(Duchesne, 2012; Komaitis, 2010)	Domain names, including generic and country-code TLDs, serve a crucial role in Internet navigation and branding, evolving from technical identifiers to valuable commodities integral to global accessibility and socioeconomic dynamics.
C7.3.2. Country-code TLDs/th pop.	(Duchesne, 2012; Komaitis, 2010)	Country-code TLDs play a vital role in navigating the Internet and establishing brands, transitioning from mere technical designations to valuable assets that enhance global accessibility and influence socioeconomic trends.
C7.3.3. Wikipedia edits/mn pop.	(Graham et al., 2014; Halfaker et al., 2013; Kittur et al., 2009; Kittur et al., 2007; Yasseri et al., 2012)	Research on Wikipedia editing behavior has identified motivations like self-interest, altruism, and recognition, alongside studies on quality control mechanisms, temporal editing patterns, geographic editing distributions, and collaborative dynamics within the platform.
C7.3.4. Mobile app creation/bn PPP\$ GDP	(Gao and Bai, 2014; Lee et al., 2003; Qiang et al., 2012; West and Mace, 2010)	Studies on mobile app development highlight its economic impact, including job creation, innovation, revenue

		generation, user engagement strategies, economic growth stimulation, productivity enhancement, sector competitiveness, and fostering of innovation ecosystems.
--	--	--

The feature selection process is a crucial step in any study that aims to achieve accurate results and conclusions. One of the indicators of the accuracy of the findings and results of the study is the parallelism and compatibility of the selected parameters with the existing literature on the subject. In other words, the relevance and appropriateness of the selected features should be evaluated based on their correspondence with the parameters highlighted in the literature. By ensuring that the parameters selected for the study align with the existing literature, the validity and reliability of the findings are strengthened. This is because the literature serves as a benchmark for the study, and the selected features that are highlighted by the literature can be considered as appropriate and relevant for the study. Therefore, the parallelism and compatibility of the selected parameters with the existing literature is an important issue that should not be overlooked in any study. It provides a strong foundation for the study, as well as a basis for comparison with other studies in the same field. In conclusion, the selection of appropriate parameters that align with the existing literature is crucial for any study that aims to achieve accurate and reliable results and conclusions. Extensive research of literature was carried out in this study to determine the parallelism and compatibility of the selected parameters with the literature. A table was prepared to summarize and explain the studies that have highlighted the significance of the selected parameters in the literature. Based on the findings, it was determined that the selected parameters are in line and compatible with the literature.

5. CONCLUSION

Connell University, INSEAD and the WIPO have jointly published the GII since 2007. The GII measures the innovation performance of 131 economies around the world, covering 93.4% of the world's population and 97.6% of global GDP (Dutta, Lanvin, and Wunsch-Vincent, 2016). The GII comprises five input sub-indices, namely institutions, human capital and research, infrastructure, market complexity and business complexity. The two sub-indices of innovation output are knowledge and technology output and creative output. The GII is calculated by taking the simple average of innovation input and output sub-indices.

Innovation plays a central role in the success of economic growth and in stimulating investment in R&D in a global economy where knowledge assets and intangible assets are increasingly important. In this study, ANN methods and various regression methods were used to analyze innovation data and predict the GII score of countries, which are ML methods. Before applying ML methods, the classification and feature selection method was applied to find out which of the 79 common parameters used to determine the GII score of countries in the 2012-2022 period is more effective on the GII score. 38 parameters were selected from 79 parameters, and these parameters were used to predict the GII score.

The neural network fitting tool was used in the ANN design, which is the first of the ML methods applied. ANN was designed using 38 parameters selected as input data and the GII score of 11 years of data as output data. Among the regression methods, one of the ML methods, 38 parameters were selected as input data, as in ANN, and the GII score of 11 years of data was used as output data. When the prediction results were compared, it was seen that the MSE value of the Levenberg-Marquart with 20 Hidden Neuron model of the ANN method and the MSE value of the Medium Gaussian SVM Regression model of the Regression method were the same as 0.001. When looking at the R values, it was determined that the ANN method with an R value of 0.9899 gave better results than the regression method with an R value of 0.9879.

This way, the score of the countries can be calculated through the ANN model created by the country score calculation method, without using all the parameters used in GII score calculation. Countries will be able to determine their strategies and address their weak points according to the estimated GII scores. ANN differs from other statistical methods in the

following aspects: It is a prediction model that allows changing the parameters used as input and seeing the place of countries in the GII ranking. Thus, countries may not need to improve all input parameters and can monitor and try to improve input parameters that will benefit them more. Two issues that contribute to both the literature and practice by revealing the strengths of the study are as follows: The first is to reduce the number of parameters by the feature selection method, and the other is that the prediction results are more accurate by covering more years than the studies in the literature. Furthermore, a prototype DSS has been proposed for calculating the GII score. This system is designed to automate the entire GII calculation process, making it easier and faster for users to input data. The prototype DSS is equipped with a user-friendly interface that allows users to input data without facing any complications. The system's features and functionality are presented in detail, which includes the automation of data input, data processing, analysis, and output of the final GII score. The prototype DSS is expected to improve the accuracy and efficiency of the GII calculation process, leading to more reliable and precise results. Additionally, potential future developments of the system, such as incorporating ML algorithms for data analysis, are discussed.

This research suggests that there should be a more significant focus on analyzing the parameters that have a significant impact on the progress of the GII. It is essential to evaluate these parameters to gain a better understanding of the key drivers of innovation performance in different countries. Moreover, the study recommends using parameter selection and ANN to achieve favorable prediction outcomes. This approach can help practitioners identify critical features that require closer scrutiny and further analysis while assessing countries' innovation performance. By identifying these features, policymakers and stakeholders can make more informed decisions to promote innovation and growth in different countries.

6. RECOMMENDATIONS

The data expansion in future studies is a key objective. By experimenting with various heuristic or hybrid methods on the data, it is believed that the results can be enhanced. The model's performance can be further improved by closely monitoring the selection times of attributes and comparing the results with other methodologies. In addition, assessing the correlation values between the chosen attributes can enhance their effectiveness. For the ANN method employed in estimation, adjusting the data split percentages can contribute to boosting model performances. Additionally, there are plans to develop a more intuitive interface for the current DSS, with the system being open to further enhancements. Additionally, by examining the impact of input parameter values on the GII score performance, the designed DSS can be refined to provide valuable guidance for researchers and policymakers.

REFERENCES

- Acar, A. (2019). The effects of political stability on economic growth of the presidential government system. *Uluslararası Ekonomi ve Siyaset Bilimleri Akademik Araştırmalar Dergisi*, 3(9), 18-31.
- Ada, Ş., & Ghaffarzadeh, M. (2015). Decision making based on management information system and decision support system. *European Researcher*(4), 260-269.
- Adams, J. D. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of political economy*, 98(4), 673-702.
- Adiyaman, G., & Hayaloglu, P. (2020). Gelişmekte olan ülkelerde inovasyonun sürdürülebilir büyüme ve kalkınmaya etkisi. *Karadeniz Ekonomi Araştırmaları Dergisi*, 1(2), 113-128.
- Aghion, P., Blundell, R., Griffith, R., Howitt, P., & Prantl, S. (2004). Entry and productivity growth: Evidence from microlevel panel data. *Journal of the European Economic Association*, 2(2-3), 265-276.
- Ahmed, A. D. (2008). Financial liberalization, financial development and growth in sub-saharan Africa's economic reform: An empirical investigation. *Centre for Strategic Economic Studies, Victoria University, Australia*.
- Akbey, F. (2014). Ar-Ge, inovasyon ve kalkınma ilişkisine yönelik bir literatür taraması: Kuramsal özet. *Maliye Dergisi*(166).
- Akpansung, A. O., & Babalola, S. J. (2011). Banking sector credit and economic growth in Nigeria: An empirical investigation. *CBN Journal of Applied Statistics*, 2(2), 51-62.
- Aksenova, S. S. (2004). Weka explorer tutorial. *School of Engineering and Computer Science California State University*, 1-44.
- Albulescu, C. T., & Drăghici, A. (2016). Entrepreneurial activity and national innovative capacity in selected European countries. *The International Journal of Entrepreneurship Innovation*, 17(3), 155-172.
- Alfaro, L., Chanda, A., Kalemli-Ozcan, S., & Sayek, S. (2006). *How Does Foreign Direct Investment Promote Economic Growth?* : Weatherhead Center for International Affairs, Harvard University.
- Aliero, H. M., Abdullahi, Y. Z., & Adamu, N. (2013). Private Sector Credit and Economic Growth Nexus in Nigeria: An Autoregressive Distributed Lag Bound Approach. *Mediterranean Journal of Social Sciences*, 4(1), 83-83.

- Alkan, M. (2014). *Giriřimcilerin giriřimcilik, inovasyon yapma, inovatif dūřünce ve inovatif giriřimcilik düzeylerinin incelenmesi*. Hasan Kalyoncu Üniversitesi,
- Altıntaş, F. F. (2020). Avrupa Birlięi űlkelerinin inovasyon etkililik, etkinlik ve verimlilik performanslarının analizi. *Turkish Studies-Social Sciences*, 15(5), 2337-2361.
- Ambec, S., & Lanoie, P. (2008). Does it pay to be green? A systematic overview. *The Academy of Management Perspectives*, 45-62.
- Archibugi, D., & Coco, A. (2004). A new indicator of technological capabilities for developed and developing countries (ArCo). *World development*, 32(4), 629-654.
- Arvis, J.-F., Alina Mustra, M., Panzer, J., Ojala, L., & Naula, T. (2007). *Connecting to compete 2007*. Retrieved from
- Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American economic review*, 86(3), 630-640.
- Ayçin, E., & Çakın, E. (2019). űlkelerin inovasyon performanslarının ölçümünde Entropi ve MABAC çok kriterli karar verme yöntemlerinin bütünleşik olarak kullanılması. *Akdeniz İİBF Dergisi*, 19(2), 326-351.
- Ayyagari, M., Demirgüç-Kunt, A., & Maksimovic, V. (2011). Firm innovation in emerging markets: The role of finance, governance, and competition. *Journal of Financial Quantitative Analysis*, 46(6), 1545-1580.
- Baker, B. M., & Ayechew, M. (2003). A genetic algorithm for the vehicle routing problem. *Computers Operations Research*, 30(5), 787-800.
- Bank, W. (2001). World development report 2002: Building Institutions for Markets. Published for the World Bank by Oxford University Press. In: Oxford and New York.
- Bashir, M. (2013). Knowledge economy index (KEI) 2012 rankings for Islamic countries and assessment of KEI indicators for Pakistan. *International Journal of Academic Research in Economics and Management Sciences*, 2(6), 28.
- Bate, A. F., Wachira, E. W., & Danka, S. (2023). The determinants of innovation performance: an income-based cross-country comparative analysis using the Global Innovation Index (GII). *Journal of Innovation Entrepreneurship*, 12(1), 20.
- Bayraktar, N. (2015). Importance of investment climates for inflows of foreign direct investment in developing countries. *Business Economic Research*, 5(1), 24-50.
- Beck, T., Demirgüç-Kunt, A., & Levine, R. (2000). A new database on financial development and structure. *World Bank economic review*, 14(3), 597-605.

- Beck, T., Demirgüç-Kunt, A., & Martinez Peria, M. S. (2008). Banking services for everyone? Barriers to bank access and use around the world. *The World Bank Economic Review*, 22(3), 397-430.
- Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*: University of Chicago press.
- Bekaert, G., Harvey, C. R., & Lundblad, C. (2005). Does financial liberalization spur growth? *Journal of financial economics*, 77(1), 3-55.
- Bénassy-Quéré, A., Fontagné, L., & Lahrèche-Révil, A. (2005). How does FDI react to corporate taxation? *International tax public finance* 12, 583-603.
- Bennett, A. (2000). *Popular music and youth culture: Music, identity and place*: Macmillan Press Ltd.
- Berkowitz, D., Pistor, K., & Richard, J.-F. J. E. e. r. (2003). Economic development, legality, and the transplant effect. 47(1), 165-195.
- Berman, S. (2013). Ideational Theorizing in the Social Sciences since “Policy Paradigms, Social Learning, and the State”. *Governance*, 26(2), 217-237.
- Bernal, J. (1939). *The social function of science* The MIT Press Cambridge.
- Berry, A., Levinsohn, J., & Pakes, A. (1995). Automobile Prices In Market Equilibrium. *Econometrica*, 63(4), 841-890.
- Bétila, R. R. (2021). The impact of Ease of Doing Business on economic growth: a dynamic panel analysis for African countries. *SN Business & Economics*, 1(10), 144. doi:10.1007/s43546-021-00143-9
- Bhatnagar, S. (2000). Social implications of information and communication technology in developing countries: Lessons from Asian success stories. *The Electronic Journal of Information Systems in Developing Countries*, 1(1), 1-9.
- Bils, M. (2000). [How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws]: Comment. *NBER macroeconomics annual*, 15, 59-68.
- Bishop, C. (1995). *Neural Networks for Pattern Recognition* (Vol. 2): Clarendon Press.
- Blanden, J., & Machin, S. (2004). Educational inequality and the expansion of UK higher education. *Scottish journal of political economy*, 51(2), 230-249.
- Blonigen, B., & Haynes, S. E. (1999). Antidumping investigations and the pass-through of exchange rates and antidumping duties. In: National Bureau of Economic Research Cambridge, Mass., USA.

- Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4), 1104-1144.
- Blumenthal, D., Gluck, M., Louis, K. S., Stoto, M. A., & Wise, D. J. S. (1986). University-industry research relationships in biotechnology: Implications for the university. *Science*, 232(4756), 1361-1366.
- Branstetter, L. G., Fisman, R., & Foley, C. F. (2006). Do stronger intellectual property rights increase international technology transfer? Empirical evidence from US firm-level panel data. *The quarterly Journal of economics*, 121(1), 321-349.
- Braunerhjelm, P., & Thulin, P. (2008). Can countries create comparative advantages? R&D expenditures, high-tech exports and country size in 19 OECD countries, 1981–1999. *International Economic Journal*, 22(1), 95-111.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24, 123-140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Brownlee, J. (2019). How to choose a feature selection method for machine learning. *Machine Learning Mastery*, 10.
- Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of economic perspectives*, 14(4), 23-48.
- Burmaoğlu, S., & Şeşen, H. (2011). Türk Firmalarının Organizasyonel İnovasyon Yeteneğini Etkileyen Faktörler Üzerine Bir Araştırma. *Ankara Üniversitesi SBF Dergisi*, 66(04), 1-20.
- Burstein, F., & Holsapple, C. W. (2008). *Handbook on decision support systems 2: variations*: Springer Science & Business Media.
- Campbell, T., & Slaughter, S. (1995). *Protecting the public's trust: A search for balance among benefits and conflicts in university-industry relations*. Paper presented at the AAAS Conference of University-Industry Cooperation: Results of Empirical Research.
- Caves, R. E. (1998). Industrial organization and new findings on the turnover and mobility of firms. *Journal of economic literature*, 36(4), 1947-1982.
- Caves, R. E. (2000). *Creative industries: Contracts between art and commerce*: Harvard University Press.
- Cem, I., & Keskin, G. (2013). Bilgi Ekonomilerinde Rekabet Üstünlüğü Oluşturulması Açısından İnovasyonun Önemi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 27(1), 41-57.

- Chen, C. P., Hu, J. L., & Yang, C. H. (2011). An international comparison of R&D efficiency of multiple innovative outputs: The role of the national innovation system. *Innovation-Organization & Management*, 13(3), 341-360. doi:DOI 10.5172/impp.2011.13.3.341
- Chen, Y.-y., Wu, L., & Zhai, Q.-g. (2019). Does ISO 9000 certification benefit service firms? *Sustainability*, 11(21), 5886.
- Cheng, T. (2019). Special economic zones: A catalyst for international trade and investment in unsettling times? *The Journal of World Investment Trade*, 20(1), 32-67.
- Chesbrough, H. (2006). *Open business models: How to thrive in the new innovation landscape*: Harvard Business Press.
- Chesbrough, H. W., & Appleyard, M. M. (2007). Open innovation and strategy. *California management review*, 50(1), 57-76.
- Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review*, 104(5), 141-147.
- Claessens, S., Djankov, S., & Lang, L. H. (2000). The separation of ownership and control in East Asian corporations. *Journal of financial economics*, 58(1-2), 81-112.
- Claessens, S., & Klapper, L. F. (2005). Bankruptcy around the world: Explanations of its relative use. *American Law Economics Review*, 7(1), 253-283.
- Claessens, S., & Yurtoglu, B. B. (2013). Corporate governance in emerging markets: A survey. *Emerging markets review*, 15, 1-33.
- Cleary, J. G., & Trigg, L. E. (1995). K*: An instance-based learner using an entropic distance measure. In *Machine Learning Proceedings 1995* (pp. 108-114): Elsevier.
- Coccia, M. (2008). Science, funding and economic growth: analysis and science policy implications. *World Review of Science, Technology Sustainable Development*, 5(1), 1-27.
- Cockburn, I. M., & Henderson, R. M. (1998). Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *The journal of industrial economics*, 46(2), 157-182.
- Cohen, W., Florida, R. L., & Goe, W. R. (1994). *University-industry research centers in the United States*: Center for Economic Development, H. John Heinz III School of Public Policy

- Corbett, C. J., Montes-Sancho, M. J., & Kirsch, D. A. (2005). The financial impact of ISO 9000 certification in the United States: An empirical analysis. *Management science*, 51(7), 1046-1059.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Crespo, N. F., & Crespo, C. F. (2016). Global innovation index: Moving beyond the absolute value of ranking with a fuzzy-set analysis. *Journal of Business Research*, 69(11), 5265-5271.
- Criscuolo, C., Haskel, J. E., & Slaughter, M. J. (2010). Global engagement and the innovation activities of firms. *International journal of industrial organization*, 28(2), 191-202.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press.
- Cui, R., Sun, J., Li, Y., Yang, K., & Wu, X. (2020). *Data-driven approach with artificial neural network for Global Innovation Index re-evaluation*. Paper presented at the 2020 6th International Conference on Big Data and Information Analytics (BigDIA).
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94.
- De la Fuente, A., & Ciccone, A. (2003). *Human capital in a global and knowledge-based economy* (Vol. 918): Office for Official Publications of the European Communities Luxembourg.
- De Mel, D., & Jayaratne, S. (2012). Vertical integration of industries in South Asia. In *Regional integration and economic development in South Asia*: Edward Elgar Publishing.
- Dechezleprêtre, A., Martin, R.-P., & Mohnen, M. (2014). *Knowledge spillovers from clean and dirty technologies*: Centre for Economic Performance.
- Demirguc-Kunt, A. (2008). *Finance, financial sector policies, and long-run growth*: World Bank Publications.
- Demirkaya, H., & Zengin, R. (2014). Hizmet inovasyonu ve bir uygulama örneği. *Elektronik Mesleki Gelişim ve Araştırma Dergisi (EJOİR) Cilt:2 Sayı:1, Cilt:2 Sayı:1*.
- Dietterich, T. G. (2000). *Ensemble methods in machine learning*. Paper presented at the International workshop on multiple classifier systems.
- Dietterich, T. G. (2002). *Machine learning for sequential data: A review*. Paper presented at the Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International

- Workshops SSPR 2002 and SPR 2002 Windsor, Ontario, Canada, August 6–9, 2002 Proceedings.
- Djankov, S., Freund, C., & Pham, C. S. (2010). Trading on time. *The review of economics statistics*, 92(1), 166-173.
- Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2002). The regulation of entry. *The quarterly Journal of economics*, 117(1), 1-37.
- Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The law and economics of self-dealing. *Journal of financial economics*, 88(3), 430-465.
- Dobrovolska, O., Sonntag, R., Kachula, S., Hubaryk, O., & Savanchuk, T. (2023). Financial and investment indicators for accelerating innovation development: Comparison of GII leaders and Ukraine. *Investment Management and Financial Innovations*.
- Dogan, B., & Deger, O. (2016). How globalization and economic growth affect energy consumption: Panel data analysis in the sample of BRIC countries. *International Journal of Energy Economics Policy*, 6(4), 806-813.
- Doğruel, M., & Firat, S. Ü. (2021). Veri Madenciliği Karar Ağaçları Kullanarak Ülkelerin İnovasyon Değerlerinin Tahmini ve Doğrusal Regresyon Modeli ile Karşılaştırmalı Bir Uygulama. *Istanbul Business Research*, 50(2), 465-493.
- Doidge, C., Karolyi, G. A., & Stulz, R. M. (2007). Why do countries matter so much for corporate governance? *Journal of financial economics*, 86(1), 1-39.
- Drazin, R. (1985). " Innovation and Entrepreneurship: Practice and Principles", by Peter F. Drucker (Book Review). *Human Resource Management*, 24(4), 509.
- Duchesne, J. N. (2012). Running Amuck: Using the Debacle of Recent Top Level Domain Expansions to Argue for Greater Governmental Participation in DNS Management. *CommLaw Conspectus*, 21, 151.
- Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (2016). *Global innovation index 2016: Winning the global innovation*: WIPO.
- Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (2020). *Global innovation index 2020*: Johnson Cornell University.
- Eaton, J., & Kortum, S. S. (1994). International patenting and technology diffusion. In: National Bureau of Economic Research Cambridge, Mass., USA.
- Elverdi, S. (2019). *AR-GE tabanlı ekonomik büyüme sürecinde inovasyon unsurunun rolü: Uluslararası karşılaştırmalı bir analiz*. . Erciyes Üniversitesi, Kayseri.

- ELVERDİ, S., & Hayriye, A. (2020). Türkiye’de inovasyon ölçümüne yönelik bir değerlendirme. *IBAD Sosyal Bilimler Dergisi*, 695-712.
- Ersöz, F. (2009). Avrupa inovasyon göstergeleri (EIS) ışığında Türkiye'nin konumu. *ITU Journal Series B: Social Sciences*, 6(1).
- Erumban, A. A., & Das, D. K. (2016). Information and communication technology and economic growth in India. *Telecommunications policy*, 40(5), 412-431.
- Famalika, A., & Sihombing, P. R. (2021). Implementation of K-Means and K-Medians Clustering in Several Countries Based on Global Innovation Index (GII) 2018. *Advance Sustainable Science, Engineering Technology*, 3(1), 343547.
- Florida, R. (2002). *The rise of the creative class* (Vol. 9): Basic books New York.
- Forman, C., & Goldfarb, A. (2020). *Concentration and agglomeration of IT innovation and entrepreneurship: Evidence from patenting*. Retrieved from
- Franco, C., & de Oliveira, R. H. (2017). Inputs and outputs of innovation: analysis of the BRICS: Theme 6–innovation technology and competitiveness. *RAI Revista de Administração e Inovação*, 14(1), 79-89.
- Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques*: Morgan Kaufmann.
- Gao, L., & Bai, X. (2014). An empirical study on continuance intention of mobile social networking services: Integrating the IS success model, network externalities and flow theory. *Asia Pacific Journal of Marketing Logistics*, 26(2), 168-189.
- Ge, Y., & Chen, Y. (2008). Foreign ownership and productivity of joint ventures. *Economic Development Cultural Change*, 56(4), 895-920.
- Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of international political economy*, 12(1), 78-104.
- Gilbert, D., Balestrini, P., & Littleboy, D. (2004). Barriers and benefits in the adoption of e-government. *International Journal of Public Sector Management*.
- Ginarte, J. C., & Park, W. G. (1997). Determinants of patent rights: A cross-national study. *Research policy*, 26(3), 283-301.
- Girma, S., Gong, Y., & Görg, H. (2008). Foreign direct investment, access to finance, and innovation activity in Chinese enterprises. *The World Bank Economic Review*, 22(2), 367-382.
- Goldberg, D. E. (1989). Genetic Algorithms in Search. *Optimization, Machine Learning*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*: MIT press.

- Gordon, R. J. (2012). *Is US economic growth over? Faltering innovation confronts the six headwinds*. Retrieved from
- Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers*, 104(4), 746-764.
- Grant, G., & Chau, D. (2005). Developing a generic framework for e-government. *Journal of Global Information Management*, 13(1), 1-30.
- Green, M., Björk, J., Forberg, J., Ekelund, U., Edenbrandt, L., & Ohlsson, M. (2006). Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. *Artificial intelligence in medicine*, 38(3), 305-318.
- Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5), 1153-1159.
- Greenwood, J., Hercowitz, Z., & Krusell, P. (1997). Long-run implications of investment-specific technological change. *The American economic review*, 342-362.
- Griliches, Z. (1998). Introduction to " R&D and productivity: The Econometric evidence". In *R&D and productivity: The econometric evidence* (pp. 1-14): University of Chicago Press.
- Grossman, G. M., & Helpman, E. (1993). *Innovation and growth in the global economy*: MIT press.
- Grzymala-Busse, J. W., & Siddhaye, S. (2004). *Rough set approaches to rule induction from incomplete data*. Paper presented at the Proceedings of the IPMU.
- Guadagno, F., & Wunsch-Vincent, S. (2020). Introduction to the GII 2020 theme. Who will finance innovation. *The Global Innovation Index 2020*, 67-73.
- Gupta, R. (2021). An efficient feature subset selection approach for machine learning. *Multimedia Tools Applications*, 80(8), 12737-12830.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar), 1157-1182.
- Gürtuna, F., & Polat, U. (2020). Küresel inovasyon endeksi verilerinin kümeleme analizi ile değerlendirilmesi. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 35(2), 551-566.

- Hagedoorn, J. (1993). Understanding the rationale of strategic technology partnering: Interorganizational modes of cooperation and sectoral differences. *Strategic management journal*, 14(5), 371-385.
- Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2013). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American behavioral scientist*, 57(5), 664-688.
- Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2000). Market value and patent citations: A first look. In: National bureau of economic research Cambridge, Mass., USA.
- Hall, B. H., & Lerner, J. (2010). The financing of R&D and innovation. In *Handbook of the Economics of Innovation* (Vol. 1, pp. 609-639): Elsevier.
- Hall, B. H., & Ziedonis, R. H. (2001). The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995. *rand Journal of Economics*, 101-128.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. J. A. S. e. n. (2009). The WEKA data mining software: an update. 11(1), 10-18.
- Hall, M. A., & Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge Data engineering*, 15(6), 1437-1447.
- Hall, M. A., & Smith, L. A. (1999). *Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper*. Paper presented at the FLAIRS conference.
- Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. *Review of Economics statistics*, 95(2), 347-361.
- Hamida Begum, M., & Aziz, S. I. (2019). Impact of domestic credit to private sector on gross domestic product in Bangladesh. *IOSR Journal of Economics Finance*, 10(1), 45-54.
- Hamidi, S., & Berrado, A. (2018). *Segmentation of innovation determinants: Case of the global innovation index*. Paper presented at the Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*: Morgan Kaufmann.
- Han, S. B., Chen, S. K., & Ebrahimpour, M. (2007). The impact of ISO 9000 on TQM and business performance. *The Journal of Business Economic Studies*, 13(2), 1.
- Hancıoğlu, Y. (2016). Küresel İnovasyon Endeksini Oluşturan İnovasyon Girdi Ve Çıktı Göstergeleri Arasındaki İlişkinin Kanonik Korelasyon Analizi İle İncelenmesi: Oecd

- Örneği. *Bolu Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 16(4), 131-158.
- Hanushek, E. A., & Woessmann, L. (2023). *The knowledge capital of nations: Education and the economics of growth*: MIT press.
- Hanushek, E. A., & Woessmann, L. J. J. o. e. l. (2008). The role of cognitive skills in economic development. 46(3), 607-668.
- Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics statistics*, 81(3), 511-515.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2): Springer.
- Haykin, S. (1998). *Neural networks: a comprehensive foundation*: Prentice Hall PTR.
- Head, K., & Mayer, T. (2004). Market potential and the location of Japanese investment in the European Union. *Review of Economics statistics*, 86(4), 959-972.
- Heckman, J. J., & LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *The review of economics statistics*, 92(2), 244-262.
- Helliwell, J. F., & Huang, H. (2014). New measures of the costs of unemployment: Evidence from the subjective well-being of 3.3 million Americans. *Economic Inquiry*, 52(4), 1485-1502.
- Heras, I., Casadesús, M., & Dick, G. P. (2002). ISO 9000 certification and the bottom line: a comparative study of the profitability of Basque region companies. *Managerial Auditing Journal*, 17(1/2), 72-78.
- Hesmondhalgh, D., & Baker, S. (2013). *Creative labour: Media work in three cultural industries*: Routledge.
- Hirsch, M., & Nov, A. (2004). World Investment Report 2003—FDI Policies for Development: National and International Perspectives. *The Journal of World Investment Trade*, 5(1), 215-226.
- Hızıroğlu, A., Pişirgen, A., Özcan, M., & İlter, H. K. (2022). Artificial Intelligence in Healthcare Industry: A Transformation From Model-Driven to Knowledge-Driven DSS. *Artificial Intelligence Theory Applications*, 2(1), 41-58.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *Ieee Transactions On Pattern Analysis Machine Intelligence*, 20(8), 832-844.
- Horbach, J. (2008). Determinants of environmental innovation—New evidence from German panel data sources. *Research policy*, 37(1), 163-173.

- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.
- <https://www.globalinnovationindex.org>
- Huarng, K.-H., & Yu, T. H.-K. (2022). Analysis of Global Innovation Index by structural qualitative association. *Technological Forecasting Social Change*, 182, 121850.
- Iglesias, S., & Ordenes, L. (2019). The Philippines 2018-2019: Authoritarian consolidation under Duterte. *Asia Maior*, 30.
- İmamoglu, İ. K., & Özdemir, D. (2023). Yönetişim, İnovasyon ve Ekonomik Büyüme Arasındaki İlişkiler: Gelişmiş ve Gelişmekte Olan Ülke Örnekleri. *Külliyeye*, 4(1), 100-136.
- Jacobzone, S., Choi, C.-w., & Miguet, C. (2007). Indicators of regulatory management systems. *OECD Working Papers on Public Governance*.
- Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, 957-970.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The quarterly Journal of economics*, 108(3), 577-598.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.
- Jankowska, B., Matysek-Jędrych, A., & Mroczek-Dąbrowska, K. (2017). Efficiency of national innovation systems: Poland and Bulgaria in the context of the Global Innovation Index. *Comparative Economic Research. Central Eastern Europe*, 20(3), 77-94.
- Jaramillo, C. F., Freund, C., Reis, J. G., Arvis, J. F., Wiederer, C. K., Ojala, L. M., & Kiiski, T. M. M. (2018). *Connecting to compete 2018: trade logistics in the global economy-the logistics performance index and its indicators*. Retrieved from
- Johnson, R. A., & Bhattacharyya, G. K. (2019). *Statistics: principles and methods*: John Wiley & Sons.
- Jorgenson, D. W., & Timmer, M. P. (2011). Structural change in advanced nations: a new set of stylised facts. *Scandinavian Journal of Economics*, 113(1), 1-29.
- Jovovic, R., Draskovic, M., Delibasic, M., & Jovovic, M. (2017). The concept of sustainable regional development—institutional aspects, policies and prospects. *Journal of International Studies*, 10(1).

- Kale, P., Singh, H., & Perlmutter, H. (2000). Learning and protection of proprietary assets in strategic alliances: Building relational capital. *Strategic management journal*, 21(3), 217-237.
- Karaata, E. S. (2012). *İnovasyonun Ölçümünde Yeni Arayışlar: TÜSİAD-Sabancı Üniversitesi Rekabet Forumu*. Retrieved from
- Karagöz, U. (2013). Kamu Sektöründe Yeni Bir İnovasyon Girişimi: Hizmet Envanteri Veri Tabanı (Hevt) Projesi. *Türk İdare Dergisi*, 476, 133-150.
- Keen, P. G., & Scott Morton, M. S. (1978). *Decision support systems: an organizational perspective*: Addison-Wesley Pub Co.
- Keller, K. L., & Lehmann, D. R. (2006). Brands and branding: Research findings and future priorities. *Marketing science*, 25(6), 740-759.
- Keller, W. (2004). International technology diffusion. *Journal of economic literature*, 42(3), 752-782.
- Khedhaouria, A., & Thurik, R. (2017). Configurational conditions of national innovation capability: A fuzzy set analysis approach. *Technological Forecasting Social Change*, 120, 48-58.
- Kijek, A., & Kijek, T. (2010). The analysis of innovation input-output relationships in EU member states. *Comparative Economic Research. Central Eastern Europe*, 13(3), 93-106.
- Kittur, A., Chi, E. H., & Suh, B. (2009). *What's in Wikipedia? Mapping topics and conflict using socially annotated category structure*. Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.
- Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). *He says, she says: conflict and coordination in Wikipedia*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Klapper, L., & Love, I. J. E. L. (2011). The impact of the financial crisis on new firm registration. *113(1)*, 1-4.
- Ko, W. L., Kim, S. Y., Lee, J.-H., & Song, T. H. (2020). The effects of strategic alliance emphasis and marketing efficiency on firm value under different technological environments. *Journal of Business Research*, 120, 453-461.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2), 273-324.

- Komaitis, K. (2010). *The current state of domain name regulation: domain names as second class citizens in a mark-dominated world*: Routledge.
- Kongar, E. J. (1995). Toplumsal değişme kuramları ve Türkiye gerçeği.
- Kononova, K. (2015). *Some Aspects of ICT Measurement: Comparative Analysis of E-Indexes*. Paper presented at the HAICTA.
- La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *The journal of finance*, 54(2), 471-517.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (2002). Investor protection and corporate valuation. *The journal of finance*, 57(3), 1147-1170.
- Lall, S., & Narula, R. (2004). Foreign direct investment and its role in economic development: do we need a new agenda? *The european journal of development research*, 16, 447-464.
- Lam, W. (2005). Barriers to e-government integration. *Journal of Enterprise Information Management*.
- Laudon, K. C., & Laudon, J. P. (2004). *Management information systems: Managing the digital firm*: Pearson Educación.
- Layos, J., & Peña, P. J. (2020). Can innovation save us? Understanding the role of innovation in mitigating the COVID-19 pandemic in ASEAN-5 economies. *De La Salle University Business Notes Briefings*, 8(2).
- Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for information systems*, 12(1), 50.
- Lee, Y. S. (1995). *The academic climate and technological innovation*: Iowa State University of Science and Technology.
- Lee, Y. S. (1996). 'Technology transfer' and the research university: a search for the boundaries of university-industry collaboration. *Research policy*, 25(6), 843-863.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2), 164-168.
- Levine, R., & Zervos, S. (1998). Stock markets, banks, and economic growth. *American Economic Review*, 537-558.
- Li, N., Niu, X., Chen, Q., & Zhou, H. (2020). Towards commercialization: the operational stability of perovskite solar cells. *Chemical Society Reviews*, 49(22), 8235-8286. doi:10.1039/D0CS00573H
- Lima, N., & Venables, A. J. (2001). Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review*, 15(3), 451-479.

- Lipsey, R. E. (2004). Home-and host-country effects of foreign direct investment. In *Challenges to globalization: Analyzing the economics* (pp. 333-382): University of Chicago Press.
- Luo, X., & Bhattacharya, C. B. (2006). Corporate social responsibility, customer satisfaction, and market value. *Journal of marketing*, 70(4), 1-18.
- Lutz, W., & Kc, S. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554), 2779-2791.
- Macit, N. Ş. (2024). Asya Ülkelerinin İnovasyon Performanslarının CILOS Temelli CoCoSo Yöntemi ile Değerlendirilmesi. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 15(41), 194-215.
- Mahroum, S., & Al-Saleh, Y. (2013). Towards a functional framework for measuring national innovation efficacy. *Technovation*, 33(10-11), 320-332.
- Majeed, M. T., & Ahmad, E. (2009). An Analysis of Host Country Characteristics that Determine FDI in Developing Countries: Recent Panel Data Evidence. *Lahore Journal of Economics*, 14(2).
- Malone, T. W., Yates, J., & Benjamin, R. I. (1989). The logic of electronic markets. *Harvard Business Review*, 67(3), 166-172.
- Martinez-Alier, J. (2003). *The Environmentalism of the poor: a study of ecological conflicts and valuation*: Edward Elgar Publishing.
- Maskus, K. E., & Penubarti, M. (1995). How trade-related are intellectual property rights? *Journal of International economics*, 39(3-4), 227-248.
- Matkin, G. W. (1994). Technology transfer and public policy: lessons from a case study. *Policy Studies Journal*, 22(2), 371-383.
- Mayda, B. (2019). *İnovasyon, yüksek teknoloji ve bilgi tabanlı ekonomi ile ekonomik büyüme arasındaki ilişki; Finlandiya örneğinde Türkiye üzerinde bir çalışma*. Bartın Üniversitesi, Sosyal Bilimler Enstitüsü,
- Melnyk, S. A., Sroufe, R. P., & Calantone, R. (2003). Assessing the impact of environmental management systems on corporate and environmental performance. *Journal of operations management*, 21(3), 329-351.
- Mendes, R., & Vilela, J. P. J. I. A. (2017). Privacy-preserving data mining: methods, metrics, and applications. 5, 10562-10582.

- Mentzer, J. T., Stank, T. P., & Esper, T. L. (2008). Supply chain management and its relationship to logistics, marketing, production, and operations management. *Journal of business logistics*, 29(1), 31-46.
- Merton, R. C. (1992). Financial innovation and economic performance. *Journal of applied corporate finance*, 4(4), 12-22.
- Metz, B., Davidson, O. R., Bosch, P. R., Dave, R., & Meyer, L. A. (2014). *Change, IPCC Climate %J Contribution of working group III to the fifth assessment report of the intergovernmental panel on climate change*. Retrieved from
- Meyers, R. A. (2009). *Encyclopedia of complexity and systems science* (Vol. 9): Springer New York.
- Minto-Coy, I., & McNaughton, M. (2016). Barriers to entrepreneurship and innovation: An institutional analysis of mobile banking in Jamaica and Kenya. *Social & Economic Studies*, 99-131.
- MogesEbero, E., & Begum, M. (2016). The desirability of doing business and flow of foreign direct investment nexus: The case of Ethiopia. *World*, 114, 37.81.
- Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine learning: algorithms and applications*: Crc Press.
- Mohnen, P., & Hoareau, C. (2003). What type of enterprise forges close links with universities and government labs? Evidence from CIS 2. *Managerial decision economics* 24(2-3), 133-145.
- Mohr, J. J., Fisher, R. J., & Nevin, J. R. (1996). Collaborative communication in interfirm relationships: moderating effects of integration and control. *Journal of marketing*, 60(3), 103-115.
- Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4), 525-533.
- Monteiro, P. V. P., De Oliveira Paula, F., Resende, D. N., Dos Santos, B. R. C., Da Silva, J. F., & Oliveira, M. d. A. S. (2023). *Evolution Of Brazil's Innovation Performance: An Analysis Of Possible Improvement Actions*. Paper presented at the II Conference on Data Science for the Social Sciences.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*: John Wiley & Sons.
- Moore, A., & Schneider, J. (1995). Memory-based stochastic optimization. *Advances in Neural Information Processing Systems*, 8.

- Morris, R., & Aziz, A. (2011). Ease of doing business and FDI inflow to Sub-Saharan Africa and Asian countries. *Cross Cultural Management: An International Journal*.
- Neal, R. M. (2012). *Bayesian learning for neural networks* (Vol. 118): Springer Science & Business Media.
- Nguyen, H.-L., Pham, B. T., Son, L. H., Thang, N. T., Ly, H.-B., Le, T.-T., . . . Tien Bui, D. (2019). Adaptive network based fuzzy inference system with meta-heuristic optimizations for international roughness index prediction. *Applied Sciences*, 9(21), 4715.
- Nick, Z. Z., & Themis, P. (2001). Web search using a genetic algorithm. *IEEE Internet Computing*, 5(2), 18-26.
- Niebel, T. (2018). ICT and economic growth—Comparing developing, emerging and developed countries. *World development*, 104, 197-211.
- Norris, P. (2000). *A virtuous circle: Political communications in postindustrial societies*: Cambridge University Press.
- Nyarku, K. M., & Oduro, S. (2018). Effect of legal and regulatory framework on SMEs growth in the Accra Metropolis of Ghana. *The International Journal of Entrepreneurship and Innovation*, 19(3), 207-217.
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., . . . Walsh, J. (2020). *Deep learning vs. traditional computer vision*. Paper presented at the Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1.
- Oecd. (2005). *Oslo Kılavuzu*. Ankara: Tübitak Yayınları.
- Okrah, J., & Hajduk-Stelmachowicz, M. (2020). Political stability and innovation in Africa. *Journal of International Studies*.
- Oluwatobi, S., Ola-David, O., Olurinola, I., Alege, P., & Ogundipe, A. (2016). Human capital, institutions and innovation in Sub-Saharan Africa. *International Journal of Economics and Financial Issues*, 6(4), 1507-1514.
- Omer, M., El-Amin, M., Nasr, A., & Ahmed, R. (2020). *Modeling, visualization, and analysis of African innovation performance*. Paper presented at the International Conference on Learning Representation (ICLR).
- Oreopoulos, P., Page, M. E., & Stevens, A. H. (2006). The intergenerational effects of compulsory schooling. *Journal of Labor Economics*, 24(4), 729-760.

- Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the “IT” in IT research—A call to theorizing the IT artifact. *Information systems research*, 12(2), 121-134.
- Oulton, N. (2012). Long term implications of the ICT revolution: Applying the lessons of growth theory and growth accounting. *Economic Modelling*, 29(5), 1722-1736.
- Ovali, S. (2014). Küresel rekabet gücü açısından Türkiye’nin konumu üzerine bir değerlendirme. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*(13), 17-36.
- Park, W. G., & Ginarte, J. C. (1997). Intellectual property rights and economic growth. *Contemporary Economic Policy*, 15(3), 51-61.
- Peck, H., & Christopher, M. (2004). Building the Resilient Supply Chain *The International Journal of Logistics Management*.
- Pekdemir, I., Erdoğan, U., Yasemin, Ş., & Selçuk, Y. (2015). "Yönetim ve Organizasyon" Alanında Bir Bilim Dalında 2004-2015 Yılları Arasında Yönetilmiş Lisansüstü Tezlerin İncelenmesi. *İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü Yönetim Dergisi*, 26(79), 20-56.
- Pençe, İ., Kalkan, A., & Çeşmeli, M. Ş. (2019). Estimation of the country ranking scores on the global innovation index 2016 using the artificial neural network method. *International Journal of Innovation Technology Management*, 16(04), 1940007.
- Perretti, F., & Negro, G. (2007). Mixing genres and matching people: A study in innovation and team composition in Hollywood. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational Organizational Psychology Behavior*, 28(5), 563-586.
- Phelan, T. J. (1994). Striking the mother lode in science: The importance of age, place, and time. *The Journal of Higher Education*.
- Piwonski, K. (2010). Does the ‘Ease of Doing Business’ In a Country Influence its Foreign Direct Investment Inflows? *Bryant University*.
- Platt, J. C. (1999). 12 fast training of support vector machines using sequential minimal optimization. *Advances in kernel methods*, 185-208.
- Policy, R. (2011). Governance: Supporting Economic Growth and Serving the Public Interest. In: OECD: Paris, France.
- Pop, D. M., & Pop, M. T. (2018). *Measuring the innovation of economy through global and European tools*. Paper presented at the MATEC Web of Conferences.

- Porta, R. L., Lopez-de-Silanes, F., & Shleifer, A. (2008). The economic consequences of legal origins. *Journal of economic literature*, 46(2), 285-332.
- Porta, R. L., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of political economy*, 106(6), 1113-1155.
- Porter, M. E. (1985). Technology and competitive advantage. *Journal of business strategy*, 5(3), 60-78.
- Porter, M. E. (2011). *Competitive advantage of nations: creating and sustaining superior performance*: simon and schuster.
- Postema, G. J. (2023). *Law's Rule: The Nature, Value, and Viability of the Rule of Law*: Oxford University Press.
- Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative science quarterly*, 116-145.
- Power, D. J., & Sharda, R. (2007). Model-driven decision support systems: Concepts and research directions. *Decision support systems*, 43(3), 1044-1061.
- Pradhan, R. P. (2009). The nexus between financial development and economic growth in India: Evidence from multivariate VAR model. *International Journal of Research Reviews in Applied Sciences*, 1(2).
- Prüss-Üstün, A., Wolf, J., Corvalán, C., Bos, R., & Neira, M. (2016). *Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks*: World Health Organization.
- Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World development*, 22(9), 1325-1343.
- Qiang, C. Z., Kuek, S. C., Dymond, A., & Esselaar, S. (2012). *Mobile applications for agriculture and rural development*. Retrieved from
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1, 81-106.
- Quinlan, J. R. (1992). *Learning with continuous classes*. Paper presented at the 5th Australian joint conference on artificial intelligence.
- Quinlan, J. R. (1996). *Bagging, boosting, and C4. 5*. Paper presented at the Aaai/Iaai, vol. 1.
- Rahimi, A., & Recht, B. (2007). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20.
- Rahm, D. (1995). *University-firm linkages for industrial innovation*: Center for Economic Policy Research, Stanford University.

- Rainer, R. K., & Prince, B. (2021). *Introduction to information systems*: John Wiley & Sons.
- Ramasamy, B., Yeung, M., & Laforet, S. (2012). China's outward foreign direct investment: Location choice and firm ownership. *Journal of World Business*, 47(1), 17-25.
- Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning* (Vol. 1): Springer.
- Reisman, Y. (1996). Computer-based clinical decision aids. A review of methods and assessment of systems. *Medical Informatics*, 21(3), 179-197.
- Roessner, J. D., Lee, Y. O. N. G., Shapira, P. H. I. L. I. P., & Bozeman, B. A. R. R. Y. (1996). Evaluation of Iowa State University's Center for Advanced Technology Development. *School of Public Policy, Georgia Institute of Technology, Volume I*.
- Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003). Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS quarterly*, 237-263.
- Samson, D., & Terziovski, M. (1999). The relationship between total quality management practices and operational performance. *Journal of operations management*, 17(4), 393-409.
- Sandu, S., & Ciocanel, B. (2014). Impact of R&D and Innovation on High-tech Export. *Procedia Economics Finance*, 15, 80-90.
- Saranya, C., & Manikandan, G. (2013). A study on normalization techniques for privacy preserving data mining. *International Journal of Engineering Technology*, 5(3), 2701-2704.
- Sarker, I. H., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, 7, 1-29.
- Savaşçı, D. (2017). *Küresel İnovasyon Endeksi*. Retrieved from
- Schelhas, L. T., Li, Z., Christians, J. A., Goyal, A., Kairys, P., Harvey, S. P., . . . Berry, J. J. (2019). Insights into operational stability and processing of halide perovskite active layers. *Energy & Environmental Science*, 12(4), 1341-1348. doi:10.1039/C8EE03051K
- Schelhas, L. T., Li, Z., Christians, J. A., Goyal, A., Kairys, P., Harvey, S. P., . . . Science, E. (2019). Insights into operational stability and processing of halide perovskite active layers. 12(4), 1341-1348.
- Schumpeter, J. (1934). *The theory of economic development* Harvard University Press. Cambridge, MA.

- Schumpeter, J. A. (1939). *Business cycles* (Vol. 1): McGraw-Hill New York.
- Sellers, M. N. S. (2014). What Is the Rule of Law and Why Is It So Important? In J. R. Silkenat, J. E. Hickey Jr, & P. D. Barenboim (Eds.), *The Legal Doctrines of the Rule of Law and the Legal State (Rechtsstaat)* (pp. 3-13). Cham: Springer International Publishing.
- Shaikh, O., & Malik, Z. (2020). Rule of Law: A Literature Review on Theoretical Concepts, Challenges from Mass Media to Innovation. *Jus Corpus LJ*, 1, 31.
- Sharda, R., Delen, D., & Turban, E. (2014). *Business intelligence and analytics: systems for decision support*: Pearson.
- Shin, K.-S., & Lee, Y.-J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications*, 23(3), 321-328.
- Singh, G. (2015). *Relationship between doing business index and foreign direct investment*. Paper presented at the International Conference on Ease of Doing Business: Contemporary Issues, Challenges and Future Scope.
- Singha, K., & Singh, M. A. (2022). Political stability and its effect on economy: some lessons from Sikkim Himalaya. *Journal of Social and Economic Development*, 24(2), 340-354. doi:10.1007/s40847-022-00199-9
- Sohn, S. Y., Kim, D. H., & Jeon, S. Y. (2016). Re-evaluation of global innovation index based on a structural equation model. *Technology Analysis Strategic Management*, 28(4), 492-505.
- Soumitra, D., Lanvin, B., & Wunsch-Vincent, S. (2020). *Global innovation index 2020: who will finance innovation?* : WIPO.
- Sovacool, B. K. (2014). What are we doing here? Analyzing fifteen years of energy scholarship and proposing a social science research agenda. *Energy Research Social Science*, 1, 1-29.
- Sovacool, B. K., & Mukherjee, I. (2011). Conceptualizing and measuring energy security: A synthesized approach. *Energy*, 36(8), 5343-5355.
- Sparkman Jr, R. M. (1979). Multivariate data analysis with readings. *JMR, Journal of Marketing Research*, 16(000003), 437.
- Stephan, P. (2012). *How economics shapes science*: Harvard University Press.
- Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. *World development*, 32(8), 1419-1439.

- Sui, B., Chang, C.-P., & Chu, Y. (2021). Political Stability: an Impetus for Spatial Environmental Spillovers. *Environmental and Resource Economics*, 79(2), 387-415. doi:10.1007/s10640-021-00568-8
- Szalacha-Jarmużek, J., & Pietrowicz, K. (2018). Missing causality and absent institutionalization. A case of Poland and methodological challenges for future studies of interlocking directorates. *Economics & Sociology*, 11(4), 157.
- Şahin, Z. (2020). *AR-GE ve inovasyon faaliyetlerinin büyüme ile ilişkisi: KOBİ'ler üzerine bir inceleme*. İstanbul Gelişim Üniversitesi,
- Şimşit, Z. T., Fırat, S. Ü. O., Es, H. A., Erdem, M., Topgöl, M., & Günay, N. S. (2014). *Bilgi Ve İletişim Teknolojileri Çerçevesinden Küresel İnovasyon Endeksinin Analizi Ve Veri Madenciliği Kullanılarak Ülkelerin Kümelenmesi*. Paper presented at the Yönetim Bilişim Sistemleri Kongresi, İstanbul.
- Tezcan, N. (2015). Comparative Analysis Of EU Countries And Turkey In Terms Of Innovation Performance. *Research Journal of Business Management*, 2(2), 204-219.
- Throsby, D. (2010). *The economics of cultural policy*: Cambridge university press.
- Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., & De Vries, G. J. (2015). An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics*, 23(3), 575-605.
- Tirole, J. (1988). *The theory of industrial organization*: MIT press.
- Toker, K., & Görener, A. (2020). İnovasyon ile Küresel Rekabetçilik Düzeyi Arasındaki İlişkide Ekonomik Serbestliğin Düzenleyici Rolü. *İşletme Araştırmaları Dergisi*, 12(3), 2996-3007.
- Tung, L. L., & Rieck, O. (2005). Adoption of electronic government services among business organizations in Singapore. *The Journal of Strategic Information Systems*, 14(4), 417-440.
- Turan, E. (2018). *Küresel inovasyon indeksi kapsamında türkiye'nin konumunun değerlendirilmesi: 2011-2017* Karadeniz Teknik Üniversitesi,
- Turkmen, M., & Aynaoglu, Y. (2018). Küresel rekabet endeksi göstergelerinin küresel inovasyon endeksi üzerindeki etkisi. *Business Management Studies: An International Journal*.
- Tüfekci, Ö. K., & Tüfekci, N. (2014). Pazarlama yeniliği ile pazarlama performansı ilişkisini açıklamaya yönelik bir vaka çalışması: Süleyman Demirel üniversitesi olimpiik yüzme havuzu örneği ve bir model önerisi. *Girişimcilik ve Kalkınma Dergisi*.

- Tüysüz, H. (2020). *Araştırma-geliştirme ve inovasyon çalışmalarında devletin rolü: KOSGEB örneği*. Zonguldak Bülent Ecevit Üniversitesi Sosyal Bilimler Enstitüsü,
- Udimal, T. B., Jincai, Z., Ibn Musah, A.-A., & Hua, C. (2019). Determinants of new products innovation in Ghanaian SMEs sector. *Journal of Global Entrepreneurship Research*, 9, 1-11.
- Usman, K., & Liu, Z. (2015). Innovation index framework to measure the innovation capacity and efficiency of SAARC countries. *European Journal of Social Sciences*, 46(3), 325-338.
- Ünlükaplan, İ. (2009). Avrupa Birliği üyesi ülkelerde iktisadi kalkınma, rekabetçilik ve inovasyon ilişkilerinin kanonik korelasyon analizi ile belirlenmesi. *Maliye Dergisi*, 157(2), 235-250.
- Van Ark, B., O'Mahony, M., & Timmer, M. P. (2008). The productivity gap between Europe and the United States: trends and causes. *Journal of economic perspectives*, 22(1), 25-44.
- Van de Walle, N. (1998). *Economic globalization and political stability in developing countries*: Project on World Security, Rockefeller Bros. Fund.
- Vu, K. M. (2011). ICT as a source of economic growth in the information age: Empirical evidence from the 1996–2005 period. *Telecommunications policy*, 35(4), 357-372.
- Wang, C., Kafouros, M., Yi, J., Hong, J., & Ganotakis, P. (2020). The role of government affiliation in explaining firm innovativeness and profitability in emerging countries: Evidence from China. *Journal of World Business*, 55(3), 101047.
- Wang, L., Luo, G.-l., Sari, A., & Shao, X. (2020). Technological forecasting & social change what nurtures the fourth industrial revolution? An investigation of economic and social determinants of technological innovation in advanced economies. *Technological Forecasting & Social Change*, 161(September), 120305.
- Weisberg, S. (2005). *Applied linear regression* (Vol. 528): John Wiley & Sons.
- West, D. (2005). Fifth annual global e-government study: US and four Asian nations are world's five best online governments. *Brown University*.
- West, J., & Mace, M. (2010). Browsing as the killer app: Explaining the rapid success of Apple's iPhone. *Telecommunications policy*, 34(5-6), 270-286.
- Witten, D., & James, G. (2013). *An introduction to statistical learning with applications in R*: springer publication.

- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). *Practical machine learning tools and techniques*. Paper presented at the Data mining.
- World-Bank. (2016). *Doing Business 2016 Measuring Regulatory Quality and Efficiency*. : World Bank.
- Yachmeneva, V. (2014). Factors influencing the enterprise innovation. *ECONTECHMOD: An International Quarterly Journal on Economics of Technology and Modelling Processes*, 3(1), 133--138.
- Yang, G., & Maskus, K. E. (2001). Intellectual property rights, licensing, and innovation in an endogenous product-cycle model. *Journal of International economics*, 53(1), 169-187.
- Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PloS one*, 7(6), e38869.
- Ye Sheng, S., & Wong, R. (2012). Innovation Capacity in China: An Analysis in a Global Context. *International Journal of China Marketing*, 3(1), 88-106.
- Yu, T. H.-K., & Huarng, K.-H. (2023). Configural analysis of GII's internal structure. *Journal of Business Research*, 154, 113323.
- Zachman, J. A. (1987). A framework for information systems architecture. *IBM systems journal*, 26(3), 276-292.
- Zhang, J., Dawes, S. S., & Sarkis, J. (2005). Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. *Journal of Enterprise Information Management*.
- Zhou, K., Ebenhofer, G., Eitzinger, C., Zimmermann, U., Walter, C., Saenz, J., . . . Oriol, J. N. (2014). *Mobile manipulator is coming to aerospace manufacturing industry*. Paper presented at the 2014 IEEE international symposium on robotic and sensors environments (ROSE) proceedings.
- Zupan, B., & Demsar, J. (2008). Open-source tools for data mining. *Clinics in laboratory medicine*, 28(1), 37-54.

APPENDIX

Appendix A. Countries

Albania	Croatia	Ireland	Nepal	Slovakia
Algeria	Cyprus	Israel	Netherlands	Slovenia
Argentina	Czech Republic	Italy	New Zealand	South Africa
Armenia	Denmark	Jamaica	Niger	Spain
Australia	Dominican Republic	Japan	Nigeria	Sri Lanka
Austria	Ecuador	Jordan	North Macedonia	Sweden
Azerbaijan	Egypt	Kazakhstan	Norway	Switzerland
Bahrain	El Salvador	Kenya	Oman	Tajikistan
Bangladesh	Estonia	Kuwait	Pakistan	Thailand
Belarus	Ethiopia	Kyrgyzstan	Panama	Togo
Belgium	Finland	Lao People's Democratic Republic	Paraguay	Trinidad And Tobago
Benin	France	Latvia	Peru	Tunisia
Bosnia And Herzegovina	Georgia	Lithuania	Philippines	Turkey
Botswana	Germany	Luxembourg	Poland	Uganda
Brazil	Ghana	Madagascar	Portugal	Ukraine
Brunei Darussalam	Greece	Malaysia	Qatar	United Arab Emirates
Bulgaria	Guatemala	Mali	Republic Of Korea	United Kingdom
Burkina Faso	Guinea	Malta	Republic Of Moldova	United Republic Of Tanzania
Cambodia	Honduras	Mauritius	Romania	United States Of America
Cameroon	Hong Kong	Mexico	Russian Federation	Uruguay
Canada	Hungary	Mongolia	Rwanda	Uzbekistan
Chile	Iceland	Montenegro	Saudi Arabia	Viet Nam
China	India	Morocco	Senegal	Yemen
Colombia	Indonesia	Mozambique	Serbia	Zambia
Costa Rica	Iran (Islamic Republic Of)	Namibia	Singapore	Zimbabwe

CURRICULUM VITAE

Rabia Sultan YILDIRIM completed her primary, secondary, and high school education in Adana before graduating from Çukurova University with a degree in Industrial Engineering in 2014. She went on to complete the Industrial Engineering Master's Program at Adana Alparslan Türkeş Science and Technology University in 2020, and subsequently began her PhD Program in Industrial Engineering at the same university. From 2014 to 2021, she gained valuable experience as a Material Management Engineer in a factory before being appointed as a Lecturer at Adana Alparslan Türkeş Science and Technology University in 2021, where she currently works.

