ESTIMATION OF SUCCESS OF ENERGY EFFICIENY SUPPORT PROGRAM APPLICATIONS FOR SME'S WITH DATA MINING

(KOBİ'LERE YÖNELİK ENERJİ VERİMLİLİĞİ DESTEK PROGRAM BAŞVURULARININ BAŞARISININ VERİ MADENCİLİĞİ İLE TAHMİNİ)

by

Hande DEMİROĞLU, B.S.

Thesis

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE
in
LOGISTICS AND FINANCIAL MANAGEMENT
in the
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
of
GALATASARAY UNIVERSITY

Supervisor: Ass. Prof. Dr. İlke BEREKETLİ
Jun 2023

ACKNOWLEDGEMENTS

I would like to thank my supervisor Ass. Prof. Dr. İlke BEREKETLİ for her continuous support during preparation of this thesis. This study would not have been possible without her sincere collaboration.

My husband Bedri, my sister Eylül, my friend Abdül and my classmates were kind supporters for early reading and time sharing. I am in debt forever.

Jun 2023
Hande DEMİROĞLU

TABLE OF CONTENTS

LIST OF SYMBOLS

AI	:Artificial Intelligence
DPT	:Devlet Planlama Teşkilatı
EIA	:Environmental Investigation Agency
GEF	:Global Environment Facility
KOBİ	:Küçük ve Orta Büyüklükteki İşletmeler
KOSGEB	**:**Küçük ve Orta Ölçekli İşletmeleri Geliştirme ve Destekleme İdaresi
KTOE	**:**Kilotonne of Oil Equivalent
MTOE	**:**Million Tonnes of Oil Equivalent
ROC	**:**Receiver Operating Characteristic
SMEs	**:**Small to Medium Enterprises
SVM	**:**Support Vector Machines
UNFCCC	**:**The UN Framework Convention on Climate Change
WWF	**:**World Wide Fund for Nature

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

99.8% of the businesses operating in Turkey are small and medium business enterprises. Small and medium-sized enterprises (SMEs) also provide 76.7% of total employment. As a result, they are crucial to the economy of the nation. As a result of receiving such large payments in the economy, their energy consumption is also high. Particular attention should be paid to the energy distribution of SMEs among the employment for the limits of energy consumption. As a result, it benefits the nation's economy, the wise use of resources, and environmental consciousness.

KOSGEB is an institution operating under the Ministry of Industry and Technology in order to meet the social and economic needs of Turkey, to increase the superiority and level of SMEs above the share and performance of SMEs, and to achieve industrial control suitable for economic development.

In this study, it is aimed to estimate the approval status of the projects by analyzing the success and failure of the projects of the enterprises that applied within the scope of the SME Development Support Program 'Increasing Energy Efficiency' project call published by KOSGEB in 2017. Thus, the scientific approach will be taken as a basis in the decisions made about the projects and the estimations. Projects evaluated by KOSGEB in 2017 to be used in the study; company status, number of partners, number of years of operation, number of existing personnel, number of white-collar personnel, number of blue-collar personnel, project budget, the part of the project to be met with equity, the enterprise's equity, domestic sales, foreign sales, administrative expenses. The role of these qualifications on the approval of the project by KOSGEB will be examined. The project application data from 81 provinces of Turkey is used in the study. Applications are received online through the system and stored by KOSGEB. Applications are evaluated by the application evaluation board. With this review, it has been tried to make an estimation of the success of the enterprises that will apply for the projects in the future. The data were preprocessed and adapted to the python software. Data set; LightGBM, Logistic Regression and Support Vector Machines algorithms are classified. Classification results were compared and LightGBM algorithm was found to be the most successful algorithm with 0.93 F1 score and 0.88 accuracy.

Key words: SMEs, Energy Efficieny, Support Programs, Data Mining

ÖZET

Türkiye'de faaliyet gösteren işletmelerin %99,8'i küçük ve orta ölçekli işletmelerdir. Küçük ve orta ölçekli işletmeler (KOBİ'ler) de toplam istihdamın %76,7'sini sağlamaktadır. Bu nedenle ülke ekonomisinde önemli bir rol oynarlar. Ekonomide bu kadar büyük pay almalarının bir sonucu olarak enerji tüketimleri de yüksektir. Enerji verimliliğinin artırılmasına yönelik çalışmalar arasında özellikle KOBİ'lerin enerji verimliliğinin artırılmasına önem verilmesi gerekmektedir. Böylece hem ülke ekonomisine hem de sınırlı kaynakların verimli kullanılmasına ve çevre bilincine katkı sağlanmaktadır.

KOSGEB, Türkiye'nin sosyal ve ekonomik ihtiyaçlarının karşılanmasına yönelik, KOBİ'lerin payının ve etkinliğinin artırılması, KOBİ'lerin rekabet gücünün ve seviyesinin yükseltilmesi, ekonomik gelişmeye uygun endüstriyel entegrasyonun sağlanması amacıyla Sanayi ve Teknoloji Bakanlığı'na bağlı olarak faaliyet gösteren bir kuruluştur.

Bu çalışmada KOSGEB tarafından 2017 yılında yayımlanan KOBİ Gelişim Destek Programı 'Enerji Verimliliğinin Arttırılması' isimli proje çağrısı kapsamında başvuru yapan işletmelerin, projelerinin başarı ve başarısızlıklarının veri madenciliği algoritmaları ile analiz edilerek projelerin onay alma durumunun tahmin edilmesi amaçlanmaktadır. Böylece yapılan tahminler ile projelerle ilgili kararlarda bilimsel yaklaşım esas alınacaktır. Çalışmada Türkiye'nin 81 ilinden proje uygulama verileri kullanılmıştır. Başvurular sistem üzerinden online olarak alınmaktadır ve KOSGEB tarafından saklanmaktadır. Başvurular başvuru değerlendirme kurulu tarafından değerlendirilmektedir. Çalışmada kullanılmak üzere 2017 yılında KOSGEB tarafından değerlendirilen projeler; firma statüsü, ortak sayısı, faaliyette bulunduğu yıl sayısı, mevcut personel sayısı, beyaz yakalı personel sayısı, mavi yakalı personel sayısı, proje bütçesi, projenin özkaynaklar ile karşılanacak kısmı, işletmenin özkaynağı, yurt içi satışları, yurt dışı satışları, yönetim giderlerine göre incelenmiştir. Bu niteliklerin projenin KOSGEB tarafından onaylanması üzerindeki rolü incelenecektir. Bu inceleme ile birlikte ileride projelere başvuracak işletmelerin başarı tahmini yapılmaya çalışılmıştır. Veriler ön işleme tabi tutularak python yazılımına uyarlanmıştır. Veri seti; LightGBM, Lojistik Regresyon ve Destek Vektör Makineleri algoritmaları sınıflandırılmıştır. Sınıflandırma sonuçları karşılaştırılmış ve LightGBM algoritması 0.93 F1 score ve 0.88 Accuracy ile en başarılı algoritma olarak bulunmuştur.

Anahtar Kelimeler: KOBİ, Enerji Verimliliği, Destek Programları, Veri Madenciliği

INTRODUCTION

The 1990 law that created KOSGEB, number 3624, limited its services and assistance to SMEs in the production sector until 2009. However, given the increased potential for producing added value and creating jobs in sectors other than the production sector in our nation, as well as the requests from SMEs in those sectors, it is necessary to expand the target population of KOSGEB in order to include all SMEs..

Accordingly, the "Law numbered 5891 on Amending the KOSGEB Establishment Law numbered 3624" was released on May 5, 2009, and it had the official gazette number 27219. The legislation amendment gave KOSGEB the justification it needed to help SMEs in industries other than production. The Cabinet Decree No. 15431 on "Determination of Sector and Regional Priorities of Small and Medium Enterprises that will Benefit from the Services and Supports to be Provided by KOSGEB" was published in the Official Journal on September 18, 2009, under the number 27353, and it established the sector and regional priorities for the businesses that will benefit from the services and supports to be offered by KOSGEB.

In accordance with the national and international goals of the nation, the KOBIGEL - SME Development Support Program's goal is to increase the share and contribution of SME's to the economy and to support SME projects that will increase their competitiveness and added value.

The Call for Proposals specifies the kinds of project costs that will be financed. The Project evaluation Board determines the costs that will be covered and specifies the support amount during the evaluation process for each qualifying project. The following expenses are not supported: real estate purchase, building construction, furnishing, vehicle purchase and rental, taxes, withholding, charges and fees, social security premiums. In this study, the call issued in 2017 is approached. In the support call, activities involving one or more of the subjects of reducing energy losses in manufacturing processes, acquiring technologies that will provide energy efficiency and installing renewable energy systems will be supported.

This study has been also prepared to solve the problem of waste of resources and time caused by unsuccessful project applications of SMEs that are intended to be supported by KOSGEB. It is aimed to prevent the waste of limited government resources and to ensure that the grant budgets are distributed correctly. In addition, the reasons for the failure of unsuccessful projects of SMEs, which allocate labor and budget to project application processes, are investigated. By evaluating the characteristics that contribute to the success or failure of SMEs that have previously applied for project funding, data mining classification algorithms have been utilized to increase the effectiveness of government assistance. Data mining classification algorithms were employed to forecast success or failure based on the features that contributed to failure. The second section of the paper provides an explanation of the definition of data mining, along with associated ideas, data mining application areas, procedures, and methods. The idea of energy efficiency is covered in the third chapter. The steps of the KOSGEB SME Development Support Program and Energy Efficiency Call are described in the fourth chapter. The data are thoroughly presented in the fifth chapter, and the algorithms are then applied to the data set. The information from the algorithms is interpreted, the success factors are exposed, the success and accuracy of the algorithms are compared, and the algorithm that predicts with the highest accuracy is chosen in the conclusion section.

LITERATURE REVIEW

The studies of earlier firm failures have been discussed in this section of the thesis. When analyzing whether a corporation was successful or unsuccessful, financial ratios were initially utilized in studies that were published in the literature. First research used conventional statistical techniques for estimate. Making forecasts using traditional techniques became simpler thanks to the ample information gathered from large-scale organizations. However, due to a lack of financial information, applying these strategies to newly found and small-to-medium-sized businesses did not yield successful results. Since then, non-financial methodologies have been utilized to analyze newly founded and small-to-medium-sized businesses. The Lussier approach, which was one of the first non-financial methods, incorporated models from discriminant analysis and logistic regression. In addition to the algorithms' widespread application, data mining algorithms have begun to use categorization data generated from both financial and non-financial ways.

Particularly in applications in electronic commerce, science, medicine, industry, and education, data mining and knowledge discovery are starting to emerge as a new and fundamental research area. Data mining refers to all research projects focused on developing, examining, and putting into practice inductive techniques that will aid in the extraction of meaningful and practical information from the available unstructured data. From big data sets, it is used to extract patterns, changes, irregularities, and linkages. Important findings can be achieved in this fashion for mechanisms that influence decision-making, including web filtering, the identification of genes in DNA sequences, the identification of trends and anomalies in the economy, and consumer shopping patterns.

The amount of digital data has exploded in the last 10 years, showing an increase beyond estimates. In contrast, the number of scientists, engineers and analysts does not change. To address this disproportion, solutions to new research problems can be divided into several groups:

- Developing new algorithms and systems for large-volume and multidimensional data mining,
- Development of new algorithms, techniques and systems for mining new data types,
- Developing algorithms, protocols and infrastructures for distributed data mining,
- Advancement and development of the use of existing data mining systems,
- Creation of unique security and privacy paradigms for data mining.

In order for all these efforts to be successful and yield results, the support of multi-disciplinary and interdisciplinary business fields is necessary. Important experimental components need to be implemented, requiring the creation of relevant systems, metered infrastructures, and test environments. As tried to be summarized above, data mining studies have become a critical area in today's information society. The extraordinary developments in informatics, internet and media technologies have left us face to face with an ocean of data. In order to reach information from this ocean of data, in other words, to fish, many research groups on data mining have been established and are being established, especially in Europe and the USA.

Traditional statistics, artificial intelligence, and artificial intelligence are typically the three techniques used in data mining analysis (Girija & Srivatsa, 2006). According to David J. Hand (1998), traditional statistics is primarily used to analyze data, data linkages, and deal with numerical data in sizable databases. Regression analysis, cluster analysis, and discriminate analysis are a few examples of classical statistics. According to Girija and Srivatsa (2006), artificial intelligence (AI) uses "human-thought-like" processing to solve statistical issues. AI employs a variety of methods, including neural computing, fuzzy logic, and genetic algorithms. For data analysis and knowledge discovery, machine learning combines advanced statistical techniques with AI heuristics (Kononenko & Kukar, 2007). The groups of techniques that are used in machine learning include neural networks, symbolic learning, genetic algorithms, and swarm optimization. Data mining can benefit from these technologies because it aims to find patterns, define trends, and predict behavior.

As shown in Figure 1, a typical data mining process consists of a number of interactive steps that frequently start by merging unstructured data from different sources and formats. These raw data are cleaned to remove noise, duplicate data, and inconsistent data (Han et al., 2011). In order to retrieve summarized data, filtration and aggregation procedures are then used on the cleaned data to convert it into relevant formats that other data mining tools can comprehend. In actuality, intriguing information is gleaned from the modified data. The analysis of this data identifies the most intriguing patterns. Knowledge is eventually visualized for the user. Han et al. (2011) provide more thorough details on a data mining procedure.

When there are enormous volumes of data available, data mining techniques are used in a variety of fields to find undiscovered or hidden information. In this sense, data mining techniques used on the web are referred to as web mining, those used on text are referred to as text mining, and those used in libraries are referred to as bibliomining, according to Girija and Srivatsa (2006).

The term "bibliomining," or "data mining for libraries," was created by Nicholson and Stanton (2003a) to describe the combination of data warehousing, data mining, and bibliometrics. This expression is used in library systems to monitor transaction patterns, behavioral changes, and trends. Even though the concept is not new, the name "bibliomining" was created to make it simpler to search for the terms "library" and "data mining" in the context of libraries rather than software libraries. Bibliomining is an essential method for finding pertinent library information in historical data to enhance decision-making (Kao et al., 2003). To provide a thorough evaluation of the library system, bibliomining must be done iteratively in conjunction with other measurement and assessment approaches. As strategic information is revealed, other problems may be presented, necessitating a restart of the process (Nicholson, 2003b).

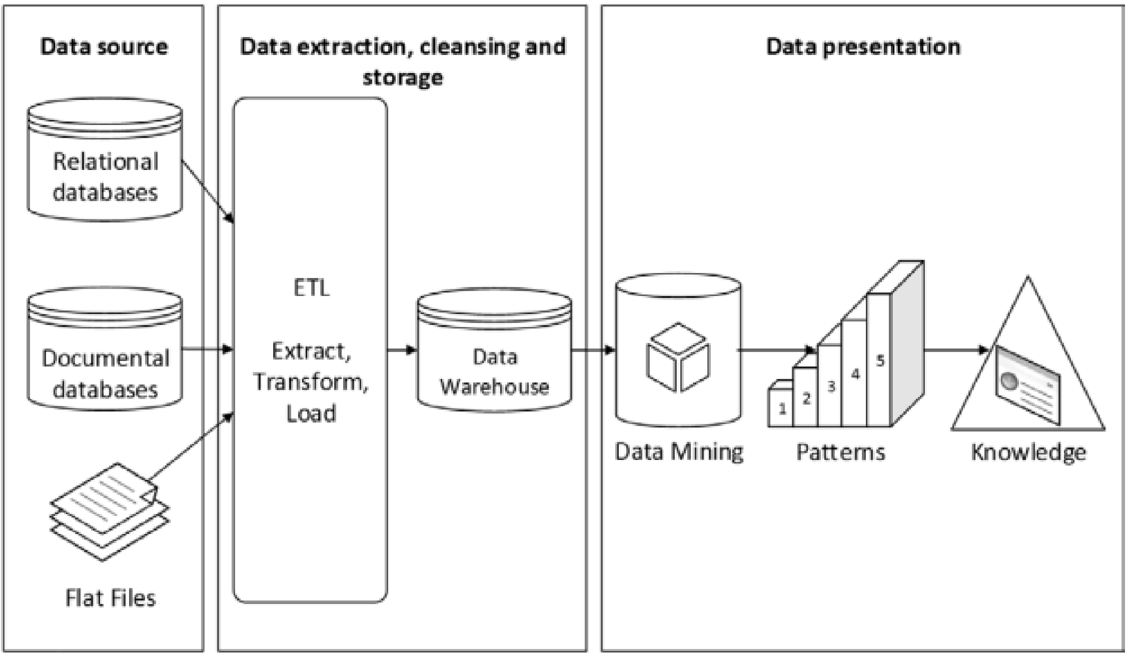

Figure 2.1: Process of Data Mining (Han et al., 2011)

When we look at the literature, it is seen that there is no estimation study about the result of the application when applying for state aid. For people who do not work in government institutions, it can be difficult to understand, interpret, request or obtain data, and analyze the obtained data. At the same time, it is an expected phenomenon that the support given on a specific issue such as energy efficiency will not be examined.

All businesses, whether SMEs or not, desire to benefit from government incentives. However, most of the time, the effort required for the application, the evaluation period and the uncertain outcome of the evaluation prevent businesses from applying for these supports. For this reason, the incentives provided by the state for the benefit of enterprises and the positive development of the economy do not find their place. With this study, businesses can predict whether their potential applications will result in a positive result by using the data of their own transactions.

3. ENERGY EFFICIENCY

3.1 Definition of Energy Efficiency

The standard of living and service quality in buildings, as well as the reduction in energy consumption per unit of service or product, can all be referred to as energy efficiency. This is true even if industry and industrial companies experience a decline in the quality and quantity of their output.

Energy is now a crucial and essential component of our daily lives. The need for energy and its utilization in daily life enhances our standard of living. As a result, there is an ever-growing need for energy. Studies on energy efficiency can help us produce our needs more quickly and more affordably.

In addition to the fact that energy is a natural and inseparable part of our daily life, the factors used in the acquisition and transmission of energy also degrade the environmental structure in an irreversible way. Difficulty in accessing resources and decreasing resources increase the cost of energy. The issues brought on by the rising energy demand have an impact on every stage of sustainable development, including the economy, the environment, and social life.

3.2 Importance of Energy Efficiency

Energy is necessary for existence and cannot be imagined without it, yet it is also finite. The availability of fuel for heating, cooling, lighting, transportation, and construction equipment is causing an increase in the demand for energy on a daily basis. The increase in requirements, such as climatic changes, is fueled by people's more contemporary lifestyles. Negative effects are also brought about by the rapid depletion of resources in response to rising needs, the degradation of the environment, and the reliance of nations on foreign sources of energy. In terms of finances, it places a heavy load on household and national budgets.

Future generations won't inherit a world that is habitable unless the modes of production and consumption habits are changed, both as producers and consumers. Because, as a result of the consumption of extremely limited resources with the ever-increasing energy need, the environmental balance is deteriorated, energy prices increase accordingly, the increase in energy prices raises the prices of the goods and services produced, and as a result, the income distribution suffers. Fossil resources that provide energy are being depleted at an increasing rate day by day. When we examine the methods of obtaining energy; it can be divided into two as non-renewable energy sources and renewable energy sources.

Minerals, fossil fuels, and nuclear energy are examples of nonrenewable energy sources. Examples of renewable energy sources include solar, wind, biomass, geothermal, hydraulic, hydrogen, and ocean energy (wave and tidal).

3.3 The Role of Energy Efficiency in Sustainable Development

Raising awareness by first fading is very important in ensuring sustainable development. Energy efficiency and savings are primarily unconscious, and the steps to support this have the quality to contribute to sustainable development. Next; There are precautions to be taken in distribution, transportation, lighting and industry. The above-mentioned causes make energy efficiency essential for sustainable growth. The focus of sustainable development, the human and environmental aspect, is immediately impacted by all manners of favorable or unfavorable developments relating to energy.

When considering how energy efficiency fits into sustainable development, it becomes clear that the work that has to be done in the areas of energy production, transmission, and consumption needs to be taken into account as a whole. This leads us to produce more energy with less primary resources and less cost. In other words, energy efficiency is doing more work with the same amount of energy or doing the same amount of work using less energy. The main objective is to produce energy more cheaply and in a way that causes less harm to the environment.

In terms of sustainable development, the goal is often for the current generation to meet their requirements in the most effective manner possible without endangering those of future generations. In fact, the question of how to meet the requirements of the current generation without endangering or reducing the opportunities for future generations to benefit from ongoing progress is one that is being sought after. Meeting current needs, repairing past resource damage and widespread environmental destruction, caring for the environment's carrying capacity, preserving and developing the renewable cores of the resources, and transferring them to the future are all examples of how economic, social, and environmental integrated development is understood. It is exceedingly difficult to restore the ecosystem after it has been destroyed without a shift in human perspective and to make it provide a good existence for future generations. If we want nature to survive, we must give it the chance to regenerate.

Each improvement in energy efficiency also reduces the production of pollution by reducing the energy output per unit and the energy input per unit for the enterprise, as it reduces energy losses. In addition, energy resources and technologies that include the life cycle offer improved efficiency opportunities by minimizing environmental impacts at the most important stages of the life cycle. Because sustainable development requires the optimal use of energy.

There is a need to make energy use more efficient not only because of the scarcity of resources and rising prices of fossil energy fuels, but also because of the protection of the environment and climate. In addition, the meaning of energy use in the economic process is great. Because energy consumption is a factor in almost every production activity. To give an example; As of 2016, Germany is one of the pioneers of energy efficiency in Europe. However, even this is not considered sufficient for Germany, and it is discussed how the country can increase its pace in this regard. Because if energy efficiency doesn't grow by at least 2.1% every year, Germany won't be able to meet its environmental goals until 2050. However, since 1990, the average energy efficiency has been about 1.8%. In 1990, the target for 2020 is 40% less, and the target for 2050 is 80% less carbon dioxide

(CO_2) emissions. Compared to 2008, it is aimed to reduce approximately 20% of primary energy consumption by 2020 and 50% by 2050. In renewable energy, it is stated that we are on the right track and that savings of up to 60% of the gross energy consumption can be achieved by 2050. For this reason, it was emphasized that Germany is already on the path to a sustainable future. (Siemens AG, Berlin und München, 2014) The following benefits of energy efficiency are included under the heading "mission" in the energy efficiency strategy created for the UK: creating a more sustainable and safe energy system, boosting economic growth, generating more jobs, lowering energy imports, and saving money on home and commercial fuel costs are all benefits of using renewable energy.

Energy efficiency will also reduce the dependence on fossil fuels with high political risk, and it will also improve the financial balance with the decrease in energy costs. The UK Secretary of State for Energy and Climate Change described energy efficiency as the heart of the low carbon economy. He also stated that energy bills can be reduced by reducing energy use and waste, and greenhouse gas emissions can be reduced by making the energy system more sustainable. (Department of Energy & Climate Change, 2012)

Gains from energy efficiency can be seen in many different areas, including reductions in greenhouse gas emissions, energy security, public finances, employment, industrial productivity, and improvements to health and welfare. In the meantime, global energy consumption is rising daily. Since the 1980s, this rise has been about 45%. An increase of about 70% is anticipated by 2030. 2016's Schneider Electric for 2040, the International Energy Agency proposes various possibilities. The three possibilities for the 13.5 billion tons of oil equivalent primary energy consumption over the world are 45%, 35%, and 12%. Up until 2040, there will be a 70% rise in the energy consumed to produce electricity, with an increase of 2% per year. In addition, 81% more primary energy is expected to be consumed by industry, accounting for 50% of the global primary energy growth rate. (2016) T.R. Ministry of Energy and Natural Resources (MENR) All aspects of sustainable development, including living circumstances, equality of opportunity, international solidarity, economic growth, productivity, competition, employment, production-consumption, and natural and cultural appearance are impacted by these increases in energy consumption. As a result, energy efficiency is a must for global sustainability. This fact necessitates a number of safeguards, which must be taken. Legal requirements, institutional design, financial support, technology (manufacturing, energy-consuming devices), and awareness-raising initiatives are some of these methods.

The benefits of increasing energy efficiency are numerous and widespread. The first of these is energy efficiency:

- Energy security,
- Greenhouse gas emissions,
- Industrial efficiency,
- Public budget,
- Health and well-being,
- Employment.

There are many different achievements in various fields. In the meantime, global energy consumption is rising daily.

3.4 Environment and Energy Efficiency

Energy efficiency and renewable energy studies have undergone significant advancements since The UN Framework Convention on Climate Change (UNFCCC) was signed in 1992 to prevent the emission of greenhouse gases from accumulating in the atmosphere, and the Kyoto Protocol in 2000 with various measures that can be followed in the Energy Efficiency Action Plan with the same target.

According to research, the use of fossil fuels in energy-related activities results in a significant portion of CO2 emissions, which makes the energy sector the first to blame for the development of greenhouse gases. Figure 3.1 (World Energy Council Turkish National Committee, 2010: 149) displays global warming emissions overall and per person.

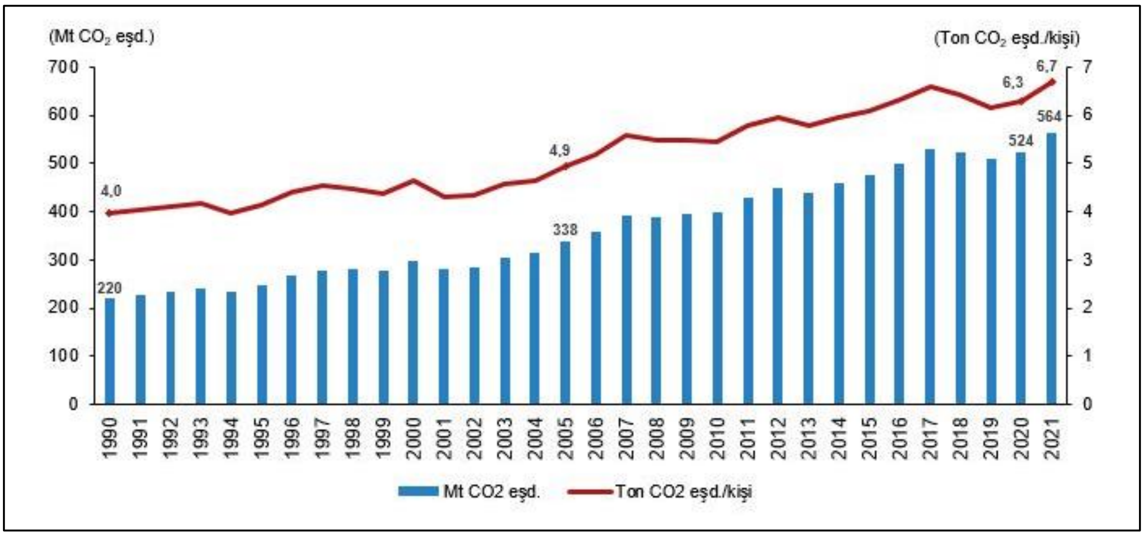

Figure 3.1: Total and Per Capita GHG Emissions, 1990-2021

The most significant contributor to climate change is generally acknowledged to be greenhouse gas emissions, which are produced in all energy-related production and consumption activities worldwide. Additionally, the burning of fuels like coal and natural gas emits pollutants that are bad for both the environment and people, in addition to greenhouse gases, and it has been noted that these pollutants have a negative impact on many aspects of human health and the ecosystem (WWF, 2011: 10).

Energy efficiency is recognized as the most cost-efficient and practical strategy to lessen the demand for energy and the environmental damage it causes. Energy efficiency is crucial in the battle against climate change and must be prioritized as the most economically advantageous strategy in order to avoid depleting natural resources to meet rising energy demand (WWF, 2011: 10).

In summary, energy efficiency and renewable energy are important concepts in terms of protecting the environment. It can be said that studies on these concepts will also contribute to the environment.

Due to factors including ongoing economic expansion, a desire to offer technical breakthroughs, a rise in the degree of welfare, and a growing population, Turkey is currently in an era of high energy consumption growth. Turkey is a country with limited energy resources and is dependent on foreign supplies, thus its main energy goal is to provide enough energy that is safe, affordable, and reliable. According to this fundamental objective, it has been suggested that effective energy use is an essential tool to accomplish this goal (DPT, 2001).

In Turkey, energy density is high and energy consumption per person is minimal. In addition, while energy intensity is decreasing all over the world, it is increasing in Turkey. All these pointed out that energy is consumed less in Turkey and that energy is not used efficiently (Çimen, 2001).

In the study conducted by Saygın (2006), it was stated that Turkey is in a similar situation to other developing countries in terms of energy. In general, it is stated that it consumes less energy and cannot use the energy it consumes efficiently and cleanly. stressed the need for research.

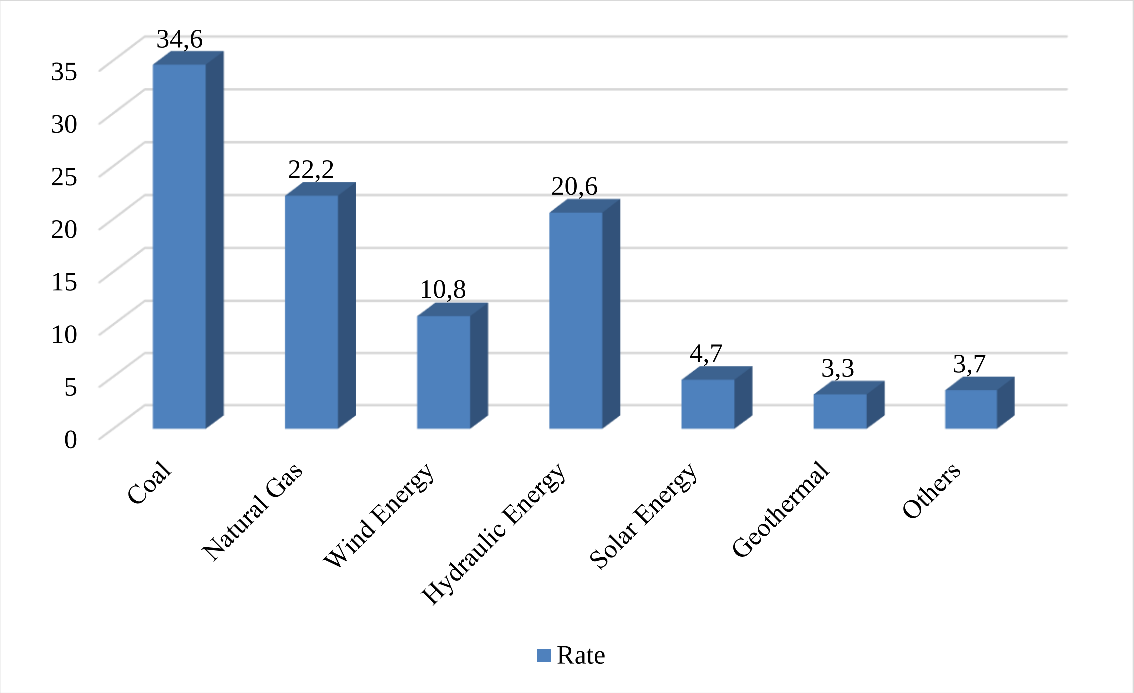

Figure 3.2: Consumption Percentages of Energy Resources in Turkey in 2022 (Ministry of Energy and Natural Resources, 2023)

As seen in Figure 10, 56.8% of the primary energy resources (Coal, Petroleum and Natural Gas) consumed by Turkey in 2022 are fossil fuel resources. These data are important because the population growth in our country, the increase in comfort standards, the increase in energy consumption with the developments in industry and technology, the increase in energy consumption also significantly increases the expenditures on resources that our country does not have, such as oil and natural gas, on which we are dependent.

According to the data of the Turkish Statistical Institute, as of 2022, Turkey's total import expenditures are 363 billion 711 million dollars. Approximately 80.5 billion dollars of these expenditures were used for energy imports. It can be stated that the share of energy imports in total import expenditures is approximately 22%. Considering that our country's current account deficit in 2022 is 48.8 billion dollars, it clearly reveals that energy should be used efficiently and that domestic-renewable energy sources should be emphasized.

If we look at the sectoral distribution of energy consumption in Turkey; According to 2021 data, approximately 24.41% of energy consumption is used in industry, 25.07% in buildings, 18.16% in transportation and 32.36% in other sectors.

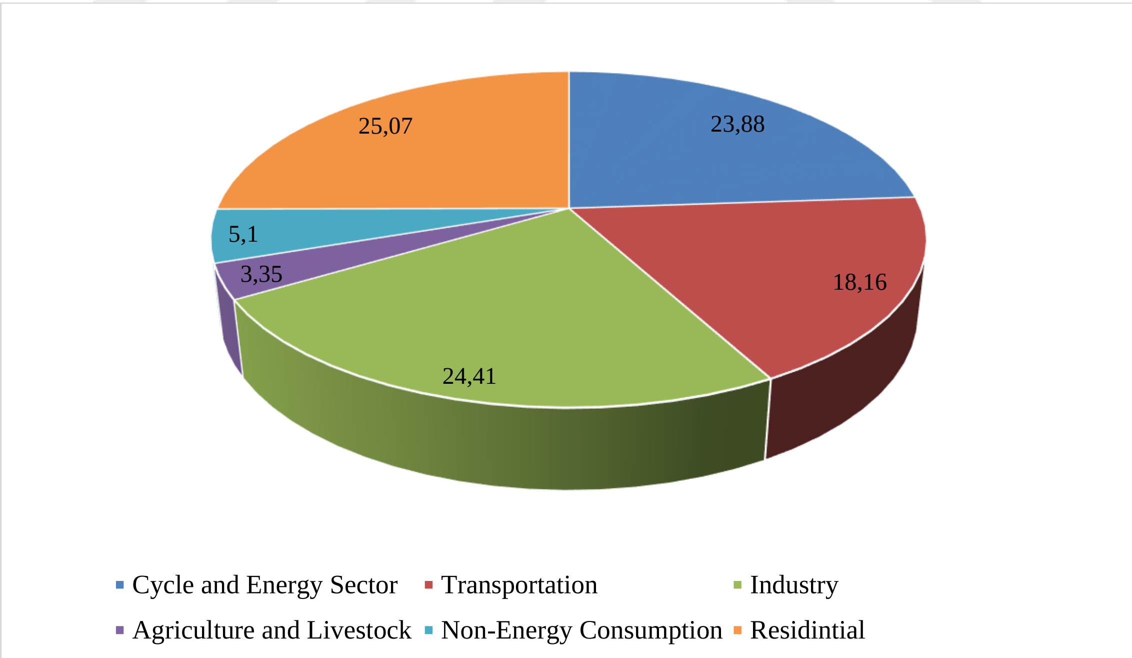

Figure 3.4: Turkey’s Energy Consumption in 2021 (Turkish Statistical Institute, 2022)

Achieving the targets set by the Ministry of Energy and Natural Resources in 2022 is of great importance. In addition, it is important to investigate this economically important concept within the scope of SMEs in Turkey and to increase the number of these studies.

3.5 Status of Energy Efficiency in Turkey

Turkey is a rapidly developing country, and in order to sustain this rapid development, it needs more and more energy every day. This requirement increases both the demand for the existing energy and its consumption.

In the analyzes made, it is estimated that a total of 635 million USD has been invested in energy efficiency for 2020 in our country, resulting in 451 KTOE primary energy savings, which is a monetary equivalent of 158 million USD. The pandemic process and its economic effects have led to the postponement of energy efficiency decisions on a global scale. It is intended to reduce Turkey's energy intensity (the amount of energy consumed per unit of national revenue) by at least 20% by 2023 compared to 2011. The National Energy Efficiency Action Plan (2017-2023), Turkey's first energy efficiency action plan, also became effective on February 1st, 2018. Up to 2023, cumulative savings of 23.9 million tons of oil equivalent (MTOE) are anticipated with the implementation of 55 actions across 6 distinct industries and an investment of USD 10.9 billion. In 2023, Turkey will use 14% less primary energy as a result of this. Up to 2033, savings of 30.2 billion dollars are anticipated. (T.R. Ministry of Natural Resources and Energy, 2022)

According to the data of 2021 among OECD countries, Turkey is one of the countries that do not use energy efficiently. However, the primary source of Turkey's current account deficit problem is energy imports.

Turkey has implemented laws and regulations to improve industrial energy efficiency. A significant development in the field of energy efficiency research is the "Regulation on Measures to be Taken by Industrial Organizations to Increase Efficiency in Energy Consumption", which was issued in the Official Gazette No. 22460 on November 11, 1995. The terms "Energy Manager" and "Energy Management Systems in Industry" originated following the publication of the legislation.

Finally, on April 18, 2007, the "Energy Efficiency Law" was passed in an effort to use energy efficiently, reduce waste, lessen the economic impact of energy expenditures, and improve the use of energy resources and energy for environmental protection. This law has made it clearer how important energy efficiency is in Turkey.

When the current sectoral composition of the Turkish industry, which is generally present in labor and energy-intensive areas, energy cost rates according to sectors, and our exports according to sectors, energy efficiency should be perceived as the struggle of Turkish industrial sectors to survive. A well-run industrial facility's effective use of energy results in a large reduction in the amount of energy purchased, lower expenses, and a higher percentage of profitability. For this reason, it is very important that everyone, from the top manager to the lowest employee, take part in studies on energy efficiency (Özdabak, 2008).

In this context, important studies on energy efficiency are carried out in our country. In the study conducted by the Technology Development Foundation of Turkey (2011), various measures are taken and dissemination activities are carried out in order to remove the financial, capacity, technology and policy barriers that prevent the widespread adoption of processes and technologies with energy efficiency in the Turkish industry, especially in recent years. For example, one of these activities is the "Increasing Energy Efficiency in Industry Project", which was actually started in 2011 with the support of the Global Environment Facility (GEF) amounting to US$ 5.9 million. The project, which is planned to be realized with a total budget of 35 million USD with the support of the GEF grant and the contributions of the implementing agency, is planned to be completed in 2015. The objectives of this project on energy efficiency are listed below;

- Development of the institutional framework and relevant legislation; Developing and disseminating the national energy management system standard,
- Increasing capacity and awareness in industrial enterprises and energy efficiency consultancy companies,
- Implementation of the energy efficiency study program,

- Realization of demonstration projects for system optimization and energy efficiency applications and
- Monitoring and evaluation of the project outputs and ensuring that the experiences gained are shared and disseminated.

It is crucial to suggest the following goals in the Ministry of Energy and Natural Resources' (2016) study on energy efficiency and saving in order to demonstrate the significance Turkey accords to these issues:

By the end of 2019, all necessary maintenance, repair, rehabilitation, and modernization projects in the public-owned electric power production plants will be finished, in line with the privatization plan,

- In the country's electricity consumption for general lighting, by the end of 2013, by replacing the existing lighting fixtures with efficient fixtures, saving at least 40% in the existing lighting poles until the end of the plan period,
- Until the conclusion of the plan term, lowering the rate of theft and loss in the distribution of electrical energy to 10%,
- Expanding existing district heating systems,
- Increasing on-site output so that it can fulfill at least 1,000 MW of total consumption by the end of 2019; and
- It is stated as the improvement in primary energy intensity in the enterprises and establishments of SEEs under the responsibility of the Ministry, according to 2013 values.

3.6 Energy Efficiency in the World

The concept of energy efficiency is a vital concept not only for Turkey but also for all countries of the world. We can show the desire of countries for economic growth as the reason for this. When the pertinent literature is read, it is clear from the previous sections that there is a connection between this goal and energy efficiency, also known as good use of energy.

Energy resources in the world are depleting rapidly. When a worldwide evaluation is made, according to estimates from the EIA (2016), the present oil reserves will endure for 42 years, the natural gas reserves for 60 years, and the coal reserves for 122 years. Fossil fuel consumption and energy resources are decreasing day by day, and high greenhouse gas emissions cause global warming and climate changes.

Energy costs in industrial enterprises are almost all of the operating costs or are very high. In industries that have done the necessary work on energy efficiency, the amount of purchased energy may decrease significantly, the costs decrease and the profit share increases. The significance of energy efficiency and the idea of it, wherever it may be found in the globe, is amply demonstrated by this situation.

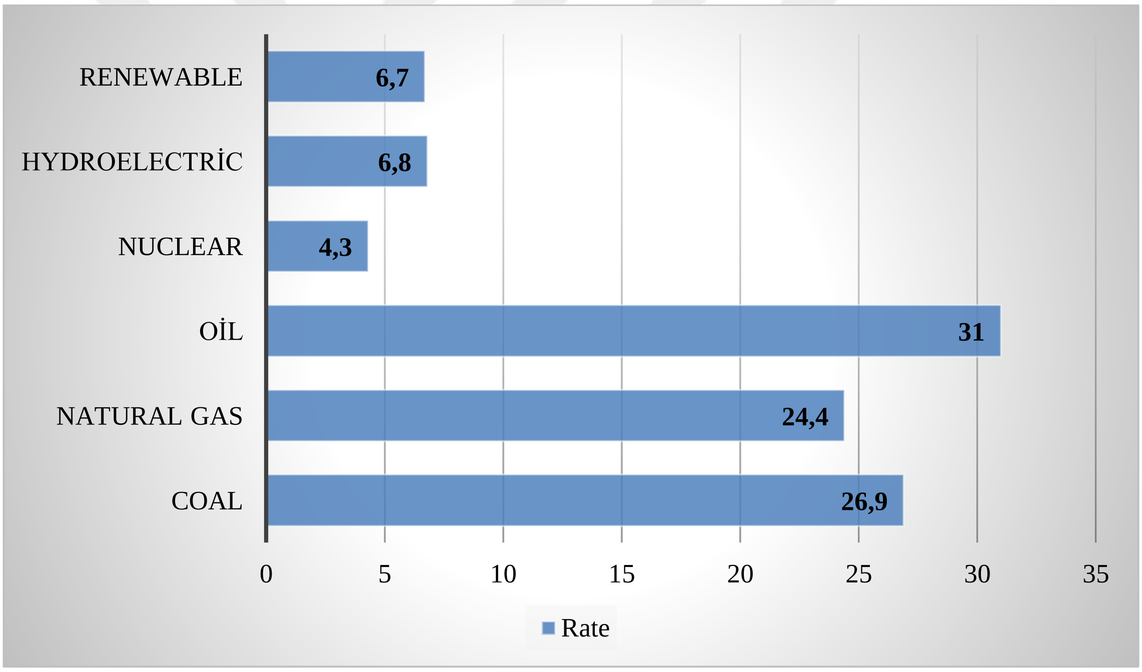

Figure 3.5: Percentages of Primary Energy Resources Consumption in the World in 2021

The most significant gain in primary energy, 31 EJ, occurred in 2021, more than making up for the substantial decline observed in 2020. In 2021, there was 8 EJ more primary energy available than in 2019. Natural gas, coal, and oil are the non-renewable energy sources that account for 86.5% of global energy consumption. It is well known that the global demand for energy derived from fossil fuels is rising.

Increasing energy efficiency is one of the most crucial issues that industrialized nations, like the United States of America (USA), pay attention to in their energy strategies. However, it's critical to focus on energy efficiency while lowering the energy density (the amount of energy required to produce a dollar's worth of gross domestic product). The objective of today's modern energy policies is not only to increase the amount of energy or electricity consumed per capita, but also to create systems that will enable the efficient use of energy and to build a structure that will produce and consume the greatest amount of energy with the least amount of energy expenditure. For instance, currently in the USA, 56% less energy is used to produce $1 of gross domestic product than it was thirty years ago. The development of effective energy policy is crucial for all nations in the world to do this (Erdabak, 2010).

Different institutions and organizations in the world make forward-looking predictions based on today's energy data and attach great importance to studies related to this. The analysis by Priddle (2002) was created in this regard by referencing the "World Energy Outlook, 2002" report released by the World Energy Agency and taking into account the anticipated developments in the global primary energy consumption and the share of energy resources through 2030. The research includes significant projections for the nations of the world in its estimate that oil, natural gas, and coal—the three fossil resources—will make up about 90% of all primary energy consumption in the future. He made important predictions with the statements that natural gas is one of the possible expectations to surpass coal after 2010.

In the study conducted by Erdabak (2010), on the condition that there is no extraordinary situation in the period between 2010 and 2030, regarding energy use in the world:

- The demand for global energy is anticipated to rise by 50%. This projection translates to a global energy consumption of 13.7 billion tons of oil equivalent in 2020, which is an increase of 2% annually on average.
- The world's energy consumption demand will be twice as high in developed countries as in developing countries.
- It is expected that the share of primary energy resources of OECD countries will decrease by 10% compared to other countries.

- In 2020, the rate of use of fossil energy resources is expected to be around 90%.
- Estimates are summarized and reported in articles.

Based on the studies in the literature, it can be concluded that energy is a power factor with its vital importance throughout the world. Another meaning of this is that the interpretation that countries that produce energy or use energy efficiently will have a strong structure together with the economic structure of the future.

4. SMALL AND MEDIUM ENTERPRISES DEVELOPMENT ORGANIZATION OF TURKEY (KOSGEB)

4.1 KOSGEB

On April 20, 1990, Law No. 3624 created the Organization for the Development of Small and Medium-Sized Enterprises. KOSGEB, the "relevant" institution of the Ministry of Industry and Technology, is a private-budget public entity that is included in Part B of Schedule II of the Public Financial Management and Control Law No. 5018.

The KOSGEB was created to increase the proportion and effectiveness of small and medium-sized enterprises (SMEs), to boost their levels of competitiveness and ability to meet the social and economic needs of the nation, and to carry out industry integration in line with economic developments.

The strategic goals of KOSGEB are;

- To improve SMEs' capacity for production and management, to create novel, high-value products and services, and to become more competitive on the global market,
- To encourage entrepreneurship, to grow entrepreneurship, and to make sure entrepreneurship is sustainable.

The goals established in accordance with these strategic goals are:

- It shall be assured that R&D innovation efforts are improved and converted into economic value.

- Financial access will be made easier.
- SMEs' management and production capabilities, culture of forming business alliances, and contribution to the rise in economic productivity levels will all be enhanced.
- By enhancing SME access to facilities and competencies, global opportunities will be secured.
- The level of support for issues with a high priority in national policy papers will increase. There will be a cultural development of entrepreneurship. Entrepreneurs with strong prospects for success, expansion, and business creation will receive support.
- Particular target groups will receive support.
- The creation of support systems for entrepreneurs will be offered.

4.2 KOBIGEL – SME Development Support Program

The program's goals are in keeping with the nation's national and international objectives and are as follows:

- To assist SME projects aimed at enhancing their competitiveness and the value they offer,
- To expand the share and contribution of SME's in the economy.

The Call for Proposals specifies the kinds of project costs that will be financed. The Project evaluation Board determines the costs that will be covered and specifies the support amount during the evaluation process for each qualifying project. The following expenses are not supported: real estate purchase, building construction, furnishing, vehicle purchase and rental, taxes, withholding, charges and fees, social security premiums.

4.3 The Program Topic "SME Energy Efficiency"

In emerging nations' economies, the manufacturing sector is extremely significant. Turkey's manufacturing sector contributes significantly to overall employment, added value, and exports. The main engine of Turkey's recent economic growth and increase in exports has been the manufacturing industry sector. For Turkey to thrive sustainably, the industrial sector needs to be strengthened and made more competitive.

Value-added and high-tech industries are anticipated to play a significant role in altering the production and international trade structures of the Turkish economy. It is foreseen that value-added and advanced technology sectors will play a leading role in changing the production and foreign trade structure of the Turkish economy. In this process, SMEs' investment, design, and skill-development efforts must be supported. increasing the level of digitalization in manufacturing as part of the transition to the fourth industrial revolution, raising the proportion of high- and medium-tech products in production and exports, switching to a production structure focused on value-added products and exports, improving the engineering and design capabilities of small- and micro-scale manufacturing businesses and ensuring their expansion, increasing energy efficiency, and going green with the project call for proposals, among other things.

With the request for bids and any further or follow-up calls that might be issued throughout the ensuing time;

- There will be an increase in the proportion of SMEs taking advantage of information technology potential in production and production logistics operations,
- There will be an increase in the number of SMEs operating in the high-value-added and advanced technology manufacturing sectors,
- The proportion of the medium-high and high-technology sectors in the total volume of production and sales will rise; The number of exporting SMEs and the volume of exports by SMEs will be increased,
- By enhancing their engineering capabilities and commercial volumes, micro-scale businesses will be able to expand up,

- SME manufacturing techniques will be developed to be energy and environmentally sustainable.

According to the call, activities that include one or more of the subjects of reducing the environmental impact of manufacturing processes, reducing the environmental impact of products throughout their life cycle, saving resources in manufacturing and disposal of wastes would be supported.

4.4 The Reason for Choosing the KOBİGEL Project in the Study

The reason why the KOBIGEL project was chosen in this study is that the program topic in 2017 was determined as "SME Energy Efficiency" Within the scope of this program, activities involving one or more of the subjects of reducing energy losses in manufacturing processes, acquiring technologies that will provide energy efficiency, installation of renewable energy systems are supported.

Success estimations of successful and unsuccessful companies applying to the program will be made by using data mining. So that the effect of the determined parameters on the probability of success will be discussed. In addition, it is aimed to be a guiding study for businesses in future project applications.

5. METHODOLOGY

5.1 Definition of Data Mining

Data mining is the process of dynamically extracting previously undiscovered, accurate, and useful information from data stacks. Clustering, learning data summarization classification rules, locating dependency networks, variability analysis, and outlier detection are just a few of the various approaches employed in this process. With the use of data mining, private information can be found in database systems with plenty of data. Statistics, mathematical fields, modeling methodologies, database technology, and numerous computer applications are used in this procedure. Since data mining is focused on the goal of studying vast amounts of data, it is strongly tied to databases. The relevant data must be kept in a way that allows quick access, and it must be obtained promptly when called upon. The goal of data warehouses, which are frequently utilized today, is to hold easier and unified overview of frequently accessed datasets. Data mining is a tool that helps the decision-making process and offers the essential knowledge to solve problems, not a solution in and of itself. Data mining aids analysts in identifying trends and connections among data collected throughout corporate operations.

Data mining is the process of using computer systems to seek for and analyze important linkages and patterns that can be used to forecast the future. Additionally, data mining is a method of data analysis that makes it possible to extract private information from database systems and aids in discovering connections between vast volumes of data by looking at their interconnections (Kalikov, 2006). These procedures have a very broad range of applications. Database systems, data visualization, artificial neural networks, statistics, machine learning, and other domains are examples of these, as shown in Figure 5.1.

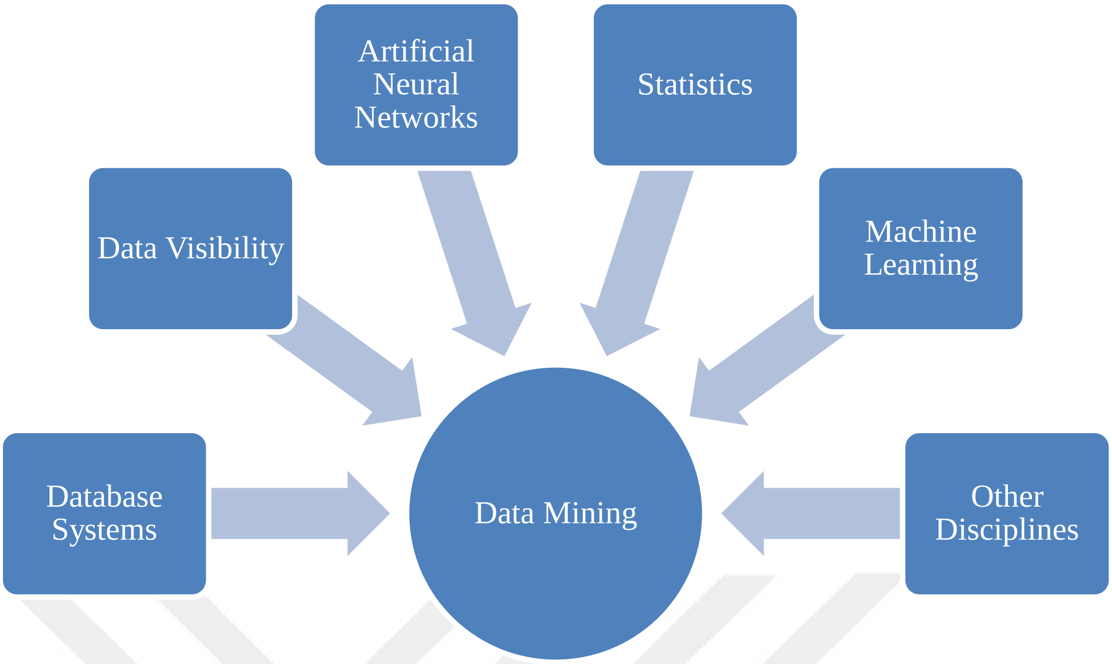

Figure 5.1: Data Mining and Other Disciplines

It is feasible to disclose the trends and behavioral patterns needed in decision support systems for firms to make more successful decisions by employing data mining techniques. Data mining has a variety of capabilities that allow for considerably more thorough and automated analysis than the previous tools that were utilized with classical decision support systems (İnan, 2003).

The ability to spot common trends and patterns of behavior between data sets is the most significant benefit data mining provides to enterprises. This procedure can also be carried out automatically. This function is frequently used, particularly in marketing campaigns that target certain populations (İnan, 2003). Another benefit is that it can make known previously hidden data that is stored in data warehouses but is initially hidden from view. For instance, by studying the items it sells or learning the connections between the things it offers, a business might influence its upcoming marketing. Finding previously unidentified data sets is the goal here.

The ability to spot common trends and patterns of behavior between data sets is the most significant benefit data mining provides to enterprises. This procedure can also be carried

out automatically. This function is frequently used, particularly in marketing campaigns that target certain populations (Nan, 2003). Another benefit is that it can make known previously hidden data that is stored in data warehouses but is initially hidden from view. For instance, by studying the items it sells or learning the connections between the things it offers, a business might influence its upcoming marketing. Finding previously unidentified data sets is the goal here.

There are numerous definitions of data mining that have been found up to this point in various sources. Some of these publications give the following definition of data mining:

- Data mining is an approach to data analysis that extracts information that raw data alone cannot (Jacobs, 1999).
- Data mining is the use of a computer software to look through massive amounts of data for connections that can help us forecast the future.
- Data mining, according to Hand (1998), is a brand-new interactive discipline that combines pattern recognition, machine learning, statistics, database technology, and secondary analysis to uncover unforeseen links in big databases.
- Data mining, according to Kitler and Wang (1998), is the process of separating out highly predictive important factors from a large pool of prospective variables.

These definitions allow for the following definition to be made: Data mining is the process of extracting and utilising relevant data from databases that include a lot of information in order to create predictions about the future that are relevant to our goals.

5.2 History of Data Mining

Nowadays, there are computers in nearly every home, and almost everyone has access to the internet. Computers can now store enormous amounts of data and analyze it much more quickly because to growing disk capacities and the ability to access information from anywhere. Data has always been interpreted, both historically and currently. To get information, equipment has been developed. Information has become transferable from the past to the present in this way. The first computers were used for counting in the 1950s. The idea of a database and data storage first appeared in the technological world in the 1960s. Scientists were able to create computers by the end of the 1960s using straightforward learning. According to Adriaans and Zantinge (1997), Minsky and Papert demonstrated that perceptrons—currently referred to as neural networks—can only learn extremely basic rules. Applications for relational database management systems first appeared in the 1970s. Computer specialists have also produced machine learning in a straightforward manner and created expert systems based on straightforward rules. Database management systems spread quickly and were employed in a variety of sectors in the 1980s, including engineering and science. Companies built databases of information about their clients, rivals, and goods throughout this time. These databases may be accessed using the SQL database query language or other similar languages, and they contain enormous amounts of data. In the 1990s, research on how to mine databases, which were accumulating data at an exponential rate, for meaningful information, began. On this, studies and publishing have begun. The process accelerated following the 1989 meeting of the KDD (IJCAI)-89 Knowledge Discovery in Databases Working Group and the 1991 publication of the KDD (Knowledge Discovery)-89 final declaration, "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop," article KDD (Knowledge Discovery) and Data Mining. Ultimately, the first data mining software was created in 1992. Data mining has continued to advance in the 2000s and has started to be used in practically every industry. The interest in this field has grown as the outcomes' advantages became apparent. Figure 5.2 depicts the historical development of data mining.

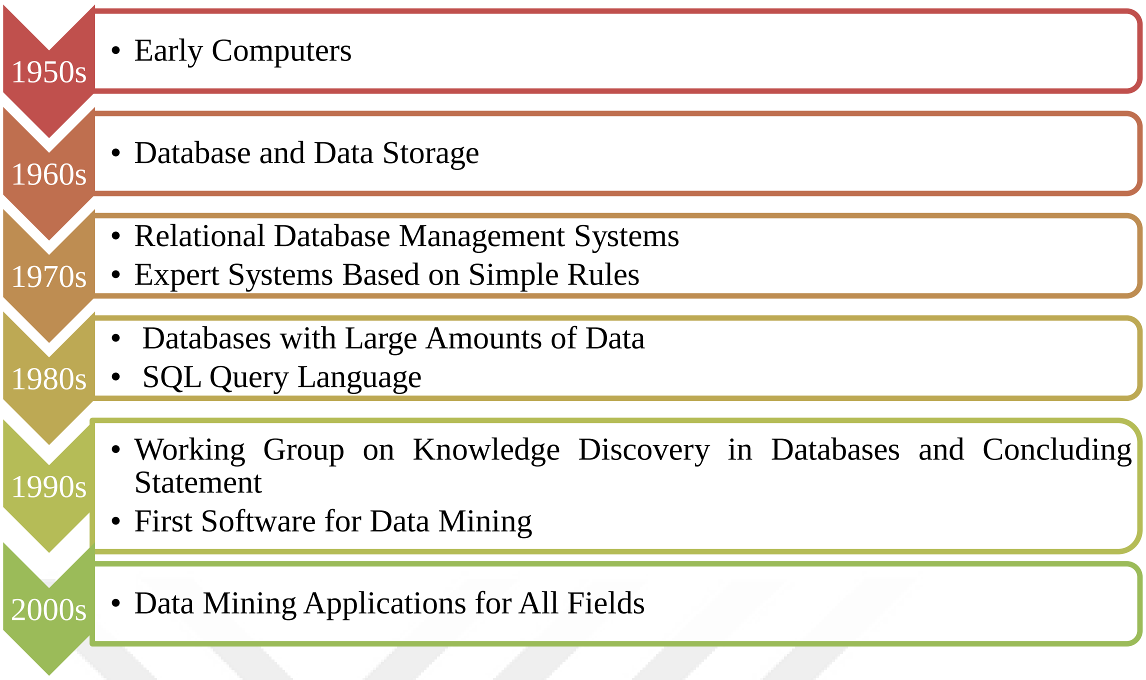

Figure 5.2 Historical Process of Data Mining

5.3 Data Mining Processes

The retrieval, compilation, organization, and subsequent analysis of data from the data warehouse mark the conclusion of the data mining process. Preprocessing the raw data may be necessary before beginning to use data mining methods. In many databases created by organizations, information is likely to be incomplete, inaccurate, repetitive and unnecessary. For this reason, after the raw data undergoes processes such as cleaning, integration, reduction and transformation, respectively, results are obtained by applying data mining algorithms. The fact that some data in the databases contain some deficiencies in terms of users prevents the data from being ready for analysis. For example, while the marital status of many people registered in the database is certain, this may be missing in some records, we can call this deficiency as missing data. Apart from that, the information in some of the records may be extreme values or incorrectly entered values; the most striking example of this would be entering a person's date of birth as 1046. Such information can be called noise or noisy data. (Han and Kamber, 2006: 89) In order to prevent these and similar problems, data pre-processing should be done. These activities might not be required if the data for the data mining application comes from a data warehouse because they are already carried out when building the data warehouse.

The data is prepared for analysis after the preparation procedures have been applied. The result is evaluated, and a pattern in the processed data is discovered by using data mining techniques. The procedures involved in data preparation can be explained as follows.

5.3.1 Identifying the Problem and Understanding the Data

Variables are chosen based on unidentified dependencies while determining the problem, and models are attempted by forming hypotheses from relationships between the data. Determining the objective for which the knowledge discovery study will be conducted and the criteria to be used to assess the level of success of the results is one of the most crucial requirements for data mining studies (Savas et Al., 2012: 8).

5.3.2 Collection of Data

With the help of preliminary processes, the data is now prepared for processing. Due to the numerous processes that must be completed, the data preparation phase takes up 50% to 80% of the time and effort required for the total data discovery process (Piramithu, 1998: 294).

Data mining places a high value on data quality. Data should be pre-processed to improve the study's dependability. Otherwise, the analyses' decision-makers' judgements will be negatively impacted by the usage of inaccurate inputs. Data preparation is a time-consuming stage in data mining that is frequently semi-automatic. Effective strategies for automatic data preprocessing have become more significant due to the increase in data and, consequently, the need to preprocess a huge amount of data (Oğuzlar, 2003: 70). Oğuzlar (2003) asserts that three factors are crucial during the preprocessing stage of data (Oğuzlar, 2003: 70). These;

- Fixing data issues that will inhibit data analysis,
- Recognizing data structure and carrying out significant data analysis,
- Taking actionable information out of the data set,

According to Famili et al. (1997), it is crucial to identify the kind of preprocessing because many applications require more than one type of data preprocessing. There are numerous data preprocessing strategies, it becomes clear when the literature is investigated. One of these, data cleaning, is used to eliminate noise and fix data discrepancies. Data integration is a separate method that merges data from several sources into an appropriate database. It is possible to use data transformation techniques like normalization. The goal of data reduction is to eliminate some of the extraneous variables, merge them, or minimize the amount of data by clustering. These aforementioned data preprocessing techniques are applied before data mining, increasing the quality of the results and/or saving time by reducing the time spent for data mining (Han and Kamber, 2006: 105). Prior to data mining, the aforementioned data pretreatment procedures are used, which improves the results' quality and/or frees up time by cutting down on the amount of time needed for data mining (Han and Kamber, 2006: 105).

5.3.3 Model Building

The foundation of data mining is model development. to properly configure the model. The purpose of the project must be clearly known. There are various algorithms available for every use. The algorithm that produces the best accurate answer is utilized in this situation after all suitable algorithms have been performed on the data at hand.

5.3.4 Model Evaluation

It is a method to assess the efficiency and caliber of one or more models used in the modelling phase as well as the model's compliance with the goals of the first phase. The evaluation process includes reviewing the employed data mining techniques and selecting how to apply the findings. The model needs to be in line with the stated goals and examined to see if it covers all pertinent ground. It is also decided at this time which specific data can be used in the future and whether the study site is suitable.

5.3.5 Scoring

The first step in every data mining effort is to look for trends. The model will mostly be used for evaluation. In data mining jargon, evaluation is also known as scoring. This requires a data collection with both novel scenarios and the trained model in order to be possible. In this manner, educated guesswork can be made for novel scenarios.

5.4 Data Mining Methods

All data science approaches, from sophisticated machine learning algorithms to categorization techniques, can be included in data mining strategies.

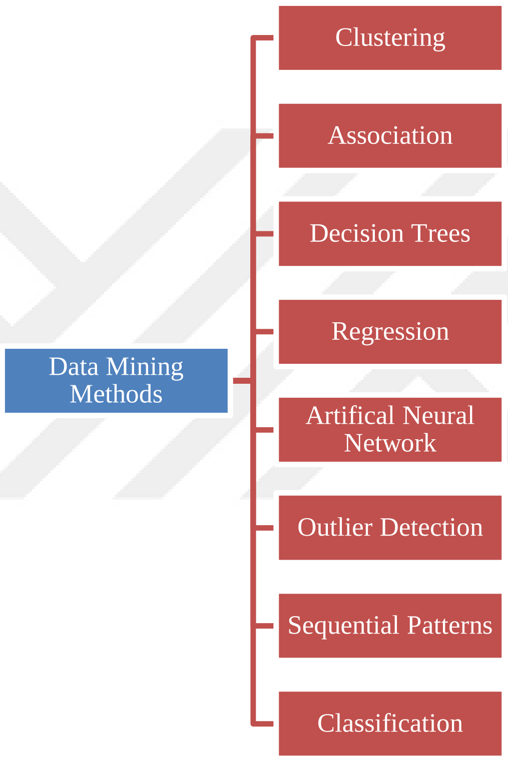

Figure 5.3: Data Mining Methods

5.4.1 Clustering

Clustering is the process of separating data into classes or clusters Karypis and Kumar, 1999). While the elements in the same set are similar to each other, they are different Data are divided into classes or clusters by the process of clustering (Karypis and Kumar, 1999). The elements of one set are distinct from those of other sets even though they are similar to one another. In several data mining disciplines, including statistics, biology, and machine learning, clustering is commonly employed. The classification model's data classes are absent from the clustering model (Ramkumar and Swami, 1998). No data class exists. The classes of the data are known in the classification model, and it can be predicted from which class fresh data may come. In contrast, the clustering model puts together data that lacks classes. Occasionally, the clustering model is used. It can serve as the classification model's preprocessor (Ramkumar and Swami, 1998). Typical clustering applications include categorizing plants and animals according to their functions in biology, identifying different customer groups in markets and revealing their shopping habits, classifying similar genes according to their functions, and categorizing houses according to their types, values, and locations in city planning. In order to categorize documents for knowledge discovery on the Web, clustering can also be used (Seidman, 2001). Data clustering is progressing well. Cluster analysis has lately emerged as a hot topic in data mining research, in conjunction with the growth in the amount of data being collected in databases.

5.4.2 Association

In order to identify events that frequently occur together, market researchers frequently use association, which searches big databases for interesting associations between variables. For instance, knowing that women between the ages of 40 and 50 love to purchase black things, product designers can incorporate this hue into a new product line. Using attribution analysis, retailers can identify product combinations that customers frequently buy together.

5.4.3 Decision Trees

Modeling methods such as decision trees provide predictions based on a set of binary rules. With the same input, a decision tree algorithm generates the same output. Regression analysis, classification models, and decision trees are all constructed using it.

5.4.4 Regression

In order to forecast continuous values depending on other dataset variables, regression analysis is performed. Regression analysis, for instance, can be used to forecast a product's future price based on demand, supply, and other variables. Linear regression and logistic regression are the two most popular regression methods.

5.4.4.1 Linear Regression

By examining other variables, it determines an unknown variable's value. For instance, the market worth of another firm can be calculated based on geography, industry, or a prospective selling date using information about businesses that have recently sold (enterprise type, location, size, sales price, date of sale, etc.).

5.4.4.2 Logistic Regression

This approach is useful for determining whether a variable favors a specific result. The variable must be "binary" for logistic regression to function. In other words, it's important to look at how a variable's existence or absence influences a "yes" or "no" response.

5.4.5 Artifical Neural Network

The human brain serves as an inspiration for modeling neural networks. It may be very intricate. Building and implementing neural networks may need companies hiring highly skilled personnel. It can be applied to circumstances that call for prompt answers, like driverless car technology.

5.4.6 Outlier Detection

You can identify anomalies in the data using outlier detection. A value or group of values that considerably deviates from predicted patterns is referred to as an anomaly or outlier. It is especially helpful for fraud detection, tracking cyberattacks, and keeping tabs on system performance.

5.4.7 Sequential Patterns

Sequential Diagram Sequential events are referred to as mining. It is frequently used to analyze transactional statistics and can be helpful in deciphering consumer behavior. It can be used to boost prospects for sales and product suggestions.

5.4.8 Classification

One of the simplest data mining approaches is classification analysis, which categorizes data. To be able to forecast behavior or provide an important response, classification analysis is done. For instance, a credit card provider might classify customers as "low risk," "medium risk," or "high risk" based on their purchasing history and annual income when determining which users in its database should receive a credit card offer.

5.4.8.1 LightGBM

A histogram-based method is used in LightGBM. Making the variables with continuous values discrete lowers the computing cost. The computation and consequently the number of divisions directly relate to the training time of the decision trees. With the help of this technique, training time and resource utilization are both decreased.

In learning decision trees, there are two methods that can be applied: level-wise, depth-wise, or leaf-wise. The balance of the tree is preserved as the tree expands in the level-oriented technique. On the other hand, division from leaves, which lowers loss, continues in the leaf-focused strategy. This characteristic sets LightGBM apart from other boosting methods. With the leaf-oriented technique, the model learns more quickly and with a lower error rate. In situations where there are few data points, the model is vulnerable to overlearning because to the leaf-oriented growth strategy. The technique is therefore better suited for application in huge data. To avoid over-learning, variables like tree depth and leaf count can also be tuned.

To avoid over-learning and to speed up learning, LightGBM's learning_rate, max_dept, num_leaves, and min_data_in_leaf parameters can be improved. The feature_fraction, bagging_fraction, and num_iteration parameters can be optimized to reduce learning time. Optimization of parameters includes:

- Num_leaves is the quantity of leaves on the tree. It is the most crucial variable used to regulate the tree's complexity. To prevent overlearning, it should be less than 2(max_dept). For instance, setting num_leaves to 127 when max_depth is 7 may result in overlearning. It could be more accurate to set it at 70 or 80.

- Max_dept is used to limit the depth of the tree to be installed. It should be optimized to avoid over-learning. Too much branching will cause over-learning, and too little branching will cause under-learning.

- Min_data_in_leaf is one of the important parameters to use to prevent over-learning. Its ideal value is influenced by num_leaves and data size. If you set it to a high value, the tree might not grow and you might only learn some things.

- Learning_rate is a value between 0-1 for scaling established trees. A smaller value will help better predictive power. However, it will increase learning time and increase the possibility of over-learning.

- Feature_fraction, variable to be used in each iteration; bagging_fraction are parameters where the number of data to be used in each iteration can be set. Num_iteration is the number of iterations to be made in the learning process. feature_fraction, bagging_fraction and num_iteration numbers are directly related to learning time. The lower these numbers are, the shorter the learning period will be. Paying close attention to incomplete learning is crucial, though. Numerous trials can be performed to determine the ideal quantity.

LightGBM employs two distinct methods. These deal with the quantity of data samples and variables and are known as the Gradient-Based Unidirectional Sampling and Custom Variable Package.

5.4.8.1.1 Gradient-based One-Side Sampling (GOSS)

GOSS tries to decrease the amount of data while keeping decision trees' accuracy. GOSS employs only the crucial data while traditional gradient boosting scans all data samples to determine the information gain for each feature. As a result, the number of data is decreased without significantly changing the distribution of the data.

5.4.8.1.2 Exclusive Feature Bundling (EFB)

EFB seeks to decrease the number of variables while maintaining accuracy rates, hence improving the effectiveness of model training. EFB processes data in two phases. With these, packages are created and variables are combined into one package. EFB combines sparse features to produce denser features. As a result, training becomes simpler and goes faster while using less memory.

In conclusion, while EFB aggregates variables to minimize dimensionality, GOSS reduces data size to compute information gain by removing data that may be viewed as less important. LightGBM improves the effectiveness of the training process with these two features.

5.4.8.2 Naive Bayes Classification

Naive Bayes classification uses a sequence of calculations that are defined in accordance with probability principles to identify the class, or category, of the data supplied to the system.
In Naive Bayes classification, the machine is given a certain amount of taught data. A class or category must be assigned to any data given for instruction. When new test data are input into the system, they are processed in accordance with the probability operations performed on the taught data in an effort to identify the category to which the given test

data belongs. The more taught data there are, the more precisely the genuine category of test data may be determined.

5.4.8.3 K-Nearest Neighbor

It is defined as the classification of the data that has not yet been assigned to a class by placing it in the best (optimal) class based on the distance determined by comparing the data of the unknown class with the other data in the training set and performing a distance measurement. It is described as a supervised machine learning technique in which the k value (similarity) is used to determine the class (learning set) and nearest neighbor (element) of the sample data point to be classified.

5.4.8.4 Support Vector Machines

A supervised learning technique known as a Support Vector Machines can be applied to classification, regression, and outlier identification tasks. They are predicated on the notion of locating a hyperplane in a high-dimensional space that best segregates the various classes.

The hyperplane with the largest margin, or distance, between the nearest data points for each class is found using the procedure. Support vectors are used to characterize the hyperplane and are the data points nearest to the hyperplane.

SVM are frequently employed in classification tasks, where the goal is to predict the class of a given data point based on its features. For instance, depending on the words and phrases in the email, a Support Vector Machine can be used to categorize emails as spam or non-spam.

In regression jobs where the objective is to predict a continuous value from a set of input properties, SVM are also employed. A SVM, for instance, can be used to calculate the cost of a house based on its size, location, and other features.

SVM are widely utilized in a variety of domains and are a potent tool for supervised learning tasks like classification and regression.

5.5 Data Mining Metrics

The results of a classification problem or procedure must be assessed using several evaluation criteria in order to determine its effectiveness. Data mining metrics are primarily concerned with the accuracy, usability, and dependability of the data. Accuracy is determined by how useful a pattern is to a user. The ease of use of the pattern and the data processing technique both affect how accurate a measurement is. It is the most important step in data mining for better outcomes and patterns that could be applied to future decision-making.

5.5.1 Confusion Matrix

The classification problem's prediction outcomes are summarized in a table called the confusion matrix. The confusion matrix can be used to get a sense of the classification mistakes that were made. The confusion matrix also provides insight into the different kinds of mistakes the model makes. The performance of the model in classification issues is typically described in the literature using the confusion matrix. The confusion matrix will appear like the table below if we evaluate for binary classification models.

Table 5.1: Confusion Matrix

	Predicted	
Actual	True Positives (TP)	False Positives (FP)
	False Negatives (FN)	True Negatives (TN)

- **Accuracy:** This metric measures how accurately the system predicts outcomes compared to all other forecasts. Accuracy is not a reliable gauge of a classifier's real performance because it will yield inaccurate results when the number of samples in the various classes varies widely (unbalanced target). For instance, the classifier might be biased to identify all the samples as cats if the data set had 95 cats but only 5 dogs. In actual use, the classifier's accuracy would be 95% overall with a 100% identification rate for the cat class and a 0% recognition rate for the dog class.

When TP is True Positive, TN is True Negative, FN is False Negative, FP is False Positive;

Accuracy = (TP+TN) / (TP+TN+FN+FP)

- **Recall:** Indicates how successfully positive states were predicted.

Recall = TP / (TP + FN)

- **Precision:** A condition in which success in a favorable scenario is indicated.

Precision = TP / (TP + FP)

- **F1 Score:** The harmonic average of the Precision and Sensitivity Recall values is displayed by the F1 Score value. Since extreme cases shouldn't be disregarded, the harmonic mean was chosen instead of the simple mean.

F1 Score = 2*Precision*Recall / (Precision + Recall)

5.6 Data Mining with Python

One of the most popular programming languages is Python. Guido van Rossum created it, and it debuted in 1991. Powerful and simple to learn, Python is a programming language. High-level data structures are useful, and approaches to object-oriented programming are straightforward but efficient. Python is the perfect language for scripting and quick application development on a number of platforms thanks to its lovely syntax, dynamic typing, and interpretable nature.

Python is frequently used to build websites and applications, automate time-consuming operations, and analyze and present data. Because it is relatively easy to learn, Python has been used by many non-programmers, including accountants and scientists, for a variety of regular activities including managing finances.

Data mining is a highly helpful and practical method that aids businesses in formulating plans based on pertinent data information. Analytics initiatives across sectors (including education, media, banking, insurance, manufacturing, etc.) are built on data mining. Due to its ease of use and wealth of potent modules, Python is a flexible tool for data mining and analysis, particularly for those searching through mountains of data for hidden nuggets.

5.7 Data Mining Usage Areas

Data mining is used in a variety of industries today, including telecommunications and banking. Here are a few of these.

5.7.1 Banking – Finance

Data mining is mostly utilized in the banking industry for customer management, profitability analysis, trend analysis, and loan fraud detection. In financial markets, it is used in situations such as portfolio management, estimating asset prices and even forecasting financial crises.

Estimating loan repayments and analyzing customer loans is an important issue for a bank. With estimation methods, it is investigated with what probability the person will default or whether the loan requesters are suitable for loan. As a result, customer-specific strategies can be determined. For customer management, customers are clustered according to their specific characteristics and each Marketing can be done by creating a separate proposal for the group.

5.7.2 Retail Selling

Data mining is a powerful tool for the retail industry that reaches the end consumer directly. Companies in the sector can collect data for customer management through plastic cards, store all kinds of information and shopping of customers in data warehouse infrastructures, and design personalized, targeted campaigns; For this, it is necessary to use clustering methods as in banks. Understanding the characteristics of the customers and assigning them to the determined lifestyles or segmenting them according to their value, scoring according to the frequency of visiting the stores, determining the customer lifespan with methods such as classification algorithms, regression, estimating those with a tendency to passivity, determining the returns of the customers or a wide variety of analyzes such as association detection can be used to capture products bought together.

With these studies, companies; It can be seen who are the most valuable and most profitable customers, which products they should sell with which shelf order, what is the most preferred product or service, what kind of campaigns its customers like, and accordingly, it can provide many benefits from budget planning to target setting.

The aforementioned analyses' primary goals are to comprehend the client, his consumption patterns, and offer suggestions to him. As a result, the data received from the customer turns into campaigns to satisfy and retain him, and for companies, it can be used as a tool to increase profitability by extracting information.

5.7.3 Telecommunication

The telecommunication sector has differentiated its service content over time and has evolved from providing only local and long-distance telephone services, and transformed into a sector that provides the infrastructure for data transfer via fax, internet access, mobile phone and other data traffic. Additionally, the sector is expanding even faster and becoming more competitive due to the restructuring of the telecommunications industry in some nations and the development of new computers and communication technologies. At this point, data mining has gained importance in terms of understanding the business context, defining communication patterns, catching frauds, using data sources better and increasing service quality.

Telecom companies can predict the communication patterns that people will create later by making association and sequential pattern analyzes with dimensions such as call duration, location, call time, call type they have in their infrastructure. Apart from this, they can search whether their customers have a tendency to passivity or not by using various classification methods and get the chance to catch them before they become passivity, and they can organize customer management campaigns specific to the determined segments by clustering people according to their characteristics.

In addition to the most widely used industries of banking, finance, telecommunications, and retail, data mining is also employed in a wide range of other fields, including astronomy, biology, insurance, medical, engineering, and many more. Classification of celestial bodies in astronomy, detailed description of gene structures in biology, cancerous in medicine understanding and classifying cells can be given as examples of these applications.

6. APPLICATION

The project applications in this area have been classified using an algorithm. The algorithms were applied to the data set, comparisons were done, and the best algorithm was identified. It has been determined that the most accurate technique for this study is the classification technique.

6.1 Dataset Description

The review is done in depth around a determined case study. As in other studies, data are collected systematically and the relationship between the variables is tried to be found. It is possible to examine oral, textual, and other information objectively and methodically using content analysis, a scientific methodology. The process of summarizing and defining the fundamental components of written information and the messages they contain is referred to as content analysis. The themes that describe the data should next be decided after conceptualizing the acquired data and organizing it logically in accordance with the developing concepts.

A four-person project evaluation board played a role in the evaluation phase of the applications. Two of the four people in question are academics appointed by universities, and two of them are KOSGEB administrative chiefs. Projects have been subjected to a qualitative evaluation in general by this committee. However, the board members clearly take into account the quantitative characteristics of the enterprises that they have written on the application form while making this evaluation. With this study, it has been tried to make an estimation of how effective the business data in the project is on the approval of the project.

177 enterprises applied to the program on "Energy Efficiency", whose applications were received by KOSGEB in 2017. 43 of these applications were approved by KOSGEB. The applications of 134 enterprises were rejected. 12 of the information of the enterprises in the project application were used within the scope of the study. These; the status of the enterprise, the number of partners, the number of years it has been operating, the number of existing personnel, the number of white collar personnel, the number of blue collar personnel, the project budget, the part of the project to be met with equity, the enterprise's equity, domestic sales, foreign sales, administrative expenses.

In terms of the number of years in which the applicant businesses have been operating, it is seen that 32 of them have been operating between 0-3 years. While 33 of them are in operation between 4-7 years, 22 of them are in operation between 8-10 years. 23 of them have been operating for 11-15 years and 67 of them have been in operation for more than 15 years. Figure 6.1 shows the operating years of the applicant companies.

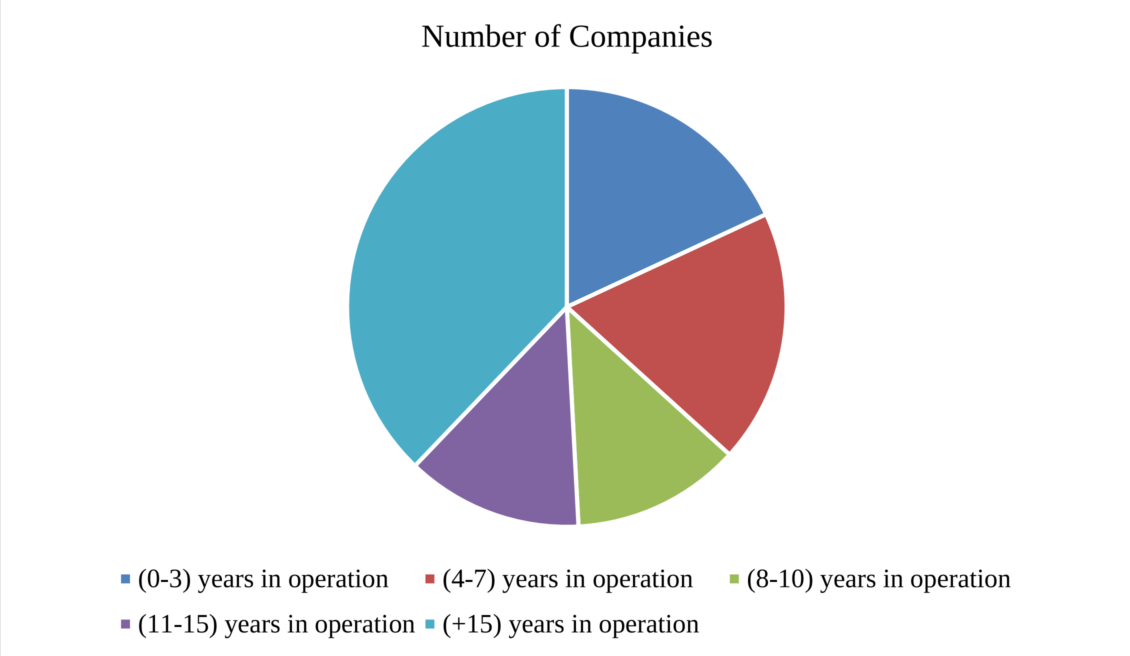

Figure 6.1: The Operating Years of the Applicant Companies

In terms of the number of years in which the approved businesses have been operating, it is seen that 3 of them have been operating between 0-3 years. While 7 of them are in operation between 4-7 years, 1 of them is in operation between 8-10 years. 4 of them have been operating for 11-15 years and 28 of them have been operating for more than 15 years. Figure 6.2 shows the operating years of the approved companies.

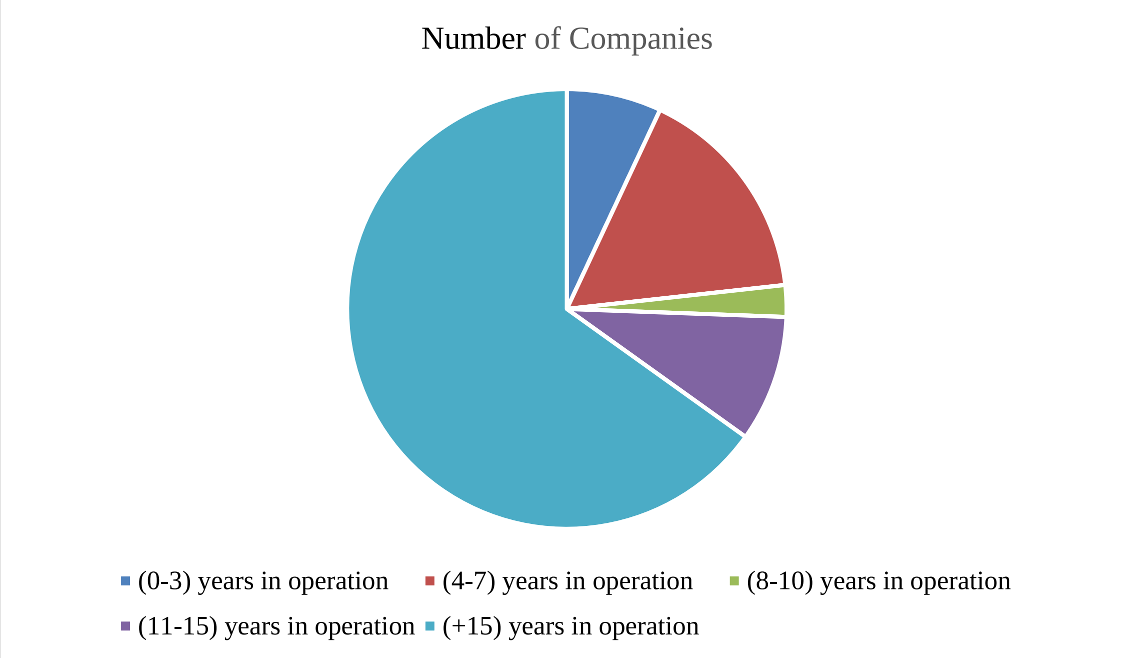

Figure 6.2: The Operating Years of the Approved Companies

41 of the applicant enterprises are joint stock companies, 106 of them are limited liability companies and 30 of them are sole proprietorships. Of the approved businesses, 23 are joint stock companies, 19 are limited liability companies, and 1 is a sole proprietorship.

Table 6.1: The Status of the Enterprises

The Status of the Enterprise	Number of Applicants	Approved	Approval Percentage	Rejected	Percentage of Rejection
Sole Proprietorship	30	1	%3,33	29	%96,66
Limited Liability Company	106	19	%17,92	87	%82,07
Joint Stock Company	41	23	%56,09	18	%43,91

The five businesses that submitted applications did not have any white-collar workers, according to an analysis of the ratio of white-collar workers to all other employees. 53 companies have a lower than 10% white-collar employee to total staff ratio. Between 10%

and 20% of 48 firms' total workforce is made up of white-collar workers. This ratio is between 20 percent and 40 percent for 66 enterprises and greater than 40 percent for 10 enterprises. This rate is above 10 percent in 34 of 43 enterprises whose projects have been approved.

Table 6.2: The Ratio of the Number of White-Collar Personnel

The Ratio of the Number of White-Collar Personnel (%)	Number of Applicants	Approved	Approval Percentage	Rejected	Percentage of Rejection
<10	53	9	%16,98	44	%83,02
10-20	48	10	%20	38	%80
20-40	66	20	%30	46	%70
>40	10	4	%40	6	%60

The project budget of the applicant enterprises is shown in Figure 6.3.

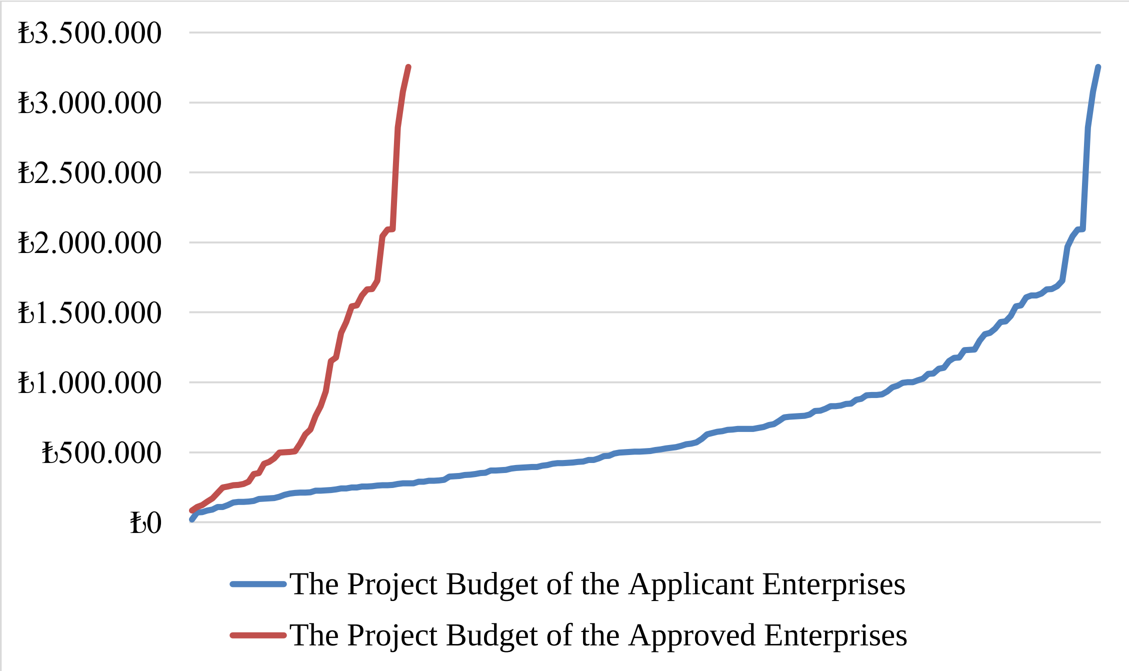

Figure 6.3: The Project Budget of the Applicant Enterprises and Approved Enterprises

6.2 Preparing the Data for Application

The study made use of the scikit learn library's methods. Testing will only use 15% of the data. The remainder was used for exercise. Figure 6.4 displays the status of the project approval.

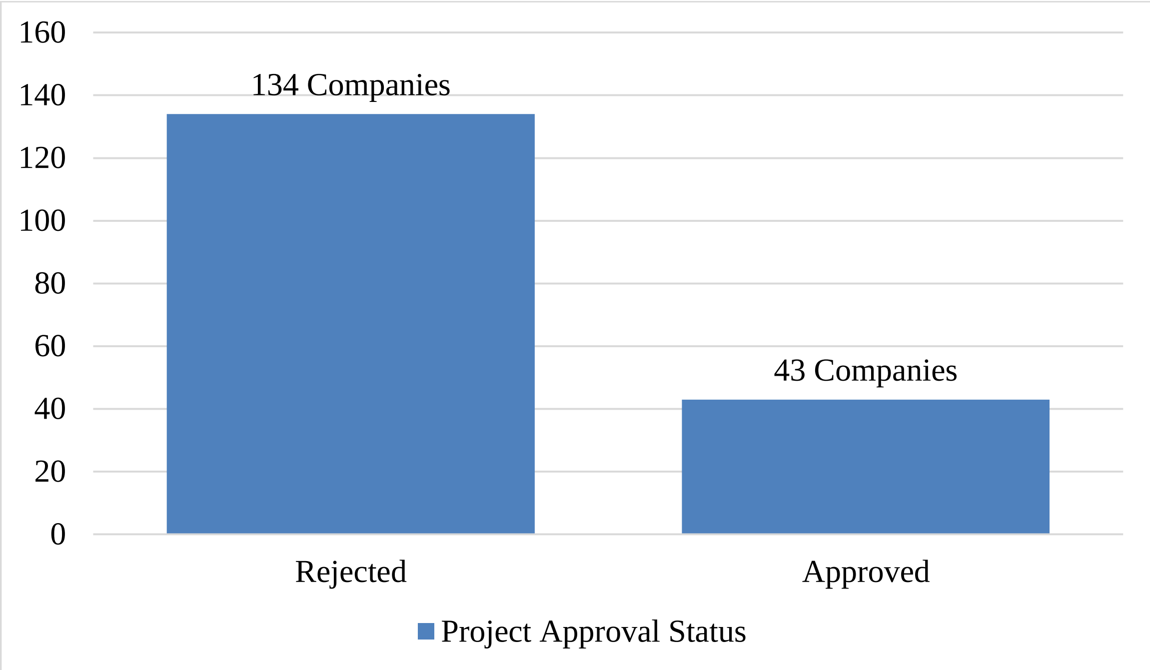

Figure 6.4: The Project Approval Status

6.2.1 Train and Test Split

The first dataset used to train machine learning algorithms is the training data. These data are used by models to develop and improve their rules. The machine learning model's primary input is training data. It imparts expectations to machines. To fully comprehend and grasp the requirements, the model performs an iterative analysis of the dataset.

Test data are used to assess the model's efficacy or performance. This sample of data offers an objective assessment of the final model that fits the training set. To assess how well the training or model performs, one turns to the test dataset.

When tested on the same dataset that the machine learning algorithms were trained on, the model is likely to perform well. It is crucial to assess the accuracy of the model using a test dataset with hidden data points in order to observe how the trained agent responds to challenges in the actual world.

6.2.2 Label Encoding in Python

In data science and machine learning processes, rewriting the categorical variable types (data in text/string format such as texts, sentences, etc.) in data sets in a format that can be understood by machines is called Encoding. Categorical variables are converted to numerical representations to make them workable by algorithms. Encoding is making changes to the way variables are represented.

The classes of a categorical variable are labels. Label encoding is one of the methods of encoding these classes in a way that makes them more useful and understandable in models. If there are two classes of a categorical variable to be label encoded, the process to be done is specifically called binary encoding (binary character encoding). For example, it is a label encoding method to express the status of being male/female represented by string expressions using the values 0 and 1. In this study, this process has been done for non-numerical business status and project approval status.

6.2.3 Feature Scaling

Before inserting the data into machine learning algorithms, while preparing it for the model, one of the preliminary operations that must be done on the data is standard scaling. When the numerical data are too different from one another, the method's goal is to compress the data into a tighter range. Feature Scaling is also known as Min-Max Normalization. The methods used in statistics to associate numbers of very different sizes with each other or to make calculations between each other are called normalization. This process was applied to all columns except the columns showing the status of the enterprise and the approval status of the project. Feature scaling is found with a formula as shown below.

$$x' = \frac{x - \bar{x}}{\sigma}$$

6.3 Implementation of Methods

In the study, the results obtained as a result of the analyzes made on the project applications of 177 enterprises were evaluated. Different performance metrics were calculated with lightgbm, logistic regression and support vector machine techniques used in the analysis.

6.3.1 Implementation of LightGBM

As a component of the Microsoft DMTK (Distributed Machine Learning Toolkit) project, the boosting method LightGBM was created in 2017. Its benefits over other boosting methods include fast processing, handling large amounts of data, using fewer resources (RAM), a high rate of prediction, parallel learning, and GPU learning support. Tables 6.3 and Figure 6.5 display the outcome of applying the data to the LightGBM model.

Table 6.3: LightGBM Result

F1 Score	0.93
Accuracy Score	0.88

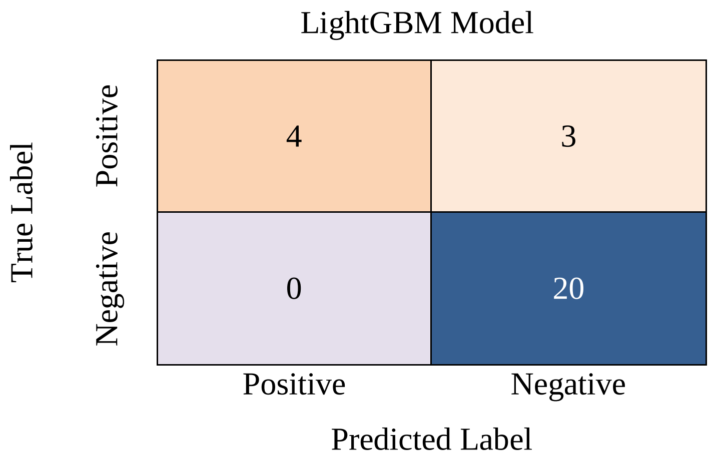

Figure 6.5: LightGBM Confusion Matrix

A high F1 score and a good accuracy score were obtained with the application of the LightGBM model in the study. The high F1 score indicated that the model achieved a successful balance, taking into account both accuracy and recall rate. The accuracy score is at an acceptable level. This shows that the model makes correct classifications in general. These results show that the LightGBM model is a strong option to use. A high F1 score indicates that the model is capable of making accurate classifications and can be effective in achieving the target of predicting acceptance of application for energy efficiency project grants.

The study examines the LightGBM elements that have the most influence on project approval success as well as the aspects that have the least. The domestic sales of the business are one of LightGBM's top criteria for project application success. The characteristics that least affect the success status are respectively the status of the enterprise, the foreign sales, the number of the existing personnel. Figure 6.6 shows how effective each feature is shown below.

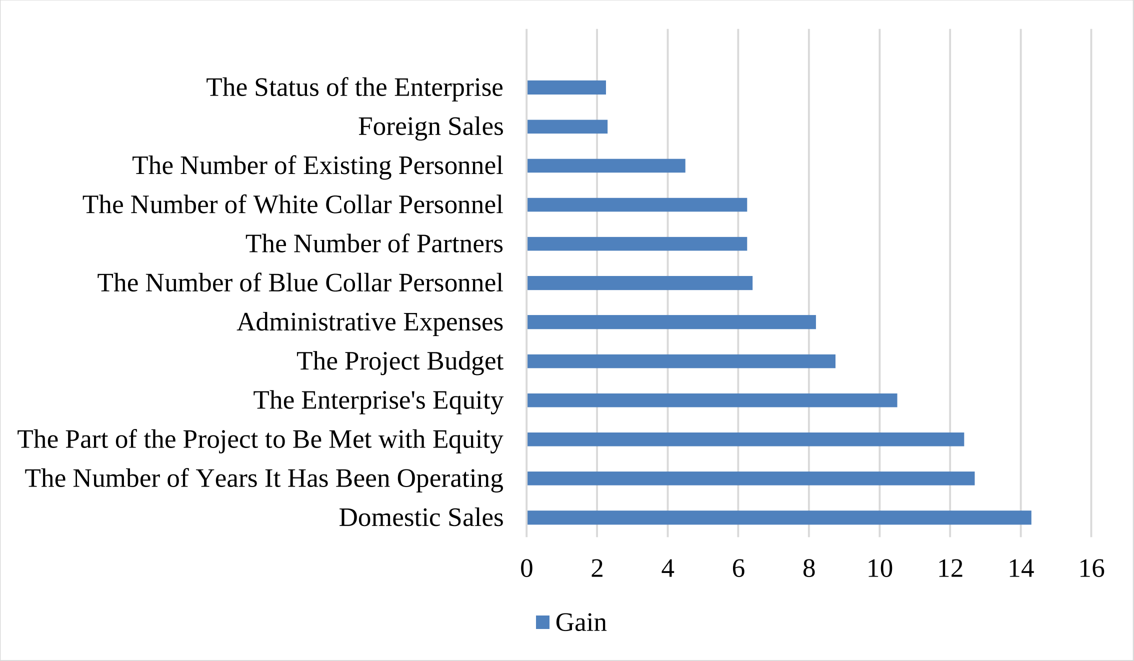

Figure 6.6.: Importance of Features in LightGBM

6.3.2 Implementation of Logistic Regression

Logistic Using a set of variables, regression analysis often seeks to build a model that predicts the value of a variable. The results of the study are used to compute the values' realization probability, which is subsequently utilized to categorize data. The results of the regression analysis' classification are shown in Table 6.4 and Figure 6.7.

Table 6.4: Logistic Regression Result

F1 Score	0.88
Accuracy Score	0.81

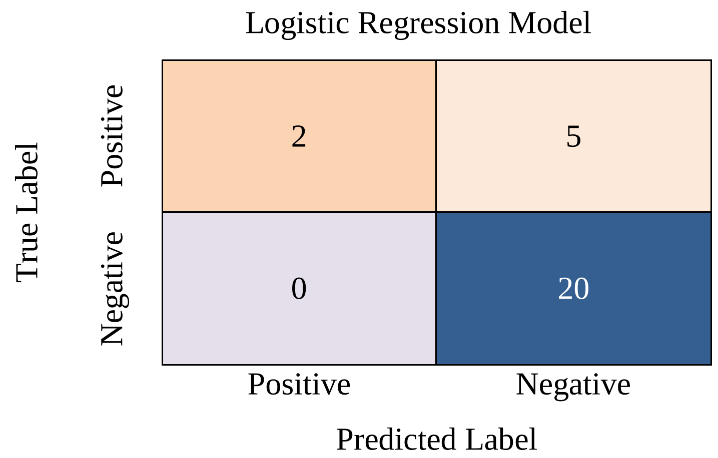

Figure 6.7: Logistic Regression Confusion Matrix

The F1 score of the logistic regression model is high and at an acceptable level. This indicates that the classification performance of the model is good. However, the accuracy score is slightly lower than the LightGBM model. This may indicate that your logistic regression model may in some cases misclassify or be less precise in identifying certain classes.

The study also uses Logistic Regression to examine which factor has a greater impact on the project's approval success than other features. The company's foreign sales are one of the most crucial elements of the project application's success. The characteristics that least affect the success status are respectively the number of partners, the part of the project to be met with equity and the number of the existing personnel. Figure 6.8 shows how effective each feature is shown below.

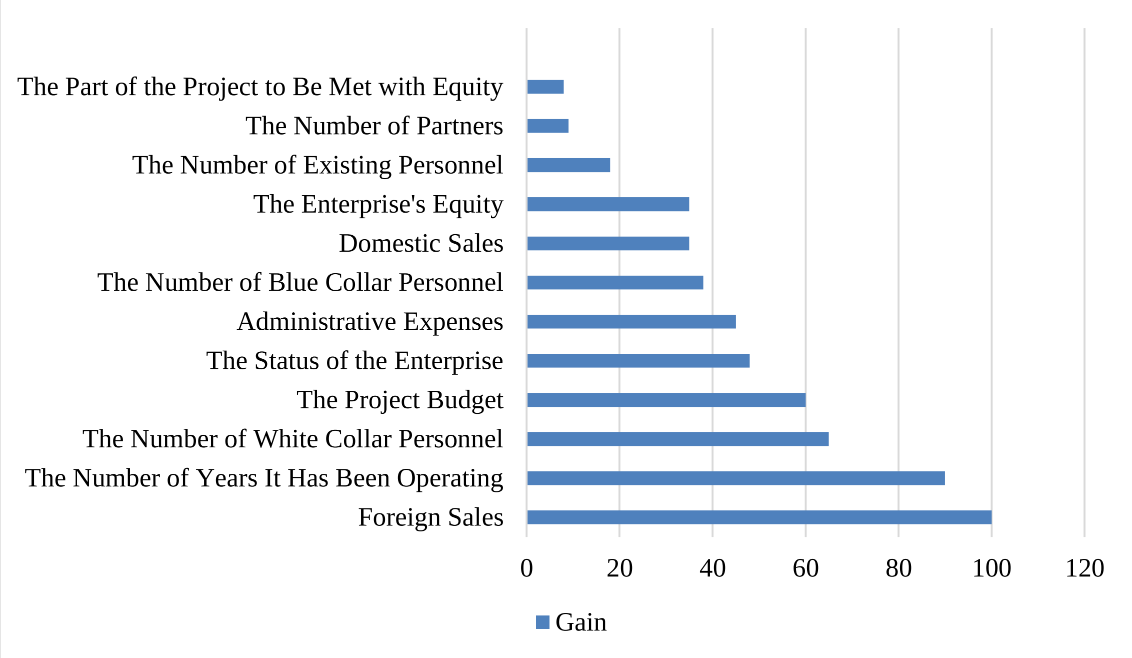

Figure 6.8: Importance of Features in Logistic Regression

6.3.4 Implementation of Support Vector Machines Algorithm

The Support Vector Machines Algorithm model has strong F1 and accuracy scores. A high F1 score shows that the model can deliver accurate results and has good classification ability. Additionally, the accuracy rating is within acceptable bounds. These findings demonstrate the utility of the Support Vector Machines algorithm paradigm. Table 6.5 and Figure 6.7 display the outcome of feeding the data into the Support Vector Machine Algorithm model.

Table 6.5: Support Vector Machine Result

F1 Score	0.88
Accuracy Score	0.81

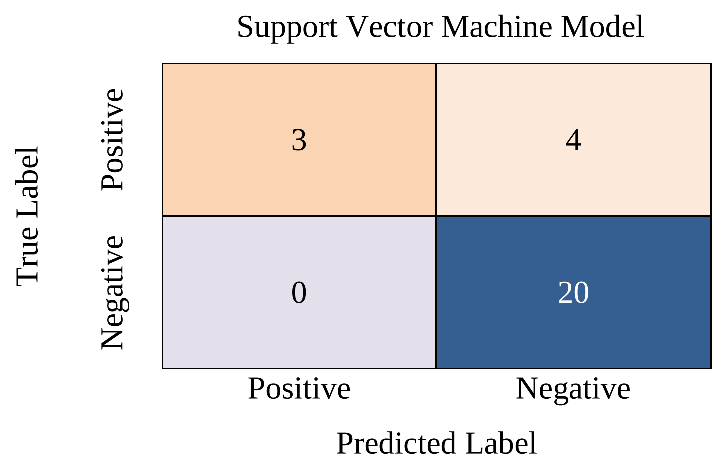

Figure 6.9: Support Vector Machine Model Confusion Matrix

7. RESULT AND DISCUSSION

The F1 score and accuracy score, which are typical measures used to assess the effectiveness of various machine learning algorithms, form the basis of the results.

It is seen that the LightGBM model performs well with an F1 score of 0.93 and an accuracy score of 0.88. The F1 score is a metric that takes into account the balance of accuracy and false positives. A high F1 score indicates that the model is able to reduce false positives as well as true positives. Therefore, we can say that the LightGBM model has the potential to predict high success with a high F1 score.

It is seen that the logistic regression model is slightly lower than the LightGBM model, with an F1 score of 0.88 and an accuracy score of 0.81. However, these scores are still at an acceptable level and demonstrate the ability of the model to make a good prediction of success.

When the F1 score of the support vector machines is evaluated as 0.90 and the accuracy score as 0.85, it is seen that it performs similarly to the LightGBM model. This shows that support vector machines can also be an effective option for future success predictions.

As a result, it is seen that the LightGBM model has the highest F1 score and performs quite well in terms of accuracy score. However, logistic regression and support vector machines are also good options for predicting success. In order to accurately anticipate the future success of the projects, it is crucial to choose the model that will perform the best.

Figure 6.8 depicts the ROC curve, one of the metrics that is most frequently used to assess how well machine learning algorithms work and to explain how well a model predicts.

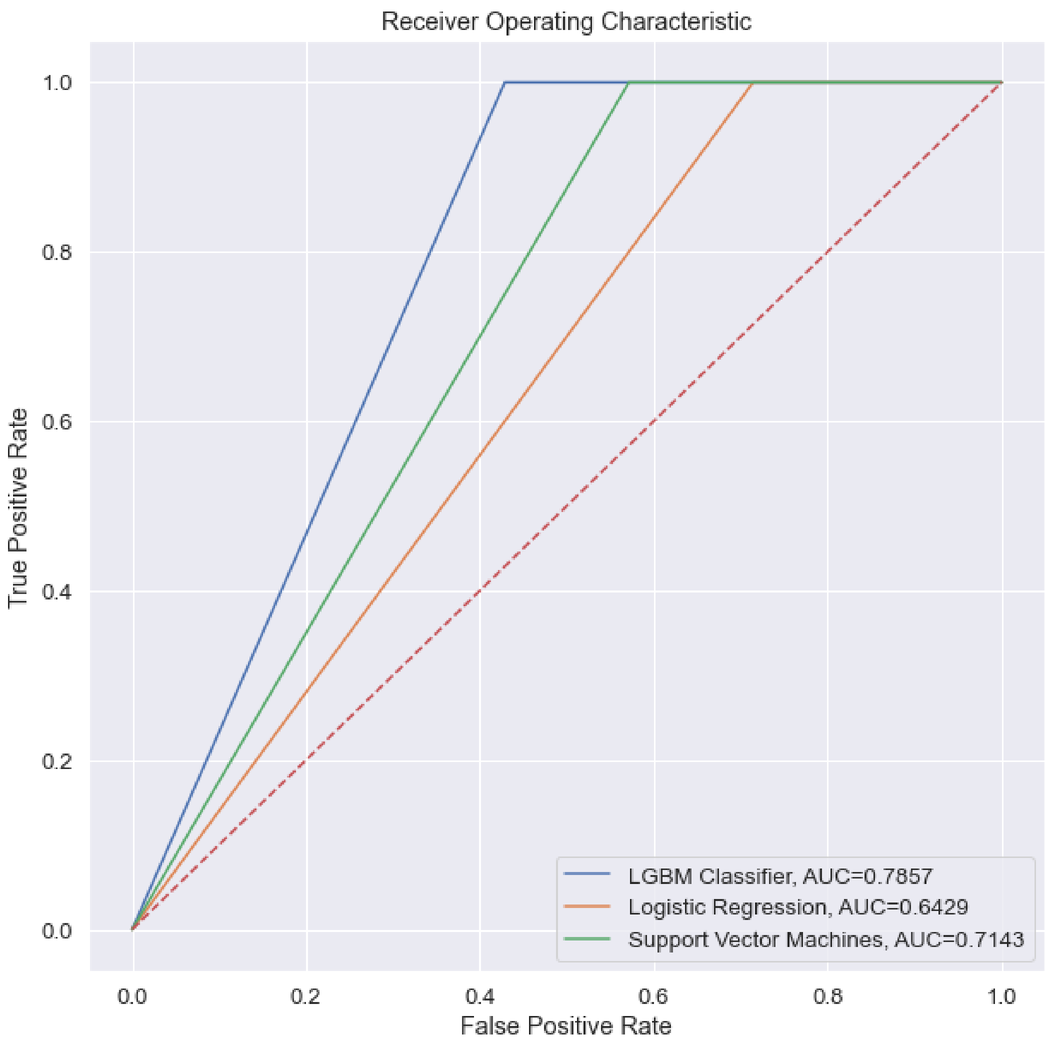

Figure 6.10: The Receiver Operating Characteristic Curve

8. CONCLUSION

The KOBIGEL - SME Development assist Program's objective is to expand the proportion and contribution of SME's to the economy and to assist SME projects that will increase their competitiveness and added value in line with the national and international goals of the country. The types of project costs that will be funded are listed in the Call for Proposals. During the evaluation process for each eligible project, the Project Evaluation Board sets the support level and expenditures that will be covered.

This study was also created to address the issue of resource and time waste brought on by KOSGEB-intended support for SMEs submitting unsuccessful project applications. It aims to prevent the squandering of scarce public resources and guarantee proper distribution of grant funding. In addition, the causes of SMEs' failing initiatives are looked into. These SMEs devote time and money to the project application processes. Data mining classification algorithms have been used to improve the efficiency of government aid by assessing the traits that influence the success or failure of SMEs that have previously applied for project funding.

This study was prepared by using project application data from 81 provinces of Turkey. Applications are received online through the system and stored by KOSGEB.

In every industry today, data size is growing. The examination of the growing amount of data gets more challenging. Big data analysis and the discovery of significant patterns are made possible by data mining.

Companies applying to governmental institutions or consulting firms set up to mediate these supports can utilize data mining, which is used in several areas. Thus, the business that wants to apply for project supports in the future can estimate the approval status of its application. Data can be categorized based on preset outputs using the classification method, one of the data mining techniques. In this study, three classification-relevant algorithms were used to the data set, and it was found that LightGBM produced the best results. The LightGBM model performs admirably in terms of accuracy score and has the highest F1 score. Support vector machines and logistic regression, however, are equally viable choices for success prediction. It is critical to select the model that will perform the best in order to accurately predict the future success of the initiatives.

In this study, the data in the project application form of 177 enterprises that applied to the KOSGEB SME Development Support Program "SME Energy Efficiency" call, which was opened for application in 2017, were examined. The status of the enterprise, the number of partners, the number of years it has been operating, the number of existing personnel, the number of white collar personnel, the number of blue The features of collar personnel, the project budget, the part of the project to be met with equity, the enterprise's equity, domestic sales, foreign sales, administrative expenses are analyzed and evaluated.

Using LightGBM and Logistic Regression algorithms, the information about which of the 12 features in the data set is more important was investigated. Although the results of the two algorithms are different from each other in this regard, in general; The characteristics of the number of the existing personnel and the number of partners turned out to be less important. The number of years it has been operating, domestic sales and foreign sales may be relatively more important features.

Future research can examine the use of data mining to estimate KOSGEB support using information on entrepreneurs from various years. The assessment of the approval status of applications for public subsidies has not been studied previously. This study has the potential to pave the way for future research, and a model for forecasting confirmation states can be developed.

REFERENCES

Adriaans, P. & Zantinge, D. (1997). Data Mining, Boston, MA, USA Addison Wesley Longman Publishing.

Akın, Y. (2008). “Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi”, PhD Thesis, Marmara University, Institute of Social Sciences, İstanbul.

Balaban M. E. & Kartal, E. (2016). Veri Madenciliği Uygulamaları, Çağlayan Kitapevi, İstanbul.

Çimen, F. (2001). “Enerjinin Etkin Kullanımı Verimlilik ve Tasarruf”, Elektrik Mühendisliği Dergisi, 407.

Department of Energy & Climate Change (2012). ‘The Energy Efficiency Strategy: The Energy Efficiency Oppurtunity in the UK’, Energy Efficiency Deployment Office Department of Energy and Climate Change, London.

Doğan, Ş. & Türkoğlu, İ. (2007). " Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees", Doğu Anadolu Bölgesi Araştırmaları Dergisi, Cilt 5, No 2, 163-169.

Energy & Natural Resources Ministry (2022) https://enerji.gov.tr/enerji-isleri-genel-mudurlugu-denge-tablolari.

Energy and Natural Resources Ministry (2016). "Dünya ve Ülkemiz Enerji ve Tabii Kaynaklar Görünümü’, Strateji Geliştirme Başkanlığı, Sayı 11, 01, January 2016, s. 9-10.

Energy and Natural Resources Ministry (2020). UEVEP ve Gelişim Raporları (2017-2020), Ankara: T.C. Enerji ve Tabii Kaynaklar Bakanlığı ulusal Enerji Verimliliği Eylem Planı, 2020, s. 1-4.

Energy Information Administration (EIA) (2016). Total Energy http://www.eia.gov/totalenergy/data/annual/index.php, December 22

Erdabak, M. (2010). Binalarda Isı Yalıtımındaki Eksikliklerin Enerji Tasarrufuna Olan Etkilerinin Uygulamalı Olarak Araştırılması, Master Thesis, Sivas Cumhuriyet University, Institute of Science and Technology, Sivas, Türkiye.

Famili, F., Weber, W.M., Simoudis, R. (1997). "Data Preprocessing and Intelligent Data Analysis", Intelligent Data Analysis, ABD, s. 3-23.

Girija, N., & Srivatsa, S. K. (2006). A research study: Using Data Mining in knowledge base business strategies. Information Technology Journal, 5(3), 590–600. http://doi.org/10.3923/itj.2006.590.600

Han, J. & Kamber, M. (2001). Data Mining Concepts and Techniques; Third Edition, Morgan Kaufman Elsevier Inc, ABD.

Hand, D. J. (1998). The American Statistician, May 1998 Vol. 52, No. 2 117.

Hand, D.J. (1998). "Data Mining: Statistics and More?", The American Statistician, Cilt 52, 112-118.

İnan, O. (2003). Veri Madenciliği, Master Thesis, Selcuk University, Institute of Science and Technology.

Jacobs, P., (1999), "Data Mining: What General Managers Need to Know", Harvard Management Update, Cilt 4, No 10, 8.

Kalıkov, A., (2006), Veri Madenciliği ve Bir E-Ticaret Uygulaması, Master Thesis, Gazi University, Institute of Science and Technology.

Kao, S.-C., Chang, H.-C., & Lin, C.-H. (2003). Decision support for the academic library acquisition budget allocation via circulation database mining. Information Processing & Management, 39(1), 133–147

Karypis, G. & Kumar, V. (1999). Chameleon: Hierarchical Clustering Using Dynamic Modeling, IEEE Computer, 1999: 68-75

Kitler R. ve Wang W. (1998). The Emerging Role of Data Mining", Solid State Technology, Cilt 42, No 11, 45.

Kononenko, I., & Kukar, M. (2007). Machine Learning and Data Mining. Elsevier.

KOSGEB (2021). KOSGEB Annual Report

KOSGEB (2023). KOBİGEL SME Support Program Implementation Principles

Maimon, O. & Rokach, L. (2007). Data Minining And Knowledge Discovery Handbook, Springer Science Business Media Inc., New York.

Nicholson, S. (2003a). Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works. Journal of the American Society for Information Science and Technology, 54(12), 1081–1090. http://doi.org/10.1002/asi.10313

Nicholson, S. (2003b). The bibliomining process: Data Warehousing and Data Mining for library decision-making. Information Technology and Libraries, 22(4), 146–151.

Oğuzlar, A. (2003).“Veri Ön İşleme”, Ege Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Sayı: 21, 2003, s. 67- 76.

Özdabak, A. (2008). “Sanayide Enerjinin Verimli Kullanılması”. VII. Ulusal Temiz Enerji Sempozyumu, UTES'2008.

Piramuthu, S. (1998). “Evaluating Feature Selection Methods For Learning in Data Mining Applications”, Thirty-First Annual Hawai International Conference on System Sciences, IEEE Computer Society, USA, 294.

Ramkumar, D. and Swami, A. (1998). Clustering Data Without Distance Functions, Ramkumar G.D. - Swami A., IEEE Bulletin of the Technical Committee on Data Engineering, Vol.21 No.l, March 1998 : 9-14. 81 İstanbul Ticaret Üniversitesi Dergisi

Roiger, R. &- Geatz, M. (2003). Data Mining: A Tutorial-Based Primer, Boston MA: Addision Wesley.

Savaş, S. & Toğaloğlu N. & - Yılmaz M. (2012). "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri", İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Sayı: 21, s. 1-7.

Saygın, H. (2006). "Küresel Enerji Politikaları ve Türkiye". Türk Harb-İş Dergisi, 219, 26-31.

Schneider Electric (2011). "Enerji Verimliliği Çözüm Kataloğu', EECAT201101TR/01/2011, s. 9, http://www.schneider-electric. com.tr/documents/solutions/EV-cozum-EECAT20110 1TR.pdf

Seidman, C. (2001). Data Mining with Microsoft SQL Server 2000, Microsoft Press, 1 st Ed.; Washington, USA.

Siemens AG. (2014). "Zukunft gemeinsam gestalten Erfolgsfaktor Energieeffizienz Der Schlüssel zu Wettbewerbsfähigkeit und Innovation (Herausgeber)", (Berlin und München: Siemens AG), s. 6-8.

Stepaniuk, J. (2008). Rough-Granular Computing in Knowledge Discovery and Data Mining, Springer Puplishing, Poland.

World Energy Council Turkish National Committee (2010). Energy Report 2010, http://www.dektmk.org.tr/upresimler/Enerji_Raporu_20106.pdf, December 15

World Wide Fund (2011). Enerji Verimliliği ve İklim Değişikliği, http://awsassets.wwftr.panda.org/downloads/wwf_enerjiverimliligi.pdf

BIOGRAPHICAL SKETCH

Hande Demiroğlu was born and raised in Turkey. She obtained her bachelor’s degree in Industrial Engineering from Sakarya University in 2013. In the same year, she started her career in aviation industry in Atlasjet International Airlines. After two years of private sector experience, she started working at KOSGEB. Since 2015, she has been working as SME’s Expert at KOSGEB. She pursued a Master of Science Degree in Logistics and Financial Management Program in Industrial Engineering Department from Galatasaray University in 2020.