REPUBLIC OF TÜRKİYE
ALTINBAŞ UNIVERSITY
Institute of Graduate Studies
Electrical and Computer Engineering

SUSTAINABLE POWER RESOURCE MANAGEMENT BASED ON GIS AND AI ALGORITHMS

Oras Fadhil Khalaf KHALAF

Doctor of Philosophy

Supervisor

Prof. Dr. Osman Nuri UÇAN

İstanbul, 2025

SUSTAINABLE POWER RECOURSE MANAGEMENT BASED ON GIS AND AI ALGORITHMS

Oras Fadhil Khalaf KHALAF

Electrical and Computer Engineering

Doctor of Philosophy

ALTINBAŞ UNIVERSITY

2025

The dissertation titled SUSTAINABLE POWER RESOURCE MANAGEMENT BASED ON GIS AND AI ALGORITHMS prepared by ORAS FADHIL KHALAF KHALAF and submitted on 10/06/2025 has been **accepted unanimously** for the degree of PhD in Electrical and Computer Engineering.

Prof. Dr. Osman Nuri UÇAN

Supervisor

Thesis Defense Committee Members:

Prof. Dr. Osman Nuri UÇAN	Department of Electrical-Electronic Engineering, Altınbaş University	______________
Assoc.Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altınbaş University	______________
Asst. Prof. Dr. Mesut ÇEVIK	Department of Electrical-Electronic Engineering, Altınbaş University	______________
Prof. Dr. Hasan Hüseyin BALIK	Department of Computerl Engineering, İstanbul Aydin University	______________
Prof. Dr. Adil Deniz DURU	Department of Sports and Health Sciences Marmara University	______________

I hereby declare that this dissertation meets all format and submission requirements for a PhD thesis.

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare that all unoriginal materials and conclusions have been cited in the text, and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the aforementioned rules and conduct.

Oras Fadhil Khalaf KHALAF

Signature

DEDICATION

To my cherished father, whose wisdom, fortitude and boundless support have been my unwavering foundation may he forever remain my guiding light .

To the revered memory of my departed mother, whose prayers and affection continue to inspire and illuminate my path .

To my dear brothers and sisters, pillars of strength and companions of joy throughout every phase of my journey .

To my life devoted wife, my partner in life and my greatest source of encouragement, whose unwavering support fuelled my perseverance.devoted wife, my partner and my greatest source of encouragement, whose steadfast support fueled my perseverance.

To my cherished children, the embodiment of hope and the legacy I strive to establish. And to my respected professors, whose counsel, mentorship, and generosity of knowledge have left an enduring mark on my academic path .

With heartfelt gratitude and profound respect, I dedicate this dissertation to all those whose presence, affection and belief in me enabled this achievement after the boundless grace of God.

ABSTRACT

SUSTAINABLE POWER RECOURSE MANAGEMENT BASED ON GIS AND AI ALGORITHMS

KHALAF, Oras Fadhil KHALAF

Ph.D., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri UÇAN

Date: 06/2025

Pages: 62

Türkiye has been experiencing rapid economic development in recent decades, which has significantly increased its energy needs, and the need for sustainable energy sources such as wind energy has emerged. This thesis aims to provide an integrated framework for selecting optimal locations for wind farms in Türkiye by integrating geographic information systems (GIS) techniques with artificial intelligence (AI) algorithms, specifically supervised and unsupervised machine learning. The study involved the spatial analysis of natural data (wind speed, slope, and elevation), socio-economic factors (proximity to roads and urban areas), and environmental factors (protected areas and water bodies). The results of the unsupervised classification algorithms (K-Means and K-Medoids) demonstrated the ability to identify clear clusters of sites with varying degrees of relevance. In contrast, the supervised algorithms demonstrated high classification accuracy, with SVM outperforming with 95.1% accuracy, followed by K-NN and Random Forest, while Naïve Bayes was the least accurate owing to its assumption of feature independence. The thesis adopted Ensemble Learning to enhance the accuracy and diminish the variance between the algorithms' results, culminating in a final identification of the optimal locations where the findings of the four algorithms converge, which were depicted cartographically using GIS. The study affirms that integrating AI techniques with spatial analysis provides a potent tool for decision-makers, expediting the transition towards renewable energy and mitigating environmental impacts. The thesis recommends further development of the models through the use of real-time data and improved multi-criteria decision-making (MCDA) methods and suggests extending the application to other sustainable energy sources.

Keywords: GIS, Wind Farm, Machine Learning, Sustainable Energy, SVM, K-Means, K-Medoids.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

1. INTRODUCTION

Türkiye has witnessed rapid economic development over the past few decades, including the modification from an agricultural economy to a more diversified economy in industry, trade, tourism and other diverse investment fields, allowing the country to maintain a stable economy and become a leading performer in the region and a destination for foreign investment [1]. With this rapid development, Türkiye's growing energy needs have increased significantly, and it needs to provide energy to support this development. The government has taken steps to manage the country's energy supply problems and promote sustainable growth, but it still faces challenges in some areas. This requires planning and forming the necessary policies to ensure energy supply. Highlights the importance of the thesis for studying and analysing the sustainable energy sector in Türkiye, future growth prospects, and finding multiple sources of clean energy that reduce the negative impact on the environment, as well as the challenges and opportunities facing the country while aiming to balance energy needs and the need for sustainable growth [2]. The thesis aims to find suitable sites for wind farms by integrating artificial intelligence algorithms with the outputs of spatial analysis using geographic information systems and using supervised and unsupervised machine learning algorithms to classify sites in the study area (Türkiye) as suitable or unsuitable for wind farms based on several criteria (environmental, economic, and natural) for selecting wind farm sites that have been collected, processed, and analyzed using geographic information systems from various and diverse sources of data. Integrating more than one technology and applying it in different fields provides more accurate and reliable results, and improves the decision-making process compared to other traditional methods. This thesis is a scientific contribution to improving and developing new methods for integrating spatial analysis with machine learning techniques to provide practical solutions in renewable energy and support global efforts in the transition to renewable and sustainable energy sources.

1.1 MOTIVATION

Deploying renewable energy technologies aims to meet the increasing energy demand and reduce the environmental effects of traditional energy sources. Renewable energy, such as wind energy, solar energy and hydroelectric energy, provides a clean and sustainable alternative to fossil fuel, which contributes to reducing greenhouse gas emissions and mitigating the effects of climate change. The motivation for using GIS and AI (Machine Learning algorithms) in wind farm site selection is to improve the accuracy and efficiency of the selection process. AI can analyze vast amounts of data to identify the most suitable sites for wind farms. In contrast, GIS technology enables spatial data representation and helps visualise the results. This integration contributes to better decision-making and ultimately leads to selecting the most suitable sites for wind farm development, increasing energy production potential and reducing environmental impacts. Selecting wind farm sites requires considering several factors and using decision support tools, such as GIS and multi-criteria analysis systems, to ensure that the selected site meets all technical, physical, economic, social and environmental requirements.. Identifying wind energy potential and assessing its spatial distribution is an essential first step in the scientific use of wind energy [3]. Wind electricity generation saw a record increase of 273 TWh (17%) in 2021. This growth was 55% higher than the previous year and was the highest among all renewable energy sources. This rapid growth was driven by a significant increase in wind capacity additions, reaching 113 GW in 2020, compared to just 59 GW in 2019. However, to achieve the net-zero emissions target by 2050, which requires around 7,900 TWh of wind power generation by 2030, there will need to be an annual capacity increase of around 250 GW, more than double the growth achieved in 2020. More significant efforts are needed to achieve this level of sustainable growth, focusing on simplifying the permitting process for onshore wind projects and reducing the costs of offshore wind [4].Renewable energy capacity growth is expected to accelerate over the next five years, accounting for around 95% of global energy growth by 2026 [5]. Renewable power generation needs to expand by over 12% annually over 2022-2030 to reach the track (Figure 1.1) [6].

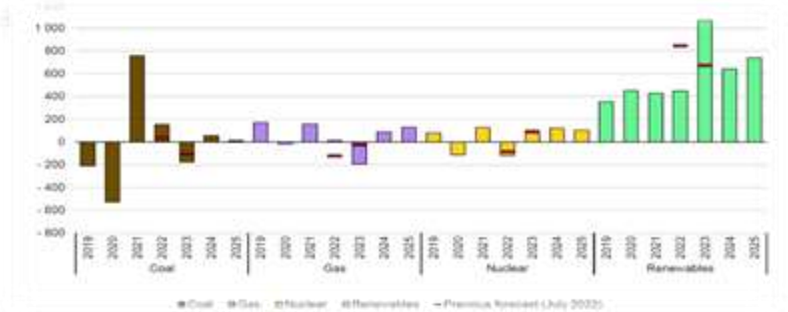


Figure 1.1: Renewables Growth Dampens Fossil Fuel-Fired Generation From 2023 to 2025.

1.2 PROBLEM STATEMENT

The global trend towards sustainable energy, including wind energy, has created many challenges, including 'location', as choosing a location is of great importance in wind energy due to the connection between several factors, mainly wind speed, slope, height, distance from urban areas and transportation routes. With population growth, urban expansion, transportation routes, used land and the presence of pre-established wind farms, the availability of land options has been reduced. That calls for answering the question "where", i.e. the appropriate location for expanding the establishment of wind farms that meet the growing need for renewable energy sources to protect the natural environment and reduce the negative impacts on residents. Integrating artificial intelligence (machine learning) and geographic information systems (GIS) to determine the suitable sites for wind farms, considering all factors, technical for wind energy generation, social and environmental, and the aim of this strategy is to enhance the accuracy and effectiveness of the selection process and thus reduce the time and cost associated with traditional human techniques. The main goal is to advance the sustainable growth of wind energy in Türkiye.

1.3 OBJECTIVES

This study aims to use Geographic Information Systems (GIS) and AI to support the decision-making process in determining the suitable site for wind farms, based on specific criteria. And reach a more accurate and efficient solution.

Objective: Selecting suitable sites for wind farms.
Purpose: Improving the process of selecting wind farm sites.
Tools: ArcGIS Pro analysis and processing tools, supervised machine learning algorithms, and unsupervised.

Expected result: Selecting the optimal sites for wind farms based on specific criteria.
Main challenge: Balancing different site and energy needs criteria while minimising environmental and social impacts.

The benefits of this approach are to reduce cost and time compared to traditional methods, and to speed up the process of obtaining results.

1.4 IMPORTANCE OF THE STUDY

a. This study aims to promote research in the field of sustainable energy.

b. The study will support decision-making related to government policies and investment in renewable energy.

c. Selecting the most suitable sites for wind farms contributes to increasing energy production and reducing environmental impact.

d. The efficient integration of technologies with each other can help solve different problems.

1.5 DISSERTATION OUTLINE.

a. Chapter 1: Introduction, Motivation, Problem Statement, Objectives, and Significance of the study.

b. Chapter 2: Background and related works. Overview of concepts including, GIS, AI (Machine Learning), wind farm.

c. Chapter 3: Describe the data set and the methodology used in the study.

d. Chapter 4: Present the results and provide a discussion of the findings.

e. Chapter 5: Conclude the study and outline future work.

2. SECOND CHAPTER

2.1 HISTORY OF WIND ENERGY

Wind power has been harnessed for thousands of years for different purposes, such as sailing ships and grinding seeds [7]. From here, the significance of wind in developing functional power emerged. At the beginning of the last century, James Blyth and Charles Brush were among the first pioneers of developing electrical power using wind, as they designed the first wind-powered electrical generator. Wind power development began as a significant source of renewable energy. Many countries and companies have supported this field, and more than 100 countries are establishing wind turbines to meet their electricity requirements [8].

2.2 RENEWABLE ENERGY AND AIR POLLUTION

Carbon dioxide emissions from fossil fuel power generation have reached 33.1 GT of carbon dioxide. That has called for alternative actions and plans to have more sustainable energy, reduce carbon dioxide emissions, and swap clean energy sources to reduce the negative environmental impact [8]. In 2015, 186 countries in France joined the climate-saving and sustainable energy agreement to reduce greenhouse gas emissions and dependence on fossil fuels [9]. On the other hand, wind energy is an attractive choice to traditional energy sources and a key component in decreasing air pollution and mitigating the impacts of global warming [10].

2.3 SIGNIFICANCE OF WIND POWER AS A SOURCE OF SUSTAINABLE ENERGY

Wind energy is a clean source of electricity, which is a reason to reduce dependence on fossil fuels and mitigate the negative impact on the climate. Wind energy is relatively inexpensive, especially in areas with substantial wind resources throughout the seasons, which makes it an essential option for generating electricity. The growth of wind energy also creates jobs and stimulates economic growth. Hence, wind energy stands out:

a. Less expensive than other traditional sources, unlimited for generating energy, clean and environmentally friendly electricity [7].

b. Individuals and communities can generate their own electricity, which reduces transmission and distribution losses from generating stations and contributes to building distributed energy and obtaining energy based on their needs [8].
c. During earthquakes, tsunamis and other natural disasters, wind farms are safe and do not cause harm to the environment, such as those that occurred in Fukushima in Japan [4].
d. The flexibility and adaptability of its facilities, such as wind turbines, can be built in a small area and placed along the coast, in mountain ranges, or on plateaus[11].
e. With relatively low initial investment costs and the quick time required to set up wind energy systems, wind energy systems have become economically viable and profitable [12].

2.4 THE GLOBAL VISION FOR WIND ENERGY FUTURE DEVELOPMENT

The world's total installed renewable energy capacity and its share of the electricity grid have increased over the past two decades. Wind power generation increased by 273 TWh (17%) in 2021. This increase was 55% higher than in 2020 and was the highest among all renewable energy sources. This rapid development was driven by the significant growth in wind power additions, which reached 113 GW in 2020, compared to just 59 GW in 2019 [13]. Today, renewables provide around 40% of the world's power. In 2023, the world will add around 510 GW of renewable energy, an increase of almost 50% from the previous year. This growth rate is the fastest in the past two decades [14]. According to the International Energy Agency (IEA), wind farms have become widespread in modern energy networks in most countries that support the transition to renewable energy, and the total global renewable energy capacity is expected to reach 10,800 GW in 2040 [15] Figure 2.1. This requires a careful analysis to choose the suitable sites for wind farms and study the factors affecting them to obtain the highest productivity [16].

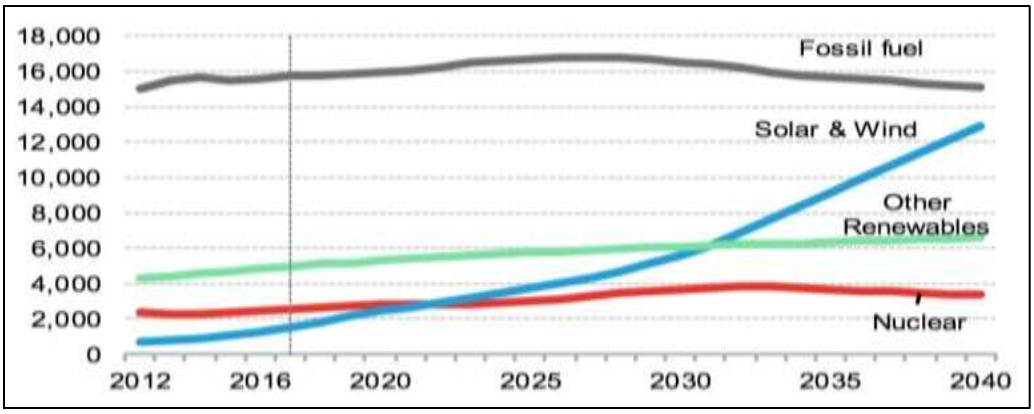


Figure 2.1: Globel Electricity Generation Mix to 2040[17].

2.5 WIND ENERGY DEVELOPMENT IN TÜRKİYE

Türkiye is one of the countries that seeks to invest heavily in the field of sustainable energy, and its wind energy production increased from 1375.80 MW in 2010 to 12482.43 MW in 2023, indicating significant growth, especially after adding 3177.41 MW between 2020 and 2023, Figure 2.2, which made Türkiye take its place in the global renewable energy ranking. The Turkish government has provided various efforts and incentives aimed at increasing investment in this field, including land allocation, tax exemption, and price guarantees [17]. Türkiye's increasing growth in several areas has led to an increasing demand for electrical energy, which is expected to continue in the future [18]. Türkiye's National Energy Plan for 2023 aims for wind energy production to reach at least 29.6 GW.

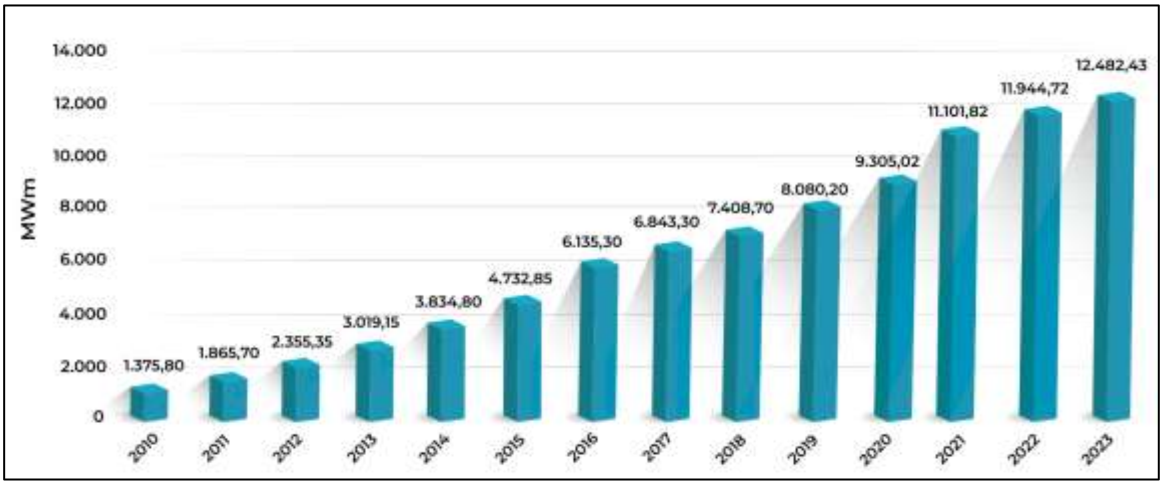


Figure 2.2: Wind Power Growth in Türkiye in MW between (2010 - 2023).

2.6 MACHINE LEARNING

Machine learning is concerned with the design and development of algorithms and techniques that allow computers to learn. The major focus of machine learning is to extract information from data automatically, using computational and statistical methods. It is thus closely related to data [19] Applications of machine learning:

a. Machine understanding and computer ideas, including entity recognition.
b. Natural language processing and syntactic pattern distinction.
c. Search engines, medical investigation, and brain-machine interfaces.
d. Catching credit card fake, supply market analysis, and classifying DNA series.
e. Speech and handwriting recognition, adaptive websites, and robot movement.
f. Computational advertising, computational finance, and health monitoring.

2.6.1 Machine Learning Algorithms

2.6.1.1 Support vector machine algorithm

Support vector machine algorithms analyze data and discover patterns. For example, given a collection of training samples, each marked in one of two classes, the SVM training algorithm creates a model that gives new examples in one class or the other, causing it a non-probabilistic binary linear classifier. SVM builds a representation of the models as a point in space, which is mapped so that the models of the individual categories are divided by a clear gap that is as wide as possible. New models are then mapped to the same space and predicted to belong to a class based on which side of the gap they fall. Support vector machines (SVM) have been developed in machine learning and applied to several applications, ranging from biological data processing to medical diagnosis, time series prediction to face recognition, geographic data classification, and other uses [20]. An advantage of a support vector machine is that the model being built is explicitly based on a subset of data points and support vectors that help interpret the model[21].Advantages (efficient in high-dimensional spaces, concentrated to over fitting) disadvantages (computational complexity, sensitive to selection of seed and hyper parameters).

2.6.1.2 Naïve bayes algorithm

The Naive Bayes algorithm calculates a set of probabilities using frequency counting and sets of values in a given data set. The Bayesian algorithm assumes that all attributes are independent, given the value of the class variable. The Naive Bayes algorithm performs well and learns quickly across different supervised classification problems [22]. Naive Bayes classification is suitable for general classification predictions. Several successful real-life applications use the Naive Bayes classification, such as weather forecasting services, customer credit ratings, weather forecasting, health status ratings, etc. [23]. Advantages (simplicity and ease of performance, quick training and prediction), disadvantages (powerful independence assumption, determined to linear decision borders).

2.6.1.3 K-nearest neighbor (k-nn) algorithm

The K-Nearest Neighbor (K-NN) algorithm is used in various fields, such as fault diagnosis, power system protection, and medical detection. Because of its high efficiency with minimal misclassify compared to other supervised learning classifiers, it can handle major class problems with comprehensive training data [24]. Advantages (simplicity and ease of performance, adaptability to multi-class problems), disadvantages (increased computational cost, intensive memory consumption).

2.6.1.4 Random forest algorithm

The Random Forest algorithm is one of the common algorithms in machine learning. It contains many decision trees to provide accurate results with the lowest error rate [25].The algorithm creates trees to choose the best classification for the test data sets and randomly selects data samples to determine the best results[26]. The non-linear method is the method used by the algorithm to discover relationships between features. This method makes it a powerful tool for classification and regression modelling. It does not cut trees, unlike other tree-based algorithms. It partitions random subsets of data at each tree node, increasing the tree set's diversity and improving performance [27]. Advantages (high accuracy, handles over fitting well) disadvantages (computational complexity, memory usage, poor interpretability). Random Forest is an effective tool for large-scale and multivariate pattern recognition [28].

2.6.1.5 K-Means algorithm

The Spatial K-Means algorithm assumes a point-assignment technique to create clusters by evaluating multivariate means within Euclidean space [29]. To effectively use the K-Means algorithm in our spatial dataset, the K-Means algorithm can be mathematically described using an objective function that depends on the distances between data points and the cluster centroids, as follows:

$$\underset{\{m_k\},1\leq k\leq K}{min} \sum_{k=1}^{K} 1 \sum_{x\in C_k} \pi_x \,\mathrm{dist}(x, m_k) \qquad (2.1)$$

"where πx is the weight of x, m_k is the number of data objects assigned to clusterC_k, $m_k = \sum_{x\in C_k} \frac{\pi_x x}{n_k}$ is the centroid of cluster Ck, K is the number of clusters set by the user, and the function "dist" computes the distance between object x and centroid m_k, 1 ≤ k ≤ K. While the selection of the distance function is optional, the squared Euclidean distance, i.e. ||x − m ||2, has been most widely used in both research and practice." The iteration process introduced in the previous paragraph is a gradient-descent alternating optimization method that helps solve Eq. (1.1), although it often converges to a local minimum or a saddle point [30]. The K-means algorithm is a widely used and well-known algorithm previously applied to solve different computational problems, such as geo-statistics in GIS [26] ,or image classification and segmentation [31] .The k-means algorithm is a robust technique used in geographic analysis for clustering and grouping spatial data and non-spatial data into meaningful clusters depending on specific criteria in multiple fields. Below are some attributes that illustrate how the K-Means algorithm erforms in geographical analysis:

a. Data Preparation: Geographic data, such as latitude and longitude coordinates, are collected for the locations of interest. These data points represent geographic features in this study: geographic data features (latitude and longitude coordinates, wind speed, slope, and elevation).

b. Initial Cluster Centers: The algorithm starts by randomly selecting a predetermined number of cluster centres known as centroids. These centroids represent the initial guesses for the cluster locations. The number of clusters is typically determined beforehand or through techniques like the elbow method.

c. Assignment of Data Points to Clusters: Each geographic data point (latitude and longitude coordinates, wind speed, slope, and elevation) is assigned to the nearest cluster centroid based on the Euclidean distance or other distance metrics that consider the geographic coordinates. This step creates initial clusters.

d. Update cluster centroids: After assigning all data points to clusters, the algorithm calculates new centroids for each cluster. These new centroids are computed as the mean (average) of all the data points assigned to each cluster. This step repositions the cluster centres based on the current member data points.

e. Iteration: Steps 3 and 4 are repeated iteratively until one of the stopping criteria is met. The most common stopping criteria include a maximum number of iterations or until the centroids no longer change significantly between iterations.

f. Final clusters: The final clusters are formed once the algorithm converges (i.e., the centroids stop changing significantly). Each data point belongs to the cluster represented by the nearest centroid.

g. Analysis and interpretation: We can now analyse the results and examine the factors of each cluster (latitude and longitude coordinates, wind speed, slope, and elevation) connected to the sites within the clusters. This analysis helps understand spatial patterns and trends within the data.

h. Decision-Making: The geographic clusters can inform decision-making processes, such as site selection for wind turbines, identifying hotspots for specific phenomena (e.g., disease outbreaks, traffic congestion), or optimising routes for delivery services.

K-means is widely used in geographic analysis because it can efficiently handle large datasets and reveal hidden patterns or structures within geographic data. It is important to note that the initial selection of cluster centroids and the choice of distance metric can significantly impact the results, and analysts often perform sensitivity analysis to ensure robustness. Additionally, other variations of K-means, such as weighted K-means or K-medoids, can be applied depending on the specific needs of the geographic analysis. The K-Means algorithm and the K-Medoids algorithm are both clustering algorithms used for grouping data points, but they have significant differences in how they operate and determine the cluster centres (medians/medoids); the differences between them:

a. Cluster Center Concept:

i. K-Means uses the average of all data points to be the centre of the cluster. A new centre is created by calculating the geometric mean of all the points in the cluster.

ii. K-Medoids uses any data point as the centre of the cluster. It is called a "medoid" and is one of the actual data points in the cluster.

b. Center Calculation:

i. K-Means: Gets the new centre by calculating the mean of coordinates (the centroid) of all data points in the cluster.

ii. K-Medoids selects a point from the actual data points to be the cluster's centre.

c. Distance Calculation:

i. K-Means: Computes the distance between the centre and other points using the Euclidean distance metric.

ii. K-Medoids uses the Euclidean distance metric but calculates the distance between the medoid and other points in the cluster.

d. Sensitivity to Outliers:

i. K-Means: Its dependence on the mean makes it sensitive to outliers.

ii. K-Medoids: more potent with the presence of outliers because it uses an actual data point as the centre.

e. Computational Efficiency

i. K-Means: It can be faster in many cases because it calculates the mean of data points.

ii. K-Medoids takes more time because the distance between every one of the points and all points in the cluster is calculated.

In our study, we chose (K-Means algorithms) because it is compatible with the nature of the data, size, and methodology used.

2.7 AI IN WIND ENERGY

Integrating AI into renewable power systems means a transformative power management and optimisation method. The principles of AI in this context revolve around improving efficiency, predicting power designs, and optimising resource allocation in wind power systems. Some applications of AI in renewable energy:

a. Classifications of artificial intelligence approaches supply a visual representation of those methods that can be used for full-fledged forecasting of wind and solar power, as a development of which a procedure was designed for using artificial intelligence technologies for full-fledged forecasting at all phases: from data preprocessing to predicting the amount of power generation for wind and solar power[32].
b. Utilising AI to enhance sustainability and power efficiency facilitates a more resilient and efficient power landscape. AI is revolutionising the renewable power sector, changing infrastructure maintenance, optimising energy generation, and combining renewables into the grid. Its advanced analytics, predictive abilities, and optimisation are required to reach global renewable power objectives [33].
c. Artificial intelligence technology has seen growth in the control and design of offshore wind power systems. The application of artificial intelligence technologies such as fuzzy logic, inferential algorithms, deep learning, and reinforcement learning has been thoroughly analysed from the viewpoint of optimisation implementation [34].
d. Employing deep learning models for time series forecasting to manage the problem of accurate wind power analysis. Which is difficult due to its unexpected and unstable nature, is a powerful tool for accurate prediction of wind speed and power, which is essential for wind power planning systems. It lays the foundation for a greener and more sustainable power future [35].
e. Machine learning (ML) algorithms are powerful tools for wind energy forecasting. Among other forecasting approaches, they can manage wind energy uncertainty, which is the main obstacle to its integration into the power grid, through accurate and efficient forecasting [36].

f. Employing machine learning can enhance decision-making abilities. Based on the geographical coordinates of wind power facilities, decision-makers can perform faster and more efficiently and select suitable sites for wind farms according to specific criteria [37]

2.8 GIS AND MACHINE LEARNING INTEGRATION

Geographic information systems (GIS) can effectively analyse and visualise spatial and non-spatial data in selecting suitable wind farm sites [38]. GIS allows the integration and processing of geospatial datasets, including topography, land use, wind speed, environmental constraints, and socio-economic characteristics, to identify suitable sites for wind farm development. It collects and analyses point and vector data and creates a buffer zone for all land uses according to the criteria adopted in the analysis. GIS allows the overlay and analysis of multiple spatial data layers, facilitating the complete assessment of potential wind farm sites. Wind resource maps, derived from recorded wind speed measurements or numerical modelling, can generally be integrated into GIS platforms to identify areas with favourable wind conditions [39]. Environmental considerations play a critical role in wind farm site selection. GIS allows the integration of different environmental datasets, including protected areas, wildlife habitats, and sensitive ecosystems, to promote sustainable wind farm development. The possibility of integrating GIS with machine learning is a branch of artificial intelligence (AI) that deals with creating programmes and strategies that can learn and measure from data, extract patterns, and make decisions based on that data. "The ability of a system to accurately analyse external data, learn from that data and utilise that learning to perform specific goals and missions through flexible adaptation", Machine learning algorithms can build a mathematical model based on a set of "training data" to produce results in the form of predictions or classifications without being explicitly programmed to do so [40]. Machine learning (ML) is a relatively new concept compared to geographic information systems (GIS). Machine learning involves a computer programmer learning from different experiences to enhance performance on specific tasks. The measurable performance of the programmer improves with increasing experience and insights from performing these tasks. The machine uses evaluations and predictions based on the data it receives [41]. Machine learning approaches are increasingly applied in different domains, including GIS [42].

2.9 LITERATURE REVIEW

[43] The study used standardisation to develop a general performance index to match the evaluation. The interrelationship between multi-criteria decision-making strategies was analyzed using Pearson's correlation coefficient in the study region of Ecuador. [44] A type 2 fuzzy analytical hierarchy approach determines suitable wind farm sites. The central idea of the model is to solve the problems related to vagueness, uncertainty, and inconsistency in wind farm site selection using fuzzy sets to describe expert linguistic judgement. Nigerian study area. [45] A Group Spatial Decision Support System (GSDSS) is presented that can be employed to build techniques in infrastructure project planning, in addition to information and tools from collective intelligence, complexity theory and geographic forecasting. This approach can be used in Mexico's Strategic Environmental Assessment (SEA). [46] A multi-criteria decision-making model established on Geographic Information Systems (GIS) is designed to select suitable wind farm sites based on biological, environmental, social and economic factors. Each selected criterion has a similar weight in the model, with Saudi Arabia as a case study. [47] A framework for the spatial selection of wind farm sites that integrated Geographic Information Systems (GIS) and Analytical Network Processing (ANP) in a neutrosophic environment was presented. After constructing a site selection evaluation model, the expert committee used a bipolar neutrosophic set in the decision-making process.[48] The Boolean method was employed with a scientific GIS software package to find suitable sites in Cameroon to construct wind farms.[38] A new approach for determining a suitable site for wind farms was suggested, which integrated the complete factorial design and the analytical hierarchy process. As a [49]. The suitability of the wind farm site was chosen using GIS and the analytical hierarchy process (AHP). The AHP was assigned appropriate weights to the guiding criteria according to their importance. GIS was used to perform geospatial analysis and generate the suitability map to determine suitable areas. [50] GIS, AHP (Analytical Hierarchy Process) and TOPSIS (Technique of Similarity to Ideal Solution) were used to evaluate the suitability and sustainability of wind farm sites in Greece. [51] Combining the fuzzy analytical hierarchy process (AHP) and the demand preference technique based on similarity to the ideal solution (F-TOPSIS) created a hybrid decision support system to evaluate potential sites for wind power plants. The evaluation takes into account factors such as wind speed and direction, land use, and distance to power lines. [52]Machine

learning techniques were used to optimise wind power plant site selection in the Philippines, and the analysis was done using Support Vector Machines (SVM). [53] Key tools include machine learning algorithms, datasets, and methods for criteria weighting and analysis. The study highlights integrating the PROMETHEE II method with GIS for suitability analysis, followed by machine learning regression algorithms to generalise suitability indices across the entire area. A spatial hybrid decision-making framework for wind farm site selection in northeastern Greece was introduced. [54]Split the factors into two separate groups: economic and environmental. A Spatial Decision Support System (SDSS) is also used to show on a map an environmental layer solution, an economic layer solution, and a mixed suitability layer. [55]This study utilised Rational Basis Function Neural Networks (RBFNN) to transparently assess criteria for wind turbine (WT) installation locations, integrating results with protected areas and the Land Fragmentation Index (LFI) for identifying socially acceptable and energy-efficient sites. [49]GIS and the Analytical Hierarchy Process (AHP) were used to do a wind farm site fit analysis. The AHP was used to give each criterion the right amount of weight based on how important it was, and the GIS was used to do a geospatial analysis, add many criteria to the proposed index, and make the end suitability map that showed where the criteria weren't met. [56] Fuzzy Scaled Weighted Ratio Analysis (F-SWARA) was used to classify the criteria and determine the priorities in the wind farm installation. In contrast, Fuzzy Measurement Alternatives and Classification by Leveling Solution (F-MARCOS) were employed to identify the most suitable location for the wind farm. The study area is Sivas province, located in central Türkiye. [57] AHP determined the relative weight of each criterion of sites, and GIS were used to incorporate environmental emissions data from the LCA into the decision-making process. An analytical model was used to estimate the transportation for each site. [58] The suitable location was selected by expectation theory and VIKOR multi-criteria ranking method based on determining the criteria weights by combining Decision-Making and Evaluation Experimental Laboratory (DEMATEL) and Entropy Weight (EW) method; the study area is China.

3. THIRD CHAPTER

3.1 STUDY AREA

Türkiye is found between 36° and 42° north and 26° and 45° east; the climatic and topographic features of Türkiye are characterised by substantial differences that indicate a characteristic feature of Türkiye, which signifies its unique nature [59]. Four seas surround Türkiye: the Marmara, Aegean, Mediterranean, and the Black Sea. On one side are Greece and Bulgaria, and on the other are Armenia and Georgia. To the southeast are Iraq and Iran, and to the south is Syria. Türkiye consists of 81 provinces. The country's seven regions are Marmara, Aegean, Mediterranean, Southeastern Anatolia, Eastern Anatolia, Central Anatolia, and the Black Sea. Its total area is 783,562 km2. Figure 3.1. There are several reasons for selecting Türkiye as the area for this thesis: -

a. Türkiye is one of the leading countries in this domain (sustainable power)[13].
b. Türkiye's topographic diversity of the Taurus Mountains, southern and northern Anatolia, Anatolian plains, and seas and its location between two continents has made it a source of wind activity throughout the seasons, making it a perfect area for wind power.
c. With Türkiye's future objectives of improving electricity production via wind and more than 239 wind farms [13] .Discovering suitable sites for new projects becomes more complex.

Figure 3.1: Map of Türkiye [60].

3.2 DATASET

3.2.1 Collection and Preprocessing of Data

Data on climatic, hydrological, topographic, and geological factors were collected based on extensive publications on wind farm site section, and various criteria reflecting local requirements were determined. These. standards are divided into three classifications: natural, socio-economic, and environmental. The collected data (raster and vector) was processed and analyzed using ArcGIS Pro software. The first section of the 'dataset' represents the input for the algorithms. In the second section, the 'dataset' for algorithm training represents the data of existing wind farm turbines. Site data (wind speed, elevation and slope) represents the measurements of this data, which are suitable for installing new farms in various regions in Türkiye.

3.2.2 Natural Factors

Critical factors affecting wind turbine placement include wind speed, slope, and elevation. Wind speed is essential for electricity generation. Wind speed is essential for optimal electricity generation from wind turbines. Higher wind speeds are essential for adequate wind energy production [61]. The World Wind Atlas collected monthly wind speed data at an altitude of 100 m above ground. The months with higher-than-average wind speeds are identified by evaluating the distribution curves in Figure 3.2, Figure 3.3 shows the relationship between wind speed distribution with slope and elevation.

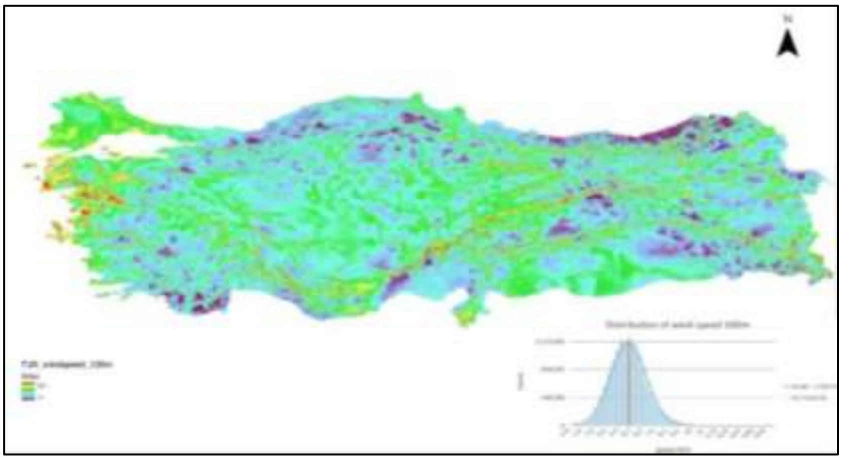


Figure 3.2: Map of Wind Speed Distribution.

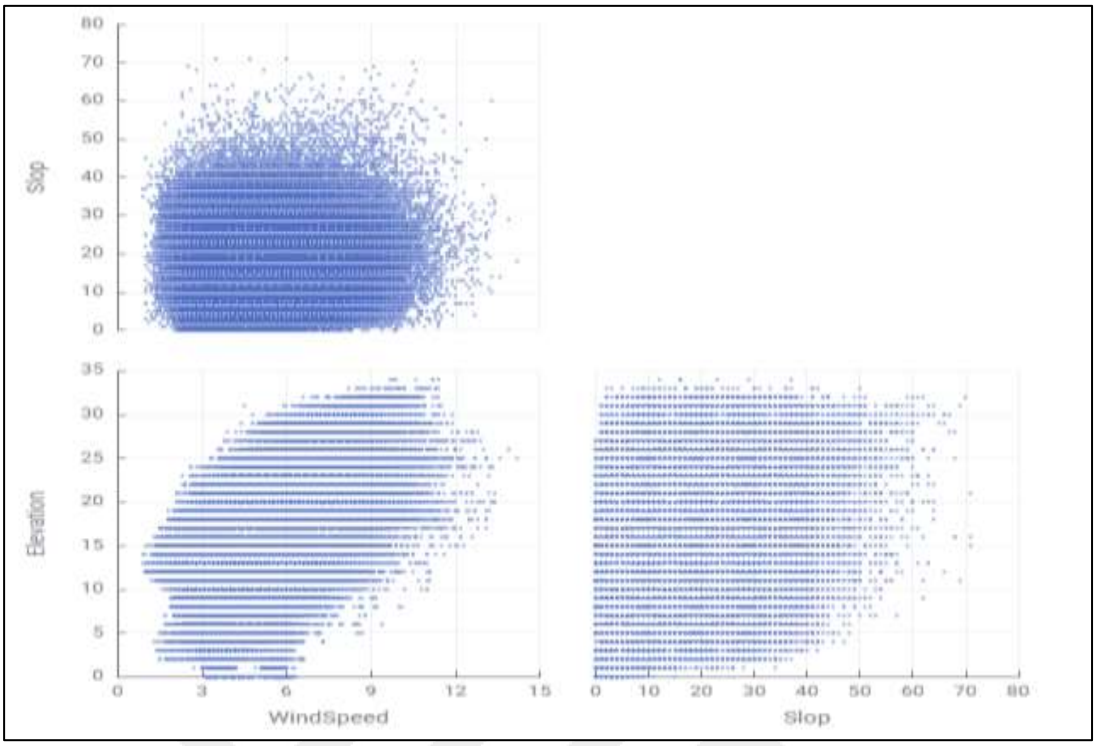


Figure 3.3: Distribution of Data (Wind Speed, Slope, and Elevation).

Figure 7 illustrates three points in which wind speed is related to the amount of energy.

a. Cut-in speed: The turbine begins to turn at the speed of 3-4 M/S.
b. Rated output power and rated output wind speed: The wind speed is between 12-17 M/S; thus, the level of power produced reaches the limitation that can be produced according to the structure of the turbine.
c. Cut-out speed: With an increase in speed beyond 25 M/S, the forces acting on the turbine structure continue to rise, and there is a danger of damage to the rotor. As a result, a braking system is operated to stop the rotor.

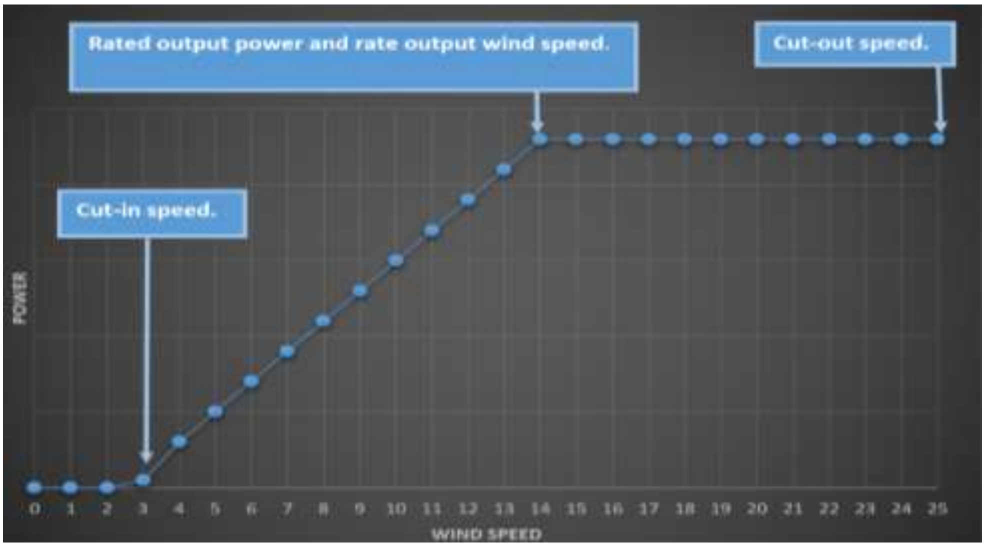


Figure 3.4: Wind Turbine Power Curve Description.

3.2.3 Socio-economic Factors

Wind farms should be located safely in urban locations and residential areas to avoid noise and provide safety. It is advisable to locate wind farms near existing transport networks, as accessibility is very significant due to their proximity to roads and railways. To lower the costs associated with installation and maintenance of a wind farm, the distance factor between wind farms and urban locations or residential areas is not fixed, due to the design and shape of the wind turbines. Some consider wind farms as visual pollution, while others view them positively due to their unique design and the slow rotation of their blades that adapt to the wind movement. These criteria are determined locally according to the laws of the country. However, the safety factor remains important [62].

3.2.4 Environmental Factors

Several environmental factors should be considered when selecting wind farm sites in areas of environmental concern and with different land cover types. Factors to be considered include distance from areas of environmental concern, land cover, and land surface use. When determining the optimal sites for installing turbines, it is important to consider the impact of wind turbines on protected areas and bird migration routes [63]. Land use/land cover is a critical factor in determining energy investments. Wind farms should be strategically located in areas that have a minimal impact on land use, marshlands, aquatic environments, and forested areas. Wind farms typically occupy an area of 4 to 6 km^2 [48].

Therefore, it is essential to mitigate negative impacts on vegetation during the construction of wind farms. The possibility of utilising the land surrounding the wind turbines for growing agricultural crops without causing any interference with their functions. Table 1 shows the site selection criteria for wind farms, taking into account natural, social, economic and environmental criteria to exclude unsuitable sites for installing wind farms by setting specific criteria for each factor.

Table 3.1: The List Values for Criteria and Buffer Zone.

Criteria	Exclusionary Criteria	Buffer Zones	Sources
Natural	Elevation	<2500m	United States Geological Survey Earth Explorer
	Slope	<30%	
	Wind Speed	>3.5m	Global Wind Atlas
Socio-Economic	Roads	0-500m	MapCruzin.com
	Urban Areas	0-2500m	
	Airports	>3500m	
Environmental	Land Cover/ Land Use	>500m	European Environment Agency.
	Water Bodies	>500m	
	Protected Areas	0-2000m	Protected Planet protectedplanet.net › country ›TUR

First, primary data on selecting suitable sites for wind farms were collected, including natural, economic, and socio-environmental data. Then, this data was processed and analysed, a database was developed, and the data was represented as vector and raster data on the study area map Figure 3.5.

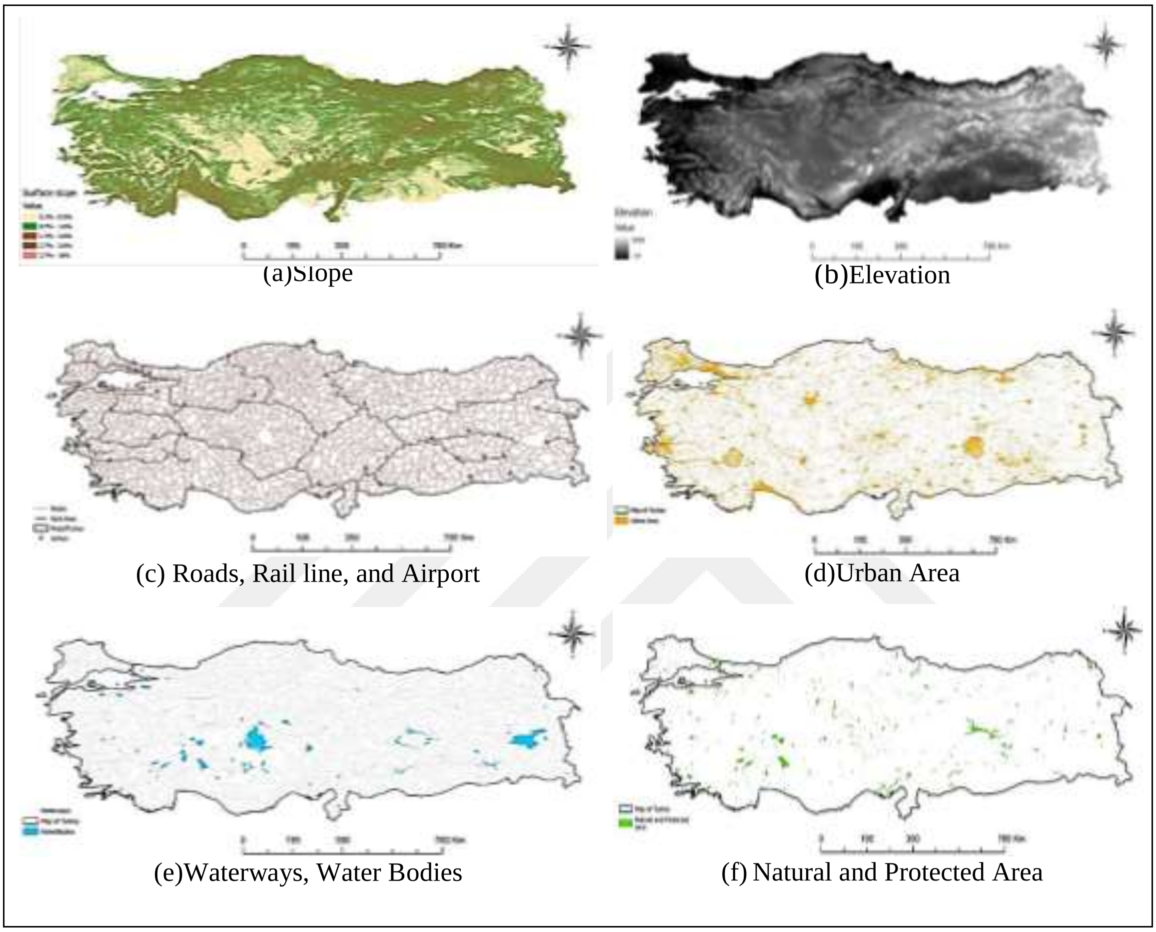
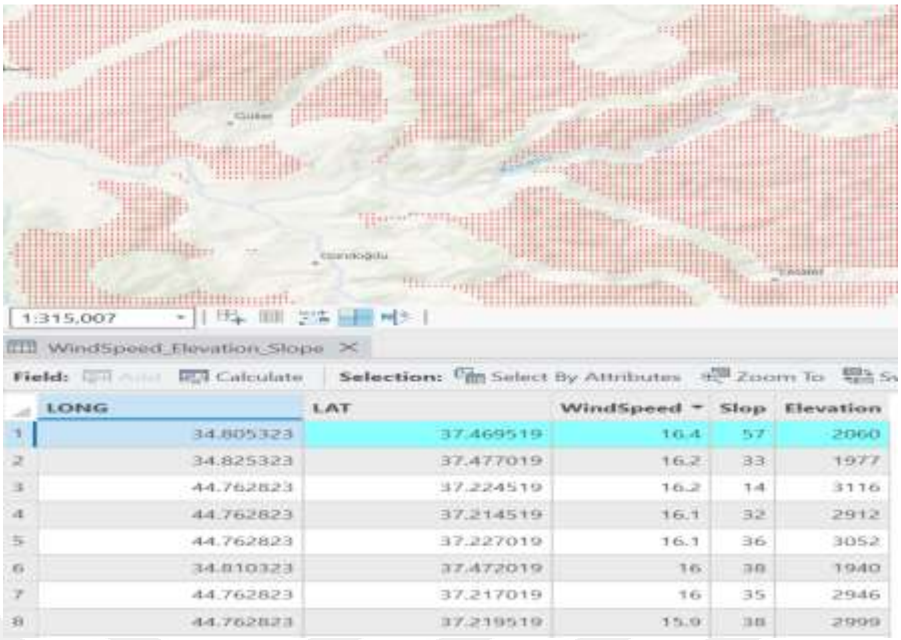


Figure 3.5: Shows Data Maps.



	LONG	LAT	WindSpeed	Slop	Elevation
1	34.805323	37.469519	16.4	57	2060
2	34.825323	37.477019	16.2	33	1977
3	44.762823	37.224519	16.2	14	3116
4	44.762823	37.214519	16.1	32	2912
5	44.762823	37.227019	16.1	36	3052
6	34.810323	37.472019	16	38	1940
7	44.762823	37.217019	16	35	2946
8	44.762823	37.219519	15.9	38	2999

Figure 3.6: Illustration Sites Related to Data (Longitude, Latitude, Wind Speed, Slope, and Elevation).

This data was utilized as input for the Multivariate Clustering algorithm to analyse the data and determine similar groups based on several criteria (wind speed, slope, and elevation). Figure 3.6 shows the methodology framework stages.

3.3 METHODOLOGY

This thesis used two methodologies to classify suitable and unsuitable sites for wind farms. The first used an unsupervised machine learning algorithm (K-Means and K-Medoids), and the second used the supervised machine learning algorithms (K-Nearest Neighbor, Random Forest, Support Vector Machines, Naive Bayes), which used the same study area and the same inputs and criteria.

3.3.1 Unsupervised Machine Learning Algorithm (K-Means and K-Medoids) Methodology

Figure 3.7 explains the methodology used in this study to select suitable wind farm sites in Türkiye. It includes Phase I, which identifies the problem and criteria that define the specifications of the suitable sites and collects data in multiple forms and sources. Phase II analyzes and processes data using GIS software tools that include maps (urban areas, transportation routes, reserves, lakes and water bodies, airports) and thus excludes them from being suitable site for wind farms. After the exclusion of all land use, a map of the nominated areas for establishing wind farms was created and converted from raster data to vector data. Each point represents coordinates on the ground and has characteristics (wind speed, elevation, and slope). Create maps of existing wind farm sites in Türkiye. Each site has the appropriate characteristics (wind speed, elevation, and slope) and is converted into a formula to train data for learning algorithms. Phase III uses four machine learning algorithms. The same dataset and training data were entered for each algorithm to classify suitable and unsuitable points, and the results and accuracy were analyzed. Phase IV uses GIS to represent the algorithm outputs, find the final result that represents suitable and unsuitable sites, and represent it on the map.

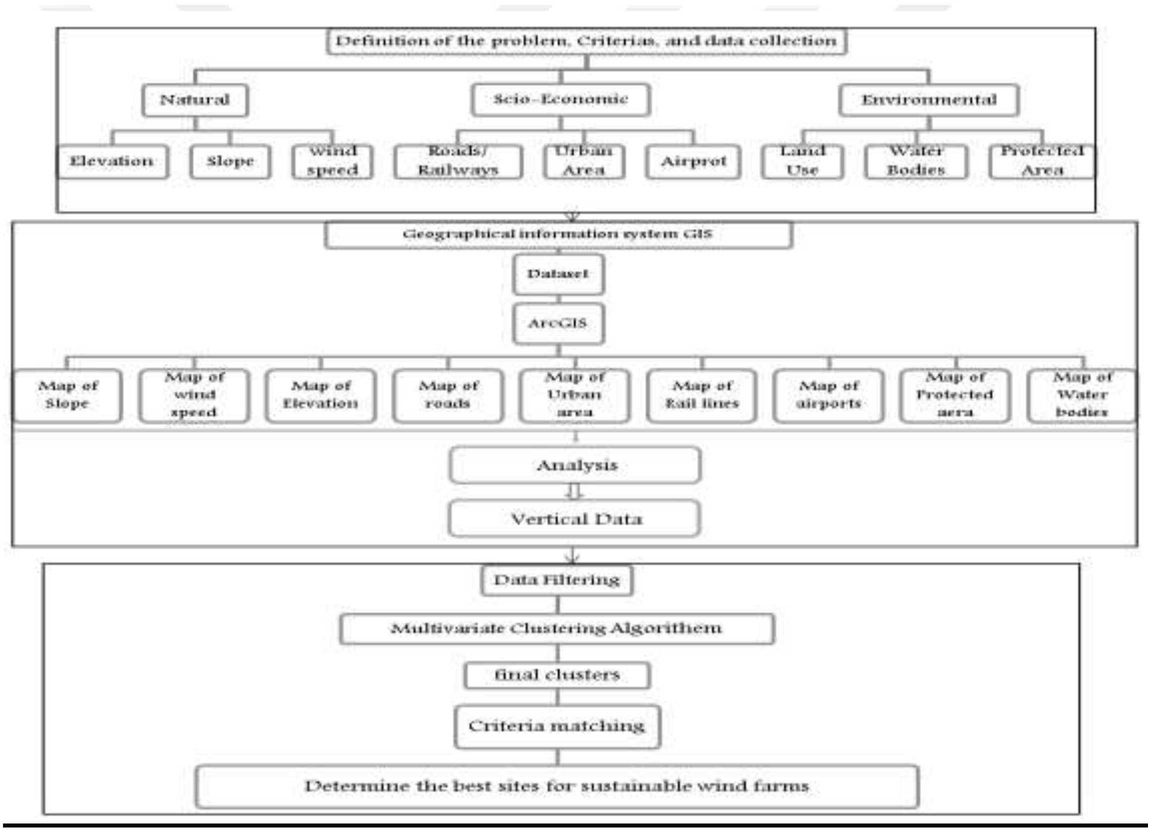


Figure 3.7: Framework Stages.

K-Means and K-Medoids are unsupervised clustering algorithms. The data is clustered as dependent on its attributes, internal form and effects [64]. The K-means and K-Medoids algorithm is widely used to solve various computational problems, such as geo-statistics in GIS [29],or image categorization and segmentation[65].

Why K-means/K-medoids?

a. Simple output structure: The two algorithms both assign each data point to a single cluster, which generates clear and discrete suitability categories (e.g., “very suitable” to “not suitable”).
b. Transparent criteria: Cluster centres/medians directly represent typical site features (e.g., high wind speed + low elevation = “very suitable”). Unlike “black box” models (e.g., neural networks), these features can be validated against domain knowledge.
c. Compatibility with GIS visualisation: Clustering results are easily plotted on ArcGIS maps, allowing for intuitive spatial interpretation
d. Trade-offs: While hierarchical clustering offers the potential for dendrogram-based interpretation, it becomes impractical for large datasets due to complexity.

The K-means and K-Medoids algorithm is a practical approach employed in geo-analysis for clustering spatial and non-spatial data into significant clusters depending on specific criteria in multiple domains. It is essential to note that the initial choice of cluster centroids and distance metrics can impact the results. Also, other variations of K-means, such as weighted K-means or K-medoids, can be used depending on the detailed requirements of the geographic analysis, but they have essential differences in how they work and choose the cluster centres the differences between them, Table 3.2.

Table 3.2: The Differences between K-means and K-Medoids Algorithm.

Center Calculation	K-Means	Gets the new centre by calculating the mean of the centroid of all data points in the cluster.
	K-Medoids	Selects a point from existing data points to be the new cluster's centre.
Distance Calculation	K-Means	Calculates the distance between the centre and other points using the Euclidean distance metric.
	K-Medoids	Utilises the Euclidean distance metric but estimates the distance between the medoid and other points in the cluster.
Sensitivity to Outliers	K-Means	Its dependence on the mean causes it to be sensitive to outliers.
	K-Medoids	It is more powerful with the existence of outliers because it utilises an actual data point as the centre.
Computational Efficiency	K-Means	It is faster in multiple cases because it calculates the mean of data points.
	K-Medoids	The computations take longer because the space between every point and other points in the cluster is calculated.

3.3.3 Geographic Information Systems (GIS)

Geographic Information Systems (GIS) can effectively analyse and visualise spatial and non-spatial data when choosing suitable sites for wind farms. GIS allows the integration and processing of geospatial datasets, including topography, land use, wind speed, environmental conditions, and socio-economic factors, to determine suitable sites for wind farm development. Figure 3.8 illustrates the model utilised in our study, which collects and processes. It analyses point and vector data and makes a buffer zone for all land uses according to the criteria adopted in the study. GIS allows for the overlay and analysis of multiple spatial data layers, facilitating the comprehensive assessment of possible wind farm sites. Wind resource maps are derived from historical wind speed measurements. Environmental considerations play a critical role in wind farm site selection. GIS enables the integration of various environmental datasets, including protected areas, wildlife habitats, and sensitive ecosystems, to facilitate sustainable wind farm development.

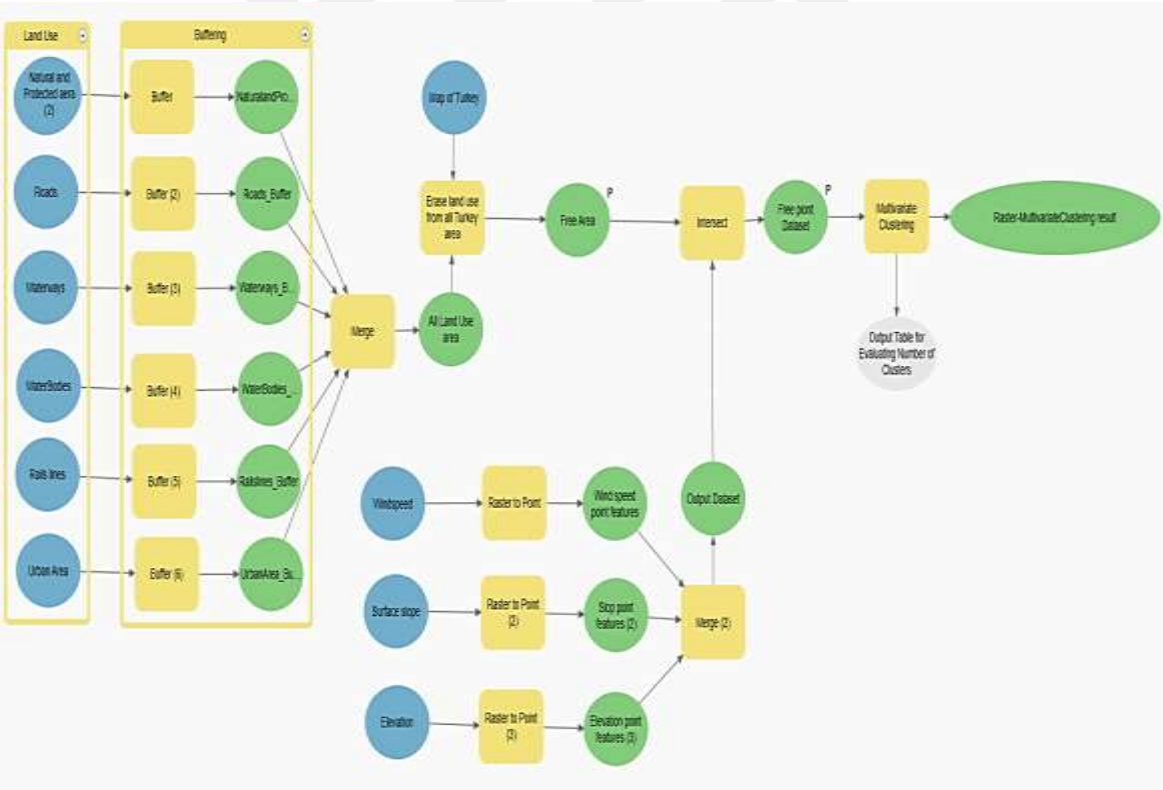


Figure 3.8: GIS Model.

3.3.3 Unsupervised Machine Learning Algorithm (K-Means) Methodology

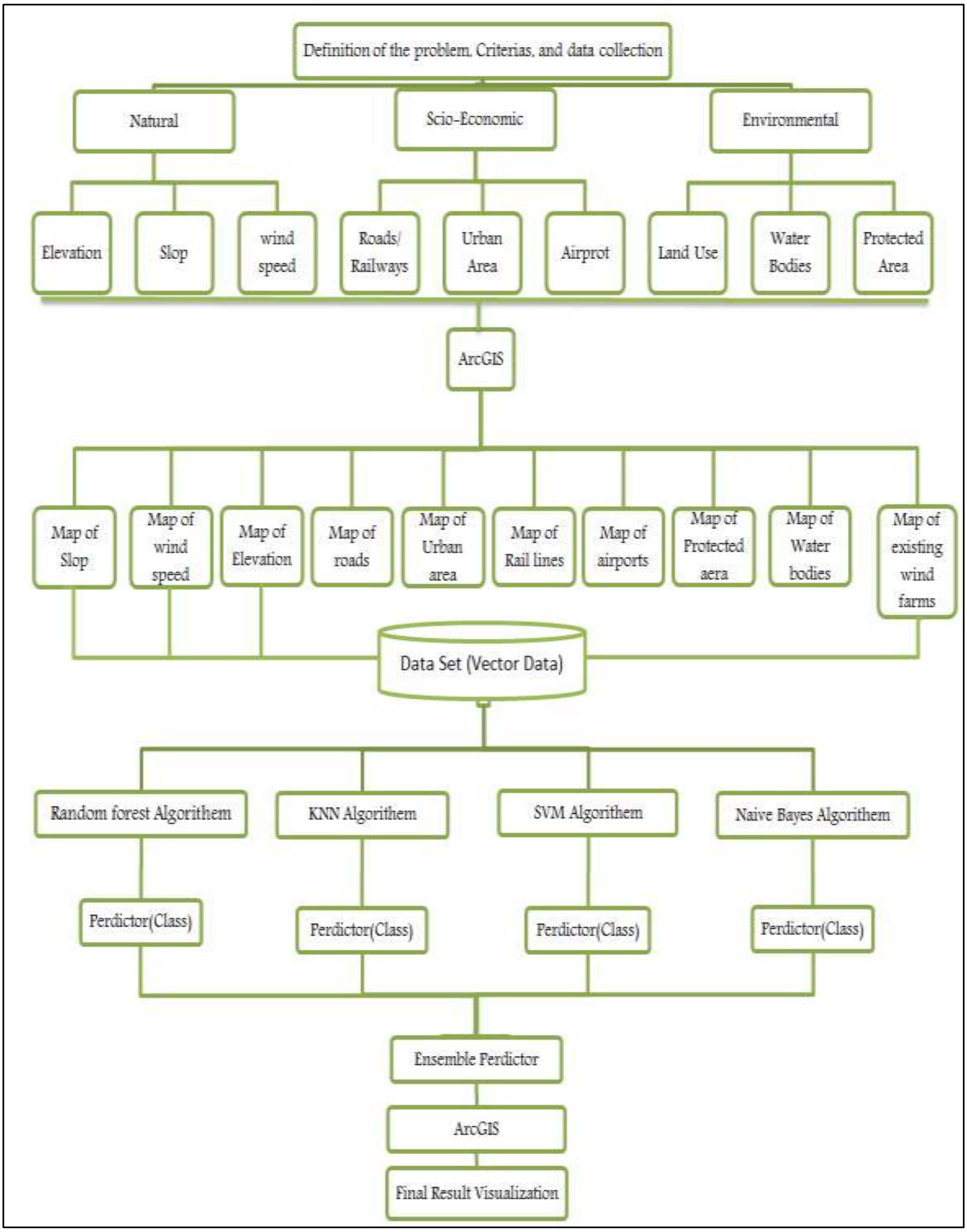


Figure 3.9: Framework Stages.

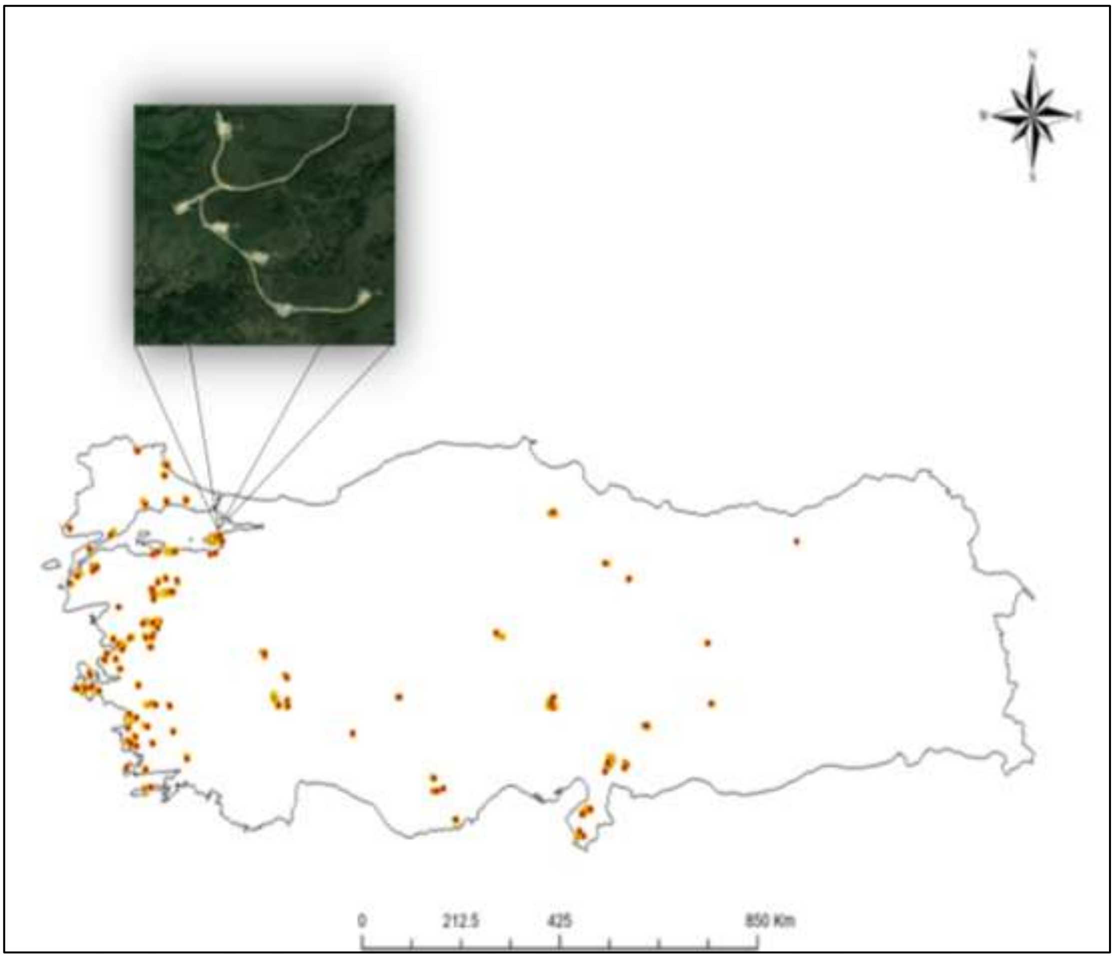


Figure 3.10: Existing wind farm sites in Türkiye.

4. FOURTH CHAPTER

4.1 MULTIVARIATE CLUSTERING ALGORITHM RESULT

The results of the algorithms (K-Means and K-Medoids) classified the study area sites into four clusters, Figure 4.1 which shows the clusters of the result output of the K-Means algorithm. The first cluster blue colour represents (39%) from the study area locations, the second red colour (14.6), the third green colour (28%), and the fourth yellow colour (18.6). Figure 4.2 shows the clusters of the K-Medoids algorithm output. The first cluster represents (31%) the study area sites, the second (18%), the third (22%), and the fourth (29%).

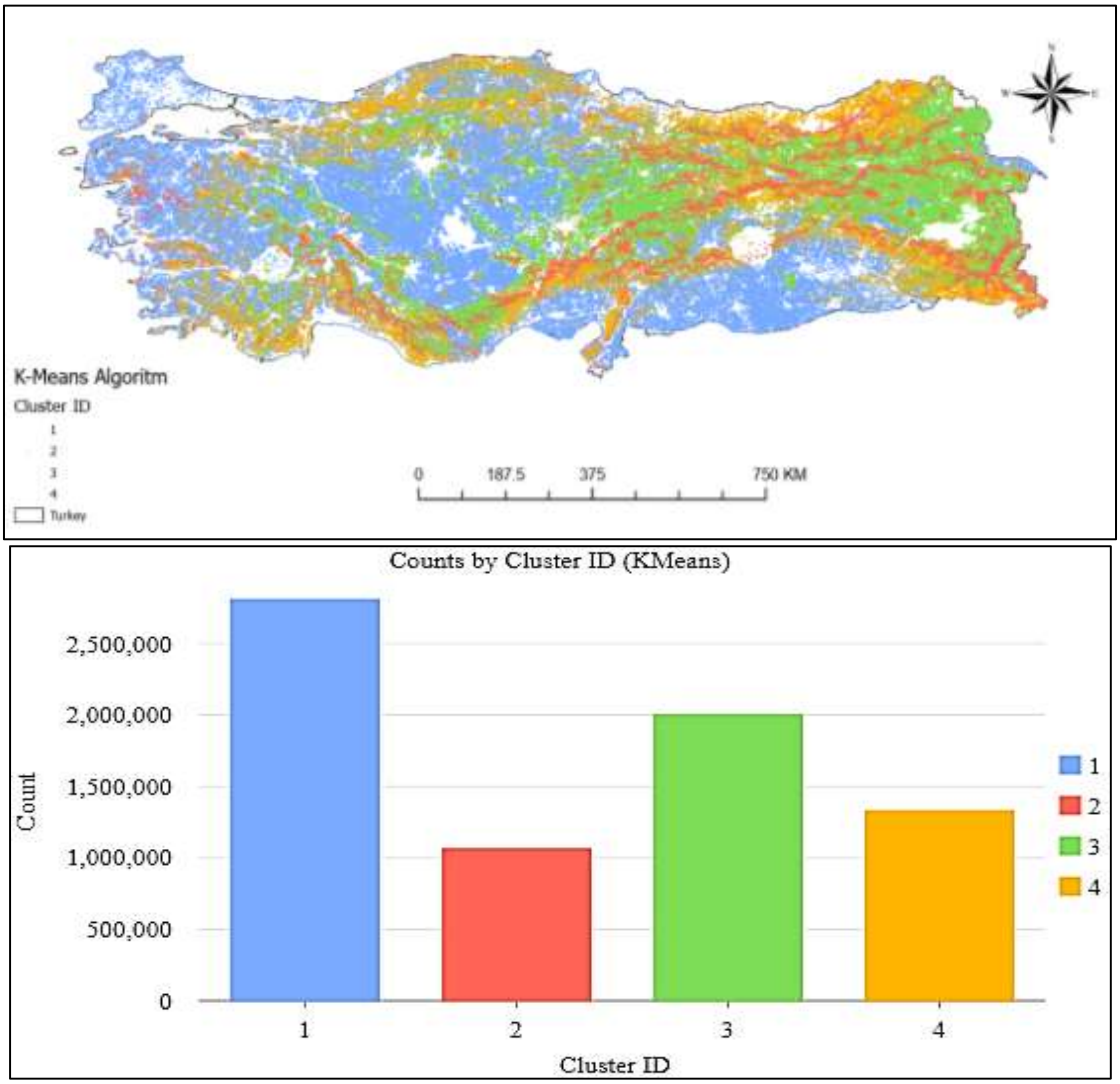


Figure 4.1: K-Means Algorithm Clusters.

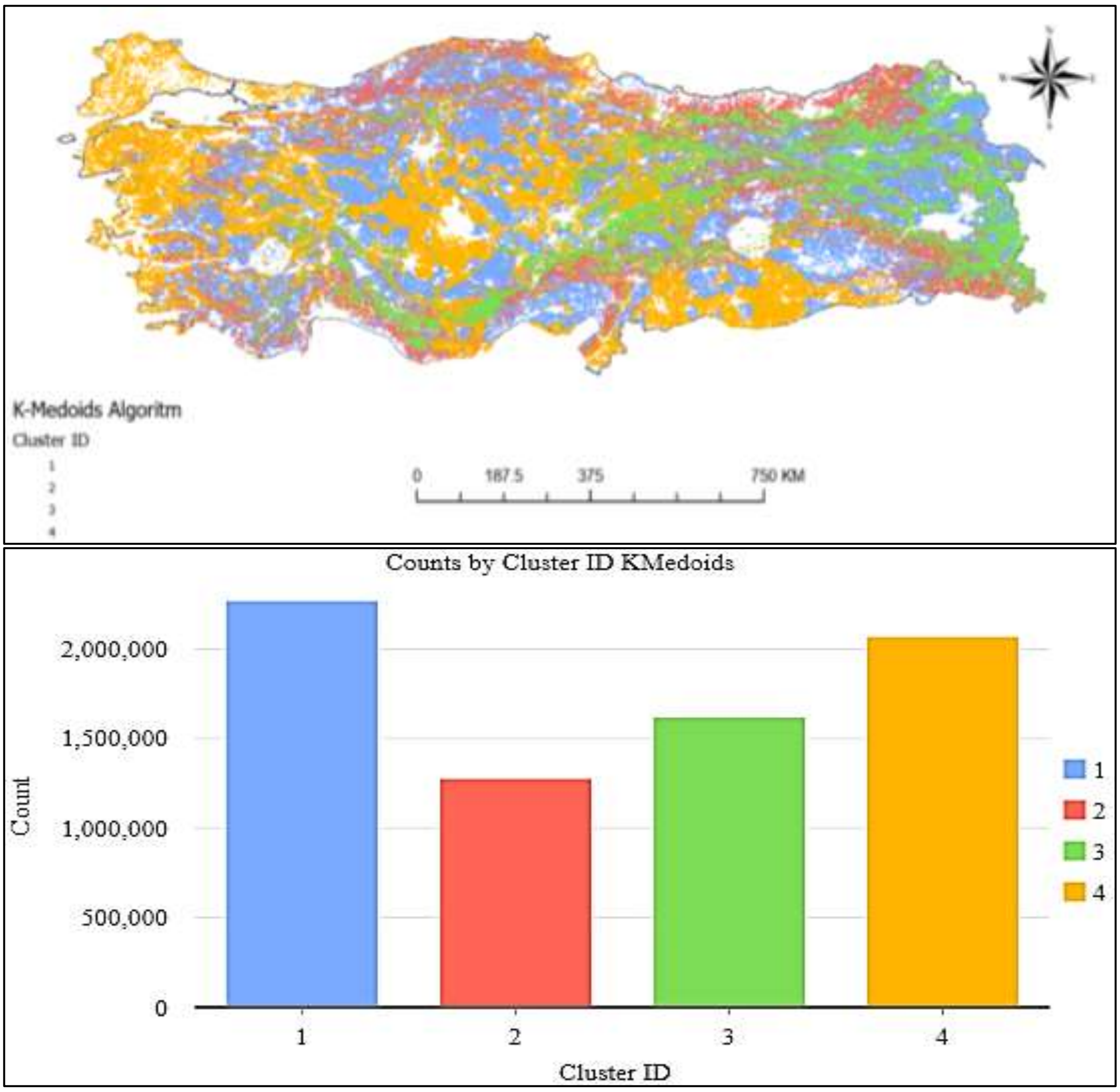


Figure 4.2: K-Medoids Algorithm Clusters**.**

Figure 4.3 and Figure 4.4 show the (Mean) values of wind speed, slope and elevation criteria for the four sets of results of the two algorithms (K-Means and K-Medoids), determining whether a site is suitable or unsuitable for wind farms.

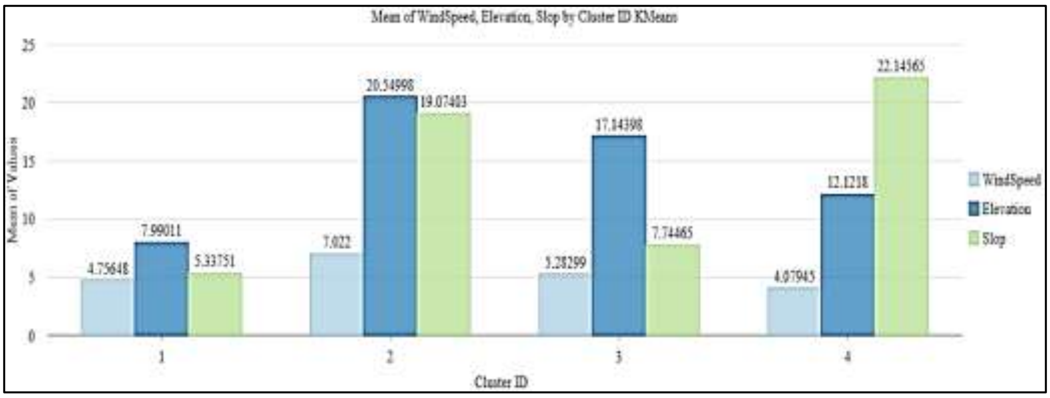


Figure 4.3: Mean Values of the Criteria K-Means Clusters**.**

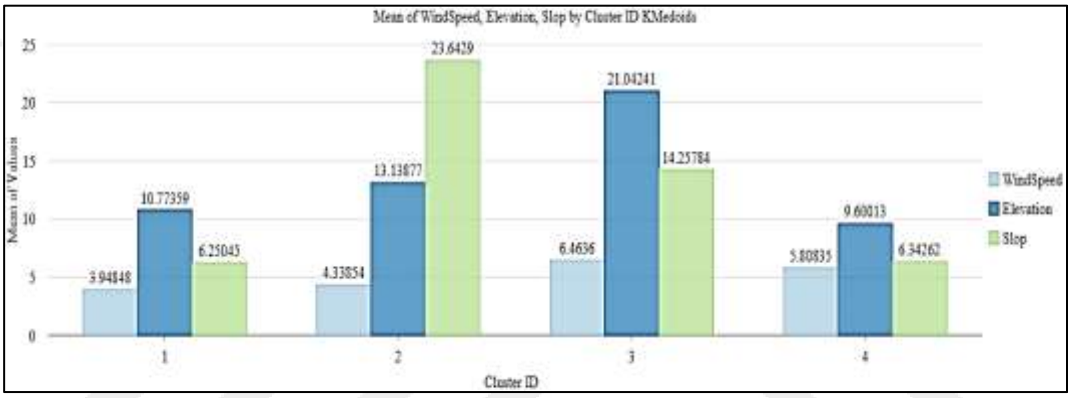


Figure 4.4: Mean Values of the Criteria K-Medoids Clusters**.**

Figure 4.5 and Figure 4.6 show the (Maximum) values of wind speed, slope and elevation criteria for the four clusters of results of the two algorithms (K-Means and K-Medoids), which help us comprehend the values of criteria that are outside or within the allowable range.

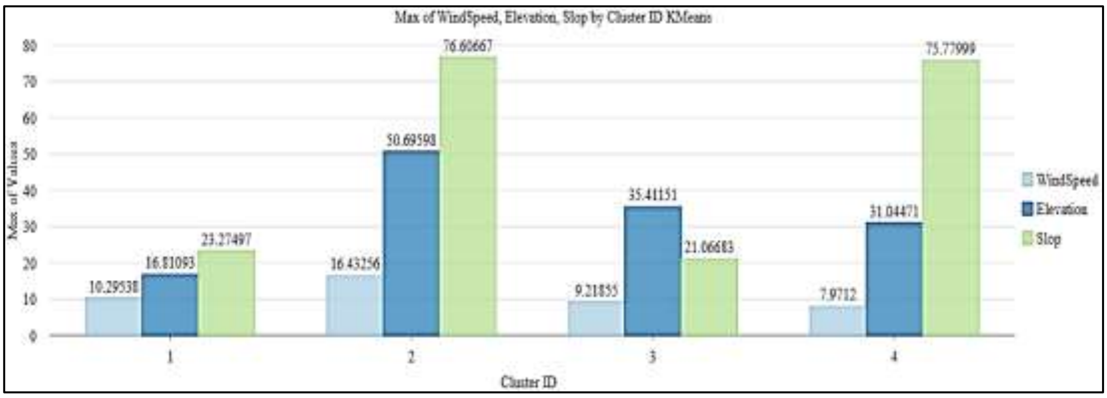


Figure 4.5: Maximum Values of the Criteria K-Means Clusters**.**

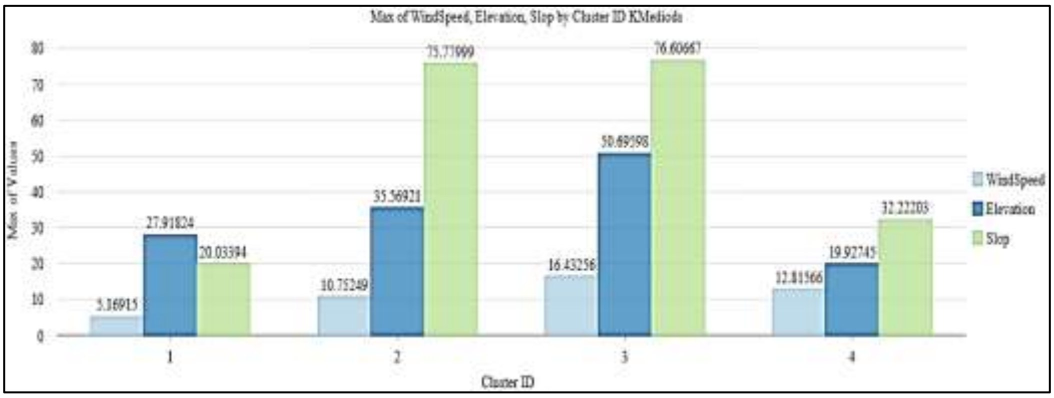


Figure 4.6: Maximum Values of the Criteria K-Medoids Clusters**.**

Figure 4.7 and Figure 4.8 show the (Minimum) values of wind speed, slope and elevation criteria for the four clusters of results of the two algorithms (K-Means and K-Medoids).

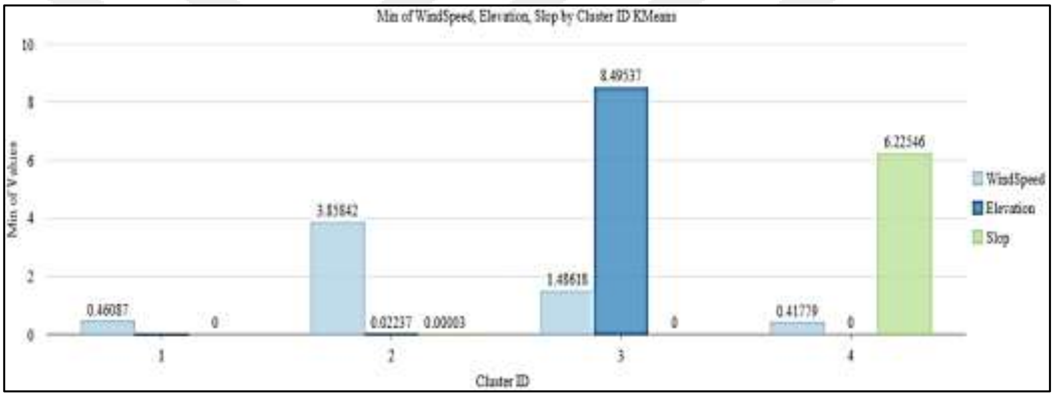


Figure 4.7: Minimum Values of the Criteria K-Means Clusters**.**

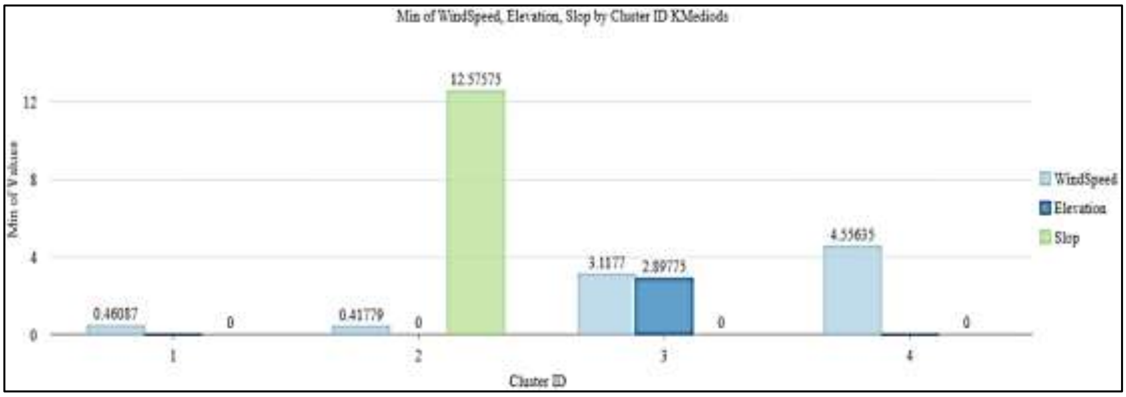


Figure 4.8: Minimum Values of the Criteria K-Medoids Clusters**.**

From the previous figures, the K-Means algorithm's results, summarised in Table 4.1, determined which clusters meet the requirements of the pre-defined criteria available at the site and are suitable or unsuitable for wind farms. Cluster No.1 averaged the criteria values within the specified criteria, especially the slope and elevation. The wind speed in the vast majority was higher than the minimum criteria specified at the sites, and therefore, the cluster is very suitable for wind farms. Cluster No. 2 wind speed was very suitable, but the slope and elevation values exceeded the criteria. However, their mean is within the specified criteria and suitable for wind farms. Cluster No. 3, according to the criteria values, the wind speed is suitable, but the elevation value is the highest. The average of the remaining values is within the criteria, and therefore, it is less suitable due to the slope rate and elevation value. Cluster 4, due to the high inclination rate, low wind rate and high height value, is unsuitable for wind farms, as shown in Figure 9.

Table 4.1: Result Criteria Value of K-Means Clusters.

Cluster	**Wind Speed**		**Mean**	**Elevation**		**Mean**	**Slop**		Mean	Percentage of area
	Max<25	Min>3.5		Max<(25*100)	Min>0		Max<30	Min>0		
1	10.29	0.46	4.75	16.81	0	7.99	23.27	0	5.33	39%
2	16.43	3.85	7.02	50.69	0.02	20.54	76.60	0	19.07	14.6%
3	9.21	1.48	5.28	35.41	8.49	17.14	21.06	0	7.74	28%
4	7.97	0.41	4.07	31.04	0	12.12.	75.77	6.22	22.14	18.4%

The K-Medoids algorithm's results are outlined in Table Cluster No. 1. Due to the low wind speed mean, this cluster is unsuitable for wind farms. . Cluster No.2, The slope and elevation values are outside the specified standards. However, their mean is within the specified criteria, so they are less suitable for wind farms. Cluster No. 3, according to the criteria values, the wind speed is suitable, and the elevation and slope values are high, but the mean is within the appropriate criteria. Therefore, they are suitable. Cluster No. 4 The wind speed and elevation values are very suitable and within the required criteria, and the slope is a suitable adjusted mean; therefore, this cluster is considered very suitable for wind farms, Figure 4.8.

Table 4.2: Result Criteria Value of K-Medoids Clusters.

Cluster	Wind Speed		Mean	Elevation		Mean	Slop		Mean	Percentage of area
	Max<25	Min>3.5		Max<(25*100)	Min>0		Max<30	Min>0		
1	5.16	0.46	3.94	27.91	0	10.77	20.03	0	6.25	31%
2	10.75	0.41	4.33	35.56	0	13.13	75.77	12.57	23.64	18%
3	16.43	3.11	6.46	50.69	2.89	21.04	76.6	0	14.25	22%
4	12.81	4.55	5.80	19.92	0	9.6	32.22	0	6.34	29%

4.2 K-NEAREST NEIGHBOR (K-NN) ALGORITHM RESULT

Using the data that represents the candidate sites for all the sites in the study area as a dataset to be the input for the algorithm, the algorithm classifies them as suitable or unsuitable sites based on the training data that represents the currently established wind farms. The result of the classification was as shown in Figure 4.9, with accuracy of 93.022%.

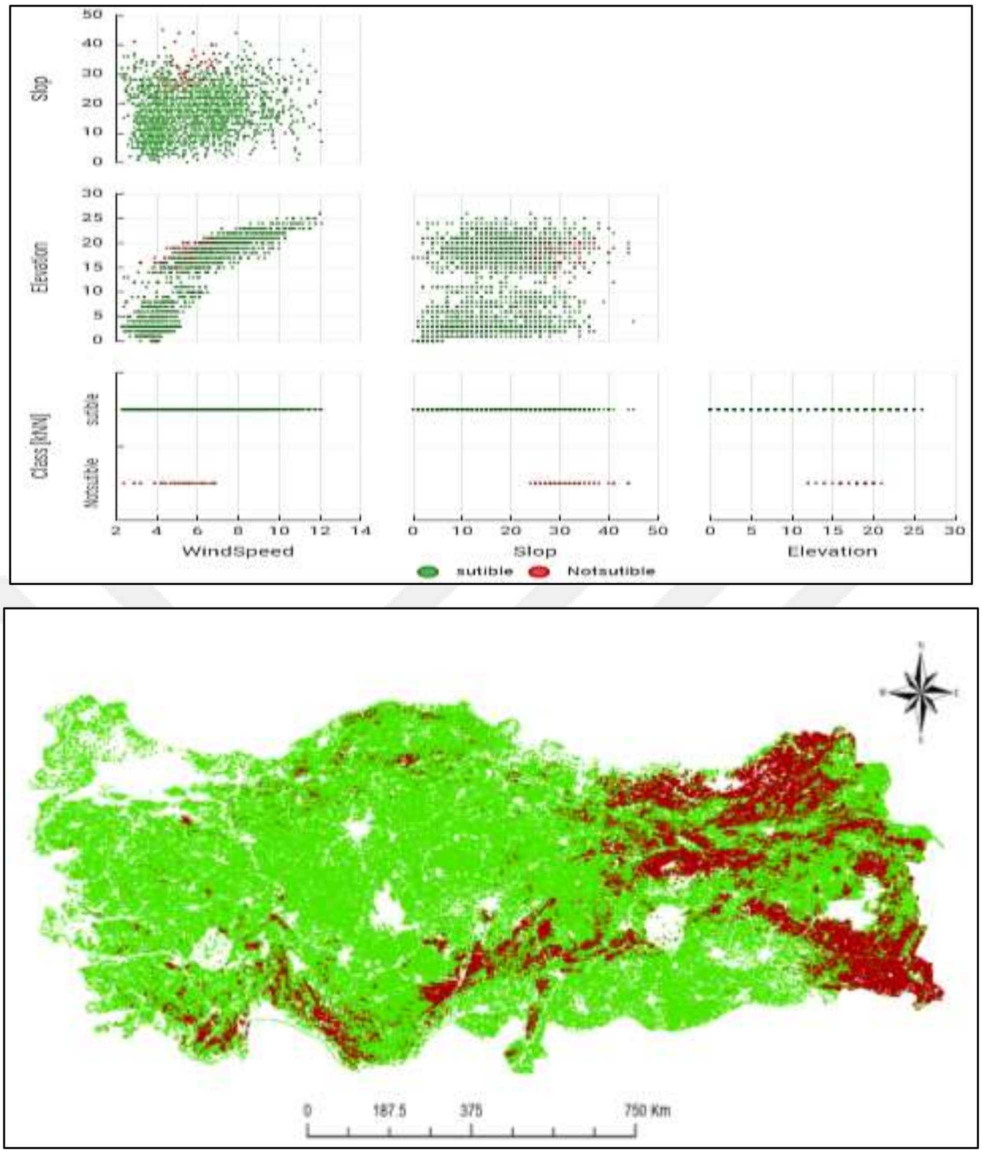


Figure 4.9: Result of K-NN Algorithm Classify.

Figure 4.9 represents the mean criteria values of (Wind Speed, Slop, and Elevation) for suitable and unsuitable sites obtained from the algorithm. In Figure 4.10, we notice that the algorithm classified some site values as suitable, even though they are outside the specified criteria. Figure 4.11 shows the classification result for the unsuitable. They were all within the criteria, as the slope criterion was decisive in making them unsuitable because its value was more than 30%. The error rate for the algorithm's classification result is 6.978%.

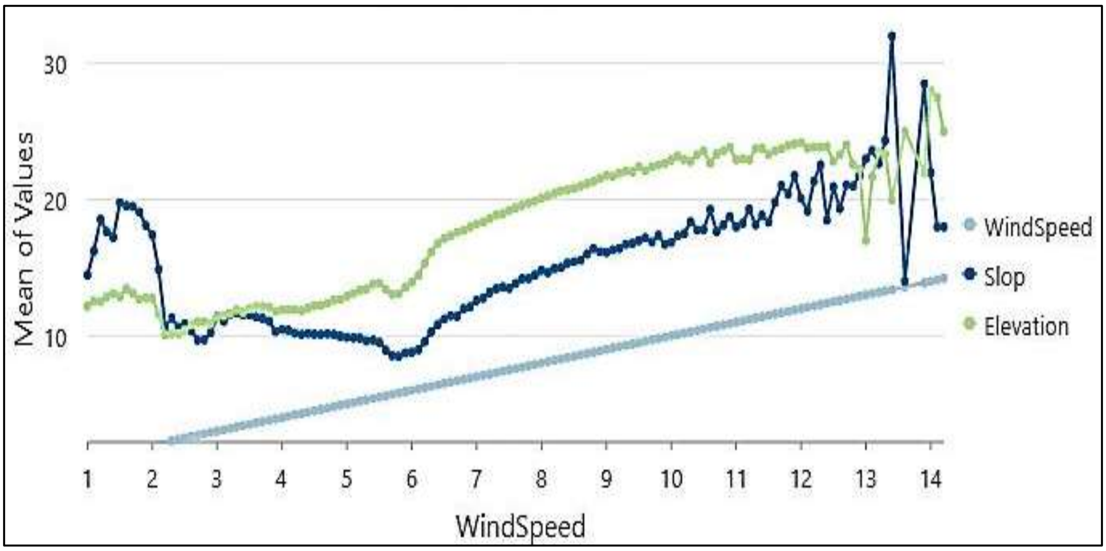


Figure 4.10: Mean of Wind Speed, Slop, and Elevation over Wind Speed, Suitable result (KNN).

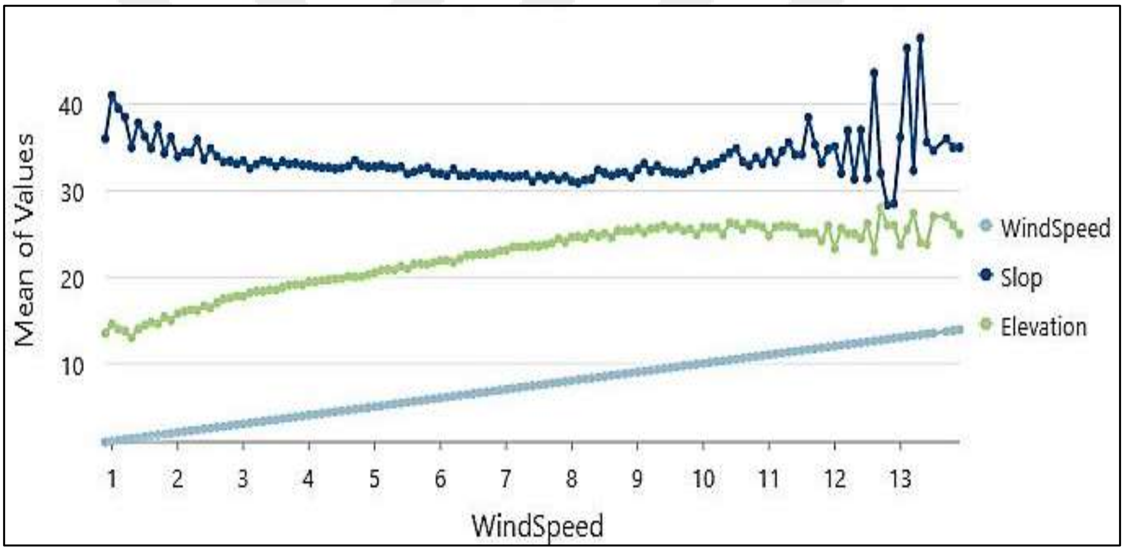


Figure 4.11: Mean of Wind Speed, Slop, and Elevation over Wind Speed, Unsuitable Result (KNN).

4.3 RANDOM FOREST ALGORITHM RESULT

Using the same training data and data set as in the previous algorithm, the results shown in Figure 4.12 were obtained with an accuracy of 93.018%.

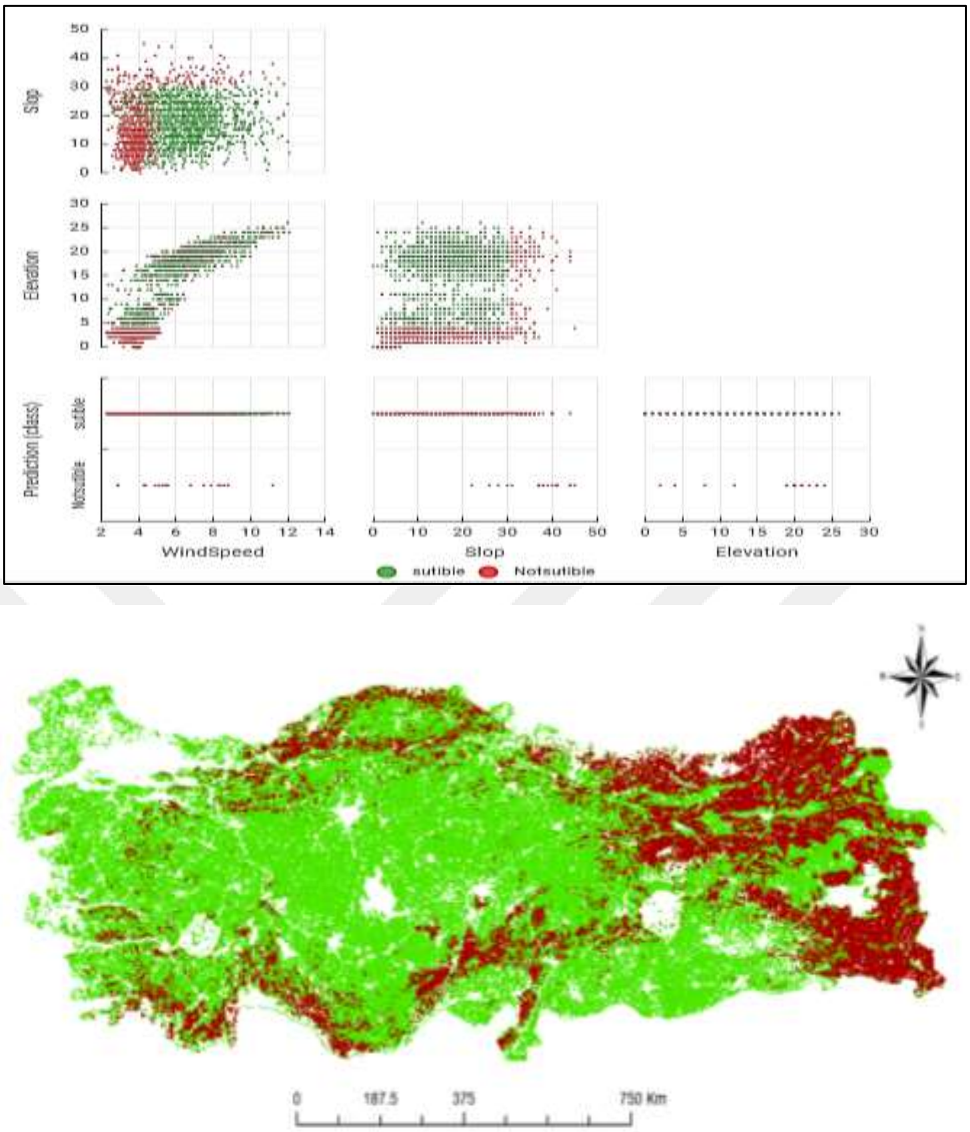


Figure 4.12: Result of Random Forest Algorithm Classify.

Figure 4.13 shows that some sites were classified as suitable even though they did not match the criteria adopted in the study, particularly in slow wind speeds. In Figure 4.14 some sites were classified as unsuitable even though their values were within the specified criteria. This normal has an error rate of 6.982% for several reasons, including the algorithm's bias, as some values are on the edge of the specified values for the criteria, making it fluctuate in choosing (suitable or unsuitable).

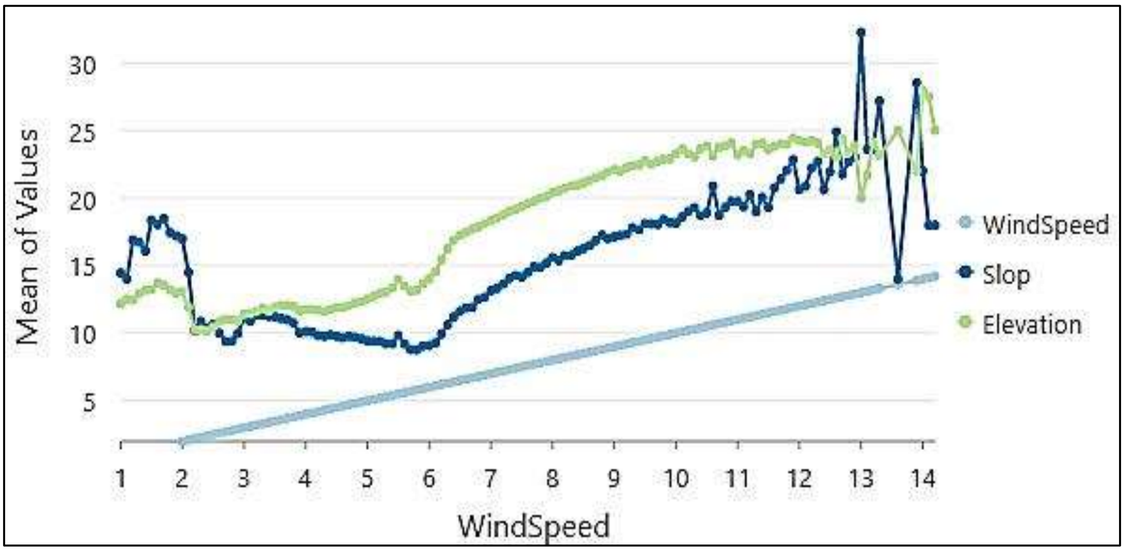


Figure 4.13: Mean of Wind Speed, Slop, and Elevation over Wind Speed, Suitable Result (RF).

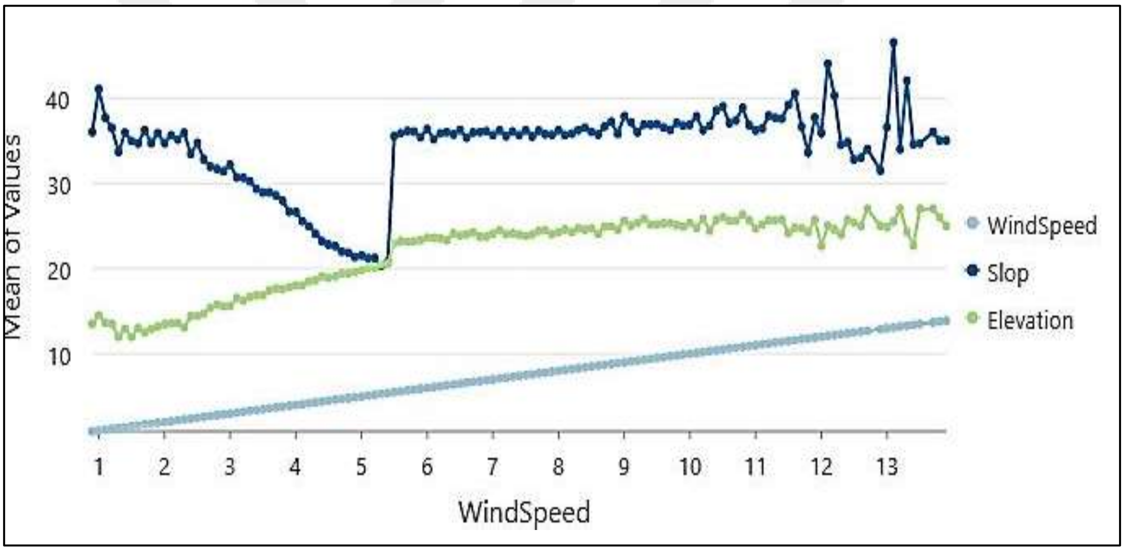


Figure 4.14: Mean of Wind Speed, Slop, and Elevation over Wind Speed, Unsuitable Result (RF).

4.4 SUPPORT VECTOR MACHINES ALGORITHM RESULT

The SVM algorithm used the same dataset and training data used as inputs in the previous algorithms, and the result is shown in Figure 4.15 with an accuracy of 95.095%.

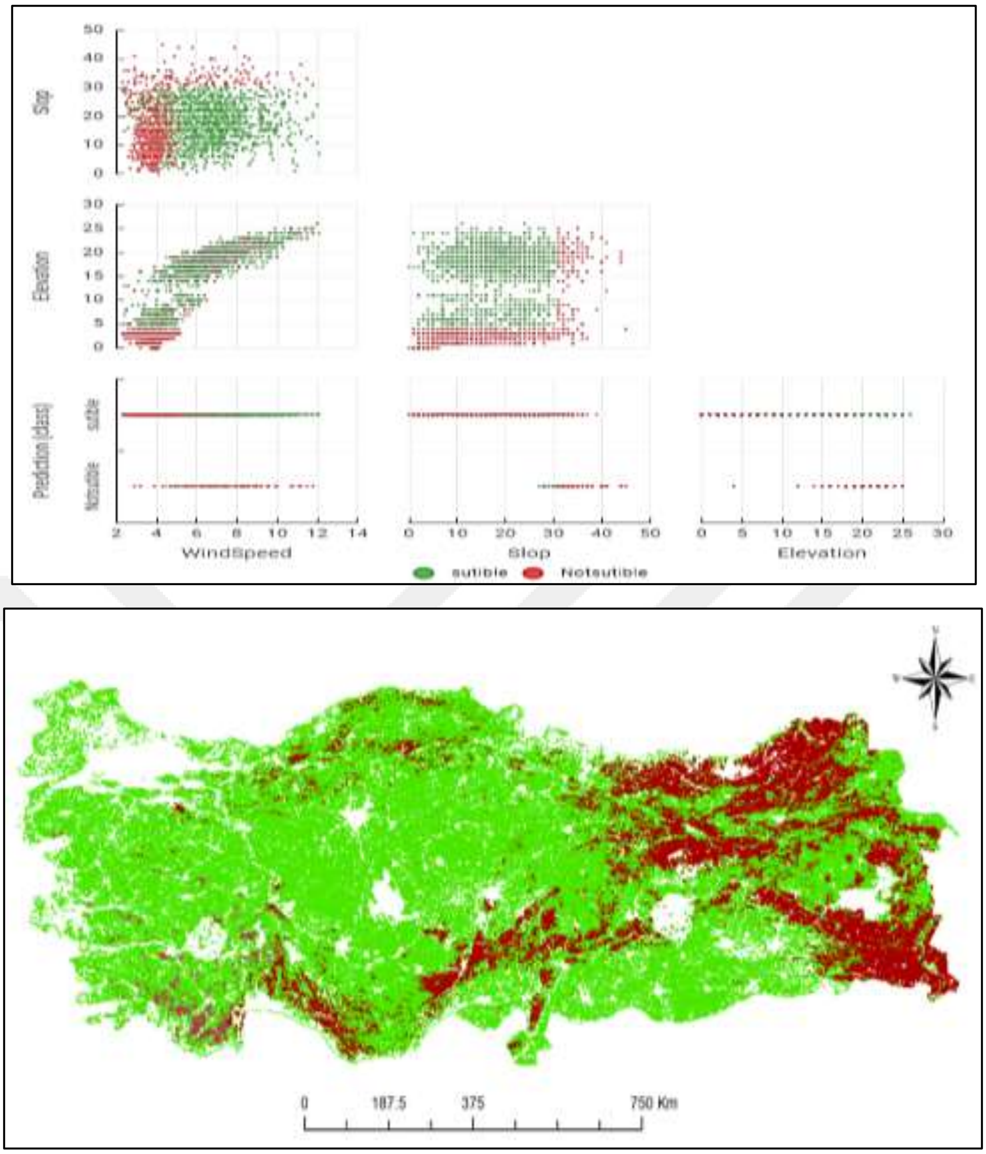


Figure 4.15: Result of Support Vector Machines Algorithm Classify.

Figure 4.16 represents the mean criteria values (Wind Speed, Slop, and Elevation) values for suitable and unsuitable sites obtained from SVM algorithm. It is close to the result of the (KNN) algorithm, with an error rate of 4.905%.

Figure 4.16: Mean of Wind Speed, Slop, and Elevation over Wind Speed, Suitable Result (SVM).

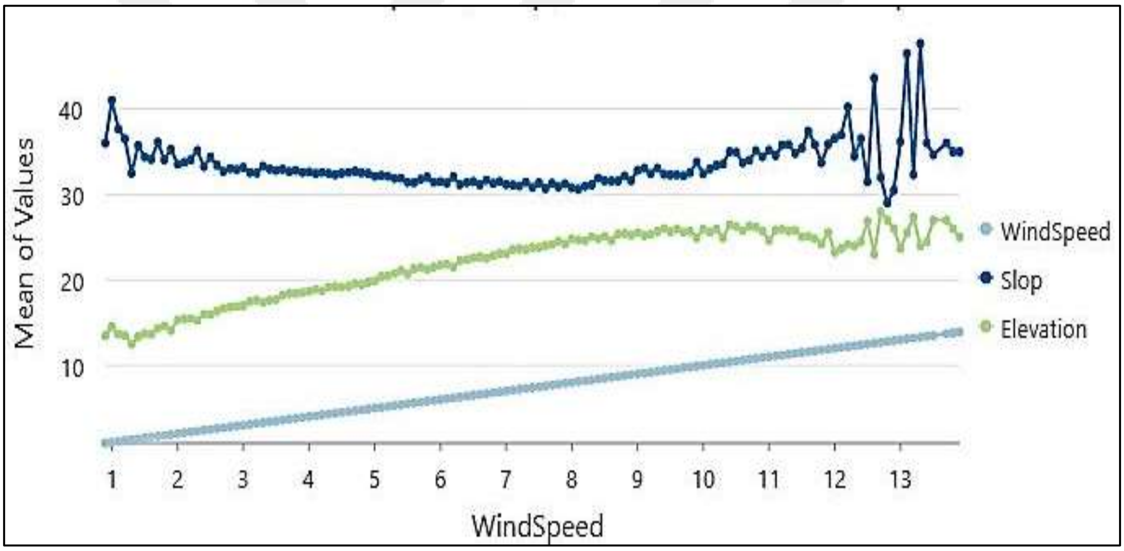


Figure 4.17: Mean of Wind Speed, Slop, and Elevation over Wind Speed, Unsuitable Result (SVM).

4.5 NAIVE BAYES ALGORITHM RESULT

The NAB algorithm used in this study to classify the data based on the same dataset and training data used as inputs in the previous algorithms, and the result is shown in Figure 4.18 with an accuracy of 89.553%,

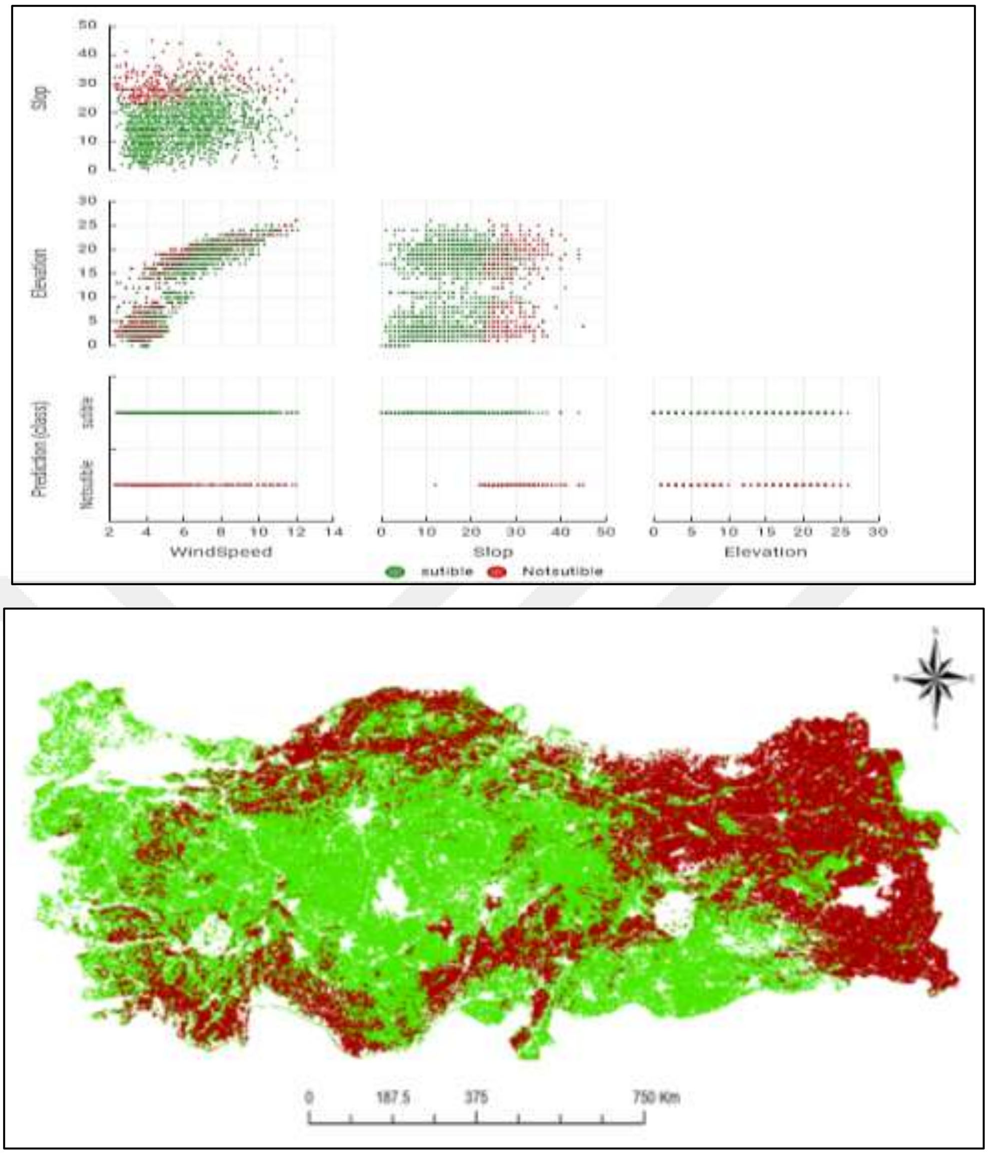


Figure 4.18: Result of Naive Bayes algorithm Classifies.

Figure 4.19 shows the results of the algorithm classify, which classified the sites as suitable, and most of the values were within the standards, except for some of them, where the low-speed wind factor was outside the criteria. Figure 4.20 shows that some sites were classified as unsuitable, although their values are ideal for suitable use. The reason is that the Naive Bayes algorithm has a conditional independence technique, which assumes that the features are independent. However, they are not independent in some cases, which leads to inaccurate results. The error rate for the algorithm's classification result is 10.447%.

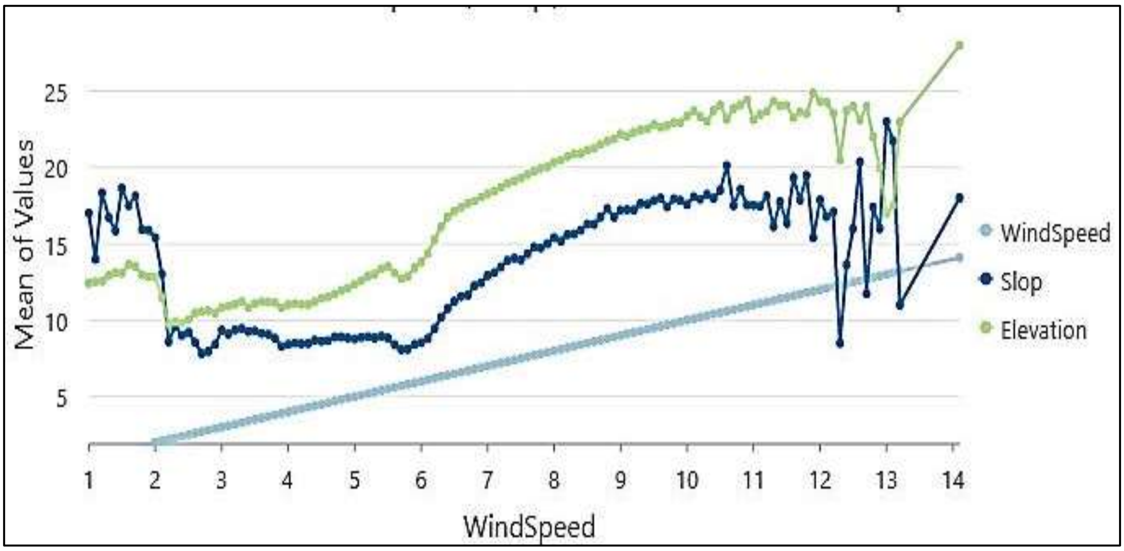


Figure 4.19: Shows the Mean of Wind Speed, Slope, and Elevation over Wind Speed, with an Unsuitable Result (NB).

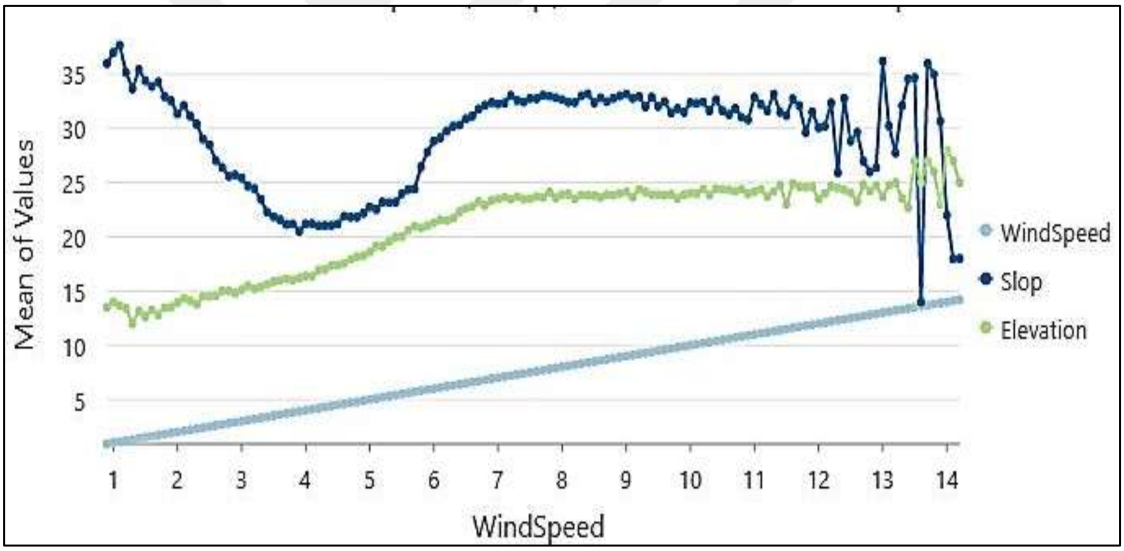


Figure 4.20: Shows the Mean of Wind Speed, Slope, and Elevation over Wind Speed, with an Unsuitable Result (NB).

Through the results we obtained from the algorithms (K-Nearest Neighbor, Random Forest, Support Vector Machines, Naive Bayes) that classified the results data into two categories, suitable and unsuitable, with varying degrees of accuracy and representation, it was found that the intersection between the results of the four algorithms has a result agreed upon by the four algorithms used, i.e. the site that was classified as suitable or unsuitable for wind farms is calculated as the final result. Figure 4.21 represents a map of the final sites, with suitable sites in green and unsuitable sites in red, which were agreed upon by the four algorithms, and this gave more accurate and reliable results than if a single algorithm was used.

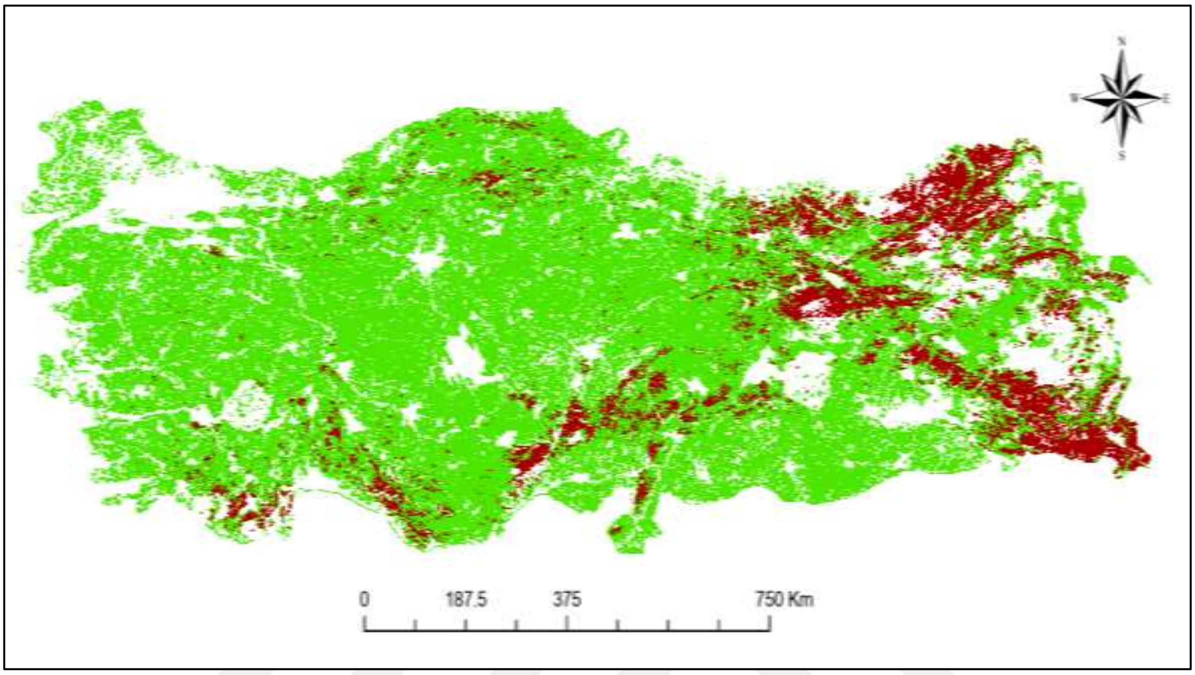


Figure 4.21: Shows a Map of Suitable and Unsuitable Sites Resulting from the Intersection of the Final Results of the Used Algorithms.

Figure 4.22 represents final result of the mean values of the criteria (wind speed, elevation, and slope) that represent suitable sites for wind farms, and their values are within the required criteria values in Table 1. The mean elevation value of suitable sites does not exceed 2500 metres (each elevation value in this study is multiplied by 100 in the real world). We note the mean slope value of suitable sites is within the required criteria, and the wind speed is not less than the speed required to start the turbine rotation, close to 3.5 metres per second and above.

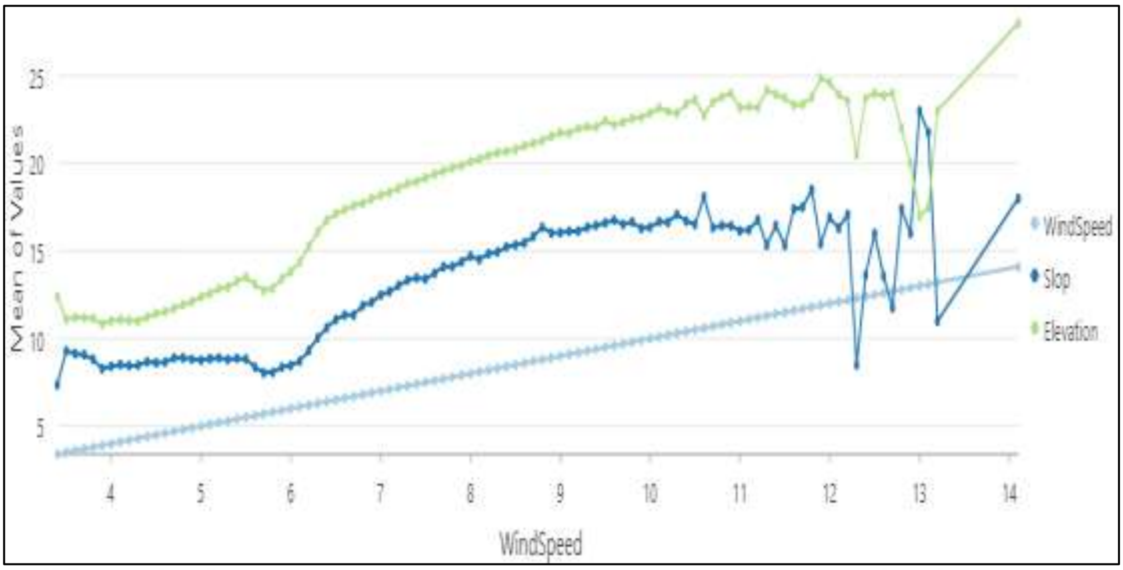


Figure 4.22: Mean Values of the Criteria of Suitable Sites.

Figure 4.23 represents the final result of the mean values of the criteria (wind speed, elevation and slope) for unsuitable sites, as the slope criterion was decisive in classifying the site as unsuitable for wind farms due to the nature of the study area's topography, while the wind speed criterion was within the appropriate criterion value in most sites. The elevation criterion was appropriate in a few sites, but the wind speed and slope criteria were not appropriate, and they were not appropriate in most sites. The site must meet all three criteria to be suitable for wind farms.

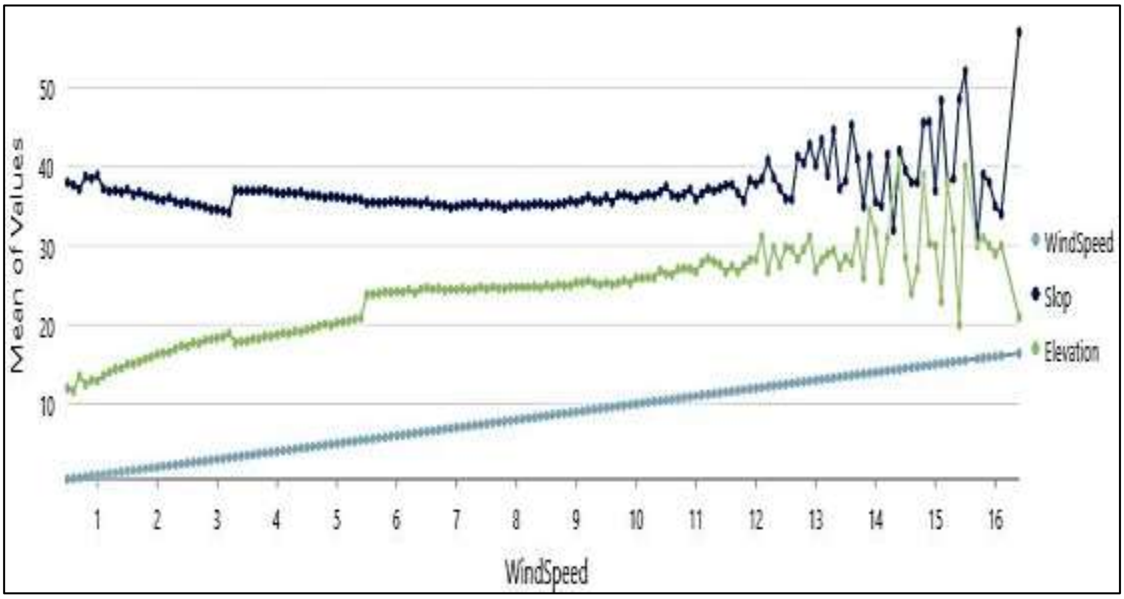


Figure 4.23: Mean values of the Criteria of Unsuitable Sites.

Figure 4.24 shows that all the sites of previously established wind farms in Türkiye fell within the areas suitable for wind farms we obtained in our study. This proves the validity and reliability of the study's results and the method used.

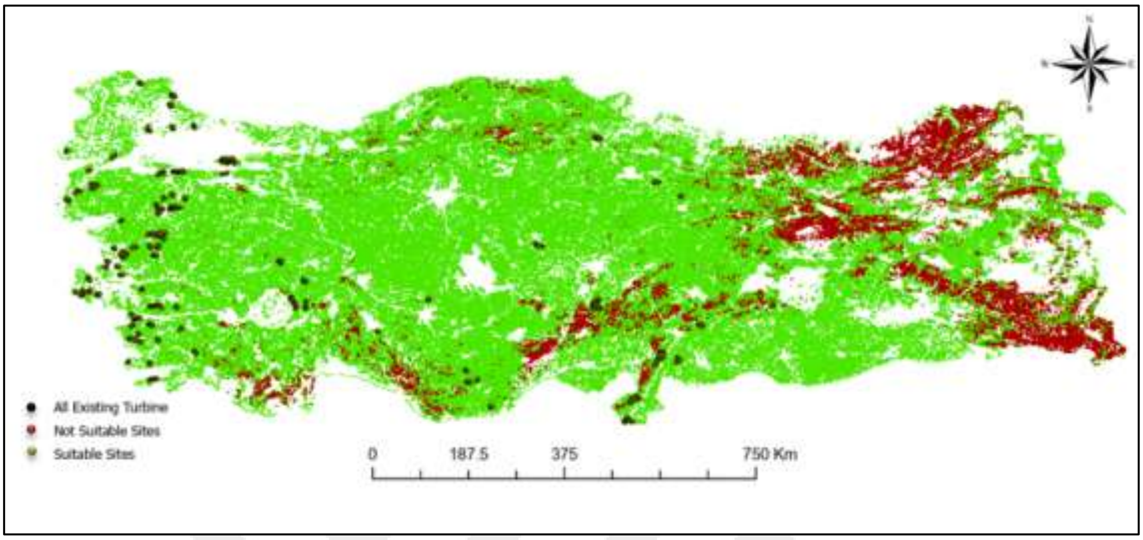


Figure 4.24: Previously Established Wind Farms in Türkiye.

5. FIFTH CHAPTER

5.1 CONCLUSION

The study showed the importance of shifting to sustainable energy with the abundance of its natural resources, including wind energy, in Türkiye. It improved the mechanism for selecting wind energy sites. Geographic information systems and their applications are essential in collecting, processing and analyzing several types of spatial data to form a data set or training data for many algorithms in the field of artificial intelligence. This study used the ensemble learning approach to find the intersection between the results of the algorithms to give accurate results that are free of variance and bias and the best results. This approach can be used in several other areas of sustainable energy, and the methodology followed in this research can be applied using supervised learning algorithms to other areas of study. The study showed that the (SVM) algorithm gave the highest accuracy (95.095%) due to its advantages in finding the optimal Decision Boundary between the features in the data set that separates the different features by a large distance. The (NB) algorithm gave the lowest accuracy (89.553%) due to the conditional independence technique, which assumes that the features are independent. Still, they are not independent sometimes, leading to inaccurate results. The possibility of using the dataset and training data generated from GIS processing of the raw data for this study, applying it using other artificial intelligence algorithms and with more criteria depending on the type of study.

5.2 FUTURE WORK

a. Integrating AI technologies with other applications opens up prospects for expanding its application in more areas

b. Combining real-time analysis with real-time data enhances prediction and choosing the right decision in regulating electricity consumption, wind speed predictions, self-maintenance, disaster prediction, etc.

c. Developing cognitive computing innovation models in sustainable energy and environmental conservation.

d. Improving multi-criteria decision analysis (MCDA) with more technology enhances ensemble learning of the site selection process.

REFRENCES

[1] A. M. Khalid, R. J. Ferguson, and M. N. Asadullah, "Economic Integration Among D-8 Muslim Countries: Prospects and Challenges," 2023.

[2] S. B. Selçuklu, D. Coit, and F. Felder, "Electricity generation portfolio planning and policy implications of Turkish power system considering cost, emission, and uncertainty," *Energy Policy*, vol. 173, p. 113393, 2023.

[3] Y. Xu *et al.*, "Site selection of wind farms using GIS and multi-criteria decision making method in Wafangdian, China," *Energy*, vol. 207, p. 118222, 2020.

[4] *Wind Energy Market in TürkiyeRoyal Danish General Consulate of Denmark in Istanbul*.

[5] I. E. Outlook, "World energy outlook special report https://www. iea. org/reports/india-energy-outlook-2021," ed: Accessed, 2022.

[6] R. Wiser *et al.*, "Land-based wind market report: 2023 edition," 2023.

[7] L. Zajac, "Eric Rosenbloom is a medical sciences editor and writer. He has been president of National Wind Watch, a clearinghouse for news and research concerning industrial-scale wind power, since 2006. THE CAREFUL PLANTING OF OFFSHORE WIND FARMS," *Wind Energy: A Reference Handbook: A Reference Handbook*, p. 163, 2014.

[8] L. Janicke, D. Nock, K. Surana, and S. M. Jordaan, "Air pollution co-benefits from strengthening electric transmission and distribution systems," *Energy*, vol. 269, p. 126735, 2023.

[9] "Wind Station Site Suitability Selection Using GIS ".

[10] W. Energy, "A Vision for Europe in 2030," *Advisory Council of the European Wind Energy Technology Platform*, pp. 1–64, 2006.

[11] N. Y. Aydın, "GIS-based site selection approach for wind and solar energy systems: a case study from Western Turkey," Middle East Technical University (Turkey), 2009.

[12] S. Parvin and M. Mayyas, "Wind energy market in USA," *European Journal of Sustainable Development Research*, vol. 7, no. 1, 2023.

[13] *Wind Energy Market in Türkiye*. [Online]. Available: https://tyrkiet.um.dk/en/-/media/country-sites/tyrkiet-en/the-trade-council/tc-Türkiye-windteaser-2021.ashx

[14] Q. Hassan *et al.*, "The renewable energy role in the global energy Transformations," *Renewable Energy Focus,* vol. 48, p. 100545, 2024.

[15] E. Sabev, R. Trifonov, G. Pavlova, and K. Rainova, "Cybersecurity analysis of wind farm SCADA systems," in *2021 International Conference on Information Technologies (InfoTech)*, 2021: IEEE, pp. 1–5.

[16] N. O. Farrar, M. H. Ali, and D. Dasgupta, "Artificial intelligence and machine learning in grid connected wind turbine control systems: A comprehensive review," *Energies,* vol. 16, no. 3, p. 1530, 2023.

[17] S. S. Yildiz, "Determining wind energy potential using geographic information system functions: a case study in Balıkesir, Turkey," *Applied Sciences,* vol. 13, no. 16, p. 9183, 2023.

[18] I. Yilmaz, "A hybrid DEA–fuzzy COPRAS approach to the evaluation of renewable energy: A case of wind farms in Turkey," *Sustainability,* vol. 15, no. 14, p. 11267, 2023.

[19] M. SvenSSon and J. Söderberg, "Machine-learning technologies in telecommunications," *Ericsson Review,* vol. 3, pp. 29–33, 2008.

[20] K.-L. Du, M. Swamy, K.-L. Du, and M. Swamy, "Support vector machines," *Neural networks and statistical learning,* pp. 469–524, 2014.

[21] N. H. Ovirianti, M. Zarlis, and H. Mawengkang, "Support Vector Machine Using A Classification Algorithm," *Sinkron: jurnal dan penelitian teknik informatika,* vol. 6, no. 3, pp. 2103–2107, 2022.

[22] T. N. Viet, H. Le Minh, L. C. Hieu, and T. H. Anh, "The Naïve Bayes algorithm for learning data analytics," *Indian Journal of Computer Science and Engineering,* vol. 12, no. 4, pp. 1038–1043, 2021.

[23] F.-J. Yang, "An implementation of naive bayes classifier," in *2018 International conference on computational science and computational intelligence (CSCI)*, 2018: IEEE, pp. 301–306.

[24] Y. Tang, Y. Chang, and K. Li, "Applications of K-nearest neighbor algorithm in intelligent diagnosis of wind turbine blades damage," *Renewable Energy,* vol. 212, pp. 855–864, 2023.

[25] A. M. Shahiri, W. Husain, and N. a. A. Rashid, "A review on predicting student's performance using data mining techniques," *procedia computer science,* vol. 72, pp. 414–422, 2015.

[26] S. Lee and J. Y. Chung, "The machine learning-based dropout early warning system for improving the performance of dropout prediction," *Applied Sciences,* vol. 9, no. 15, p. 3093, 2019.

[27] K. Fahd, S. J. Miah, and K. Ahmed, "Predicting student performance in a blended learning environment using learning management system interaction data," *Applied Computing and Informatics,* 2021.

[28] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS journal of photogrammetry and remote sensing,* vol. 67, pp. 93–104, 2012.

[29] F. Nji, R. Salvi, S. Tirumala, J. Wang, and X. Zheng, "Evaluation of Clustering Algorithms for Spatio-Temporal Multivariate Weather Data," Lawrence Livermore National Laboratory (LLNL), Livermore, CA (United States), 2022.

[30] J. Wu, *Advances in K-means clustering: a data mining thinking*. Springer Science & Business Media, 2012.

[31] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 881–892, 2002.

[32] V. Simankov *et al.*, "A solar and wind energy evaluation methodology using artificial intelligence technologies," *Energies*, vol. 17, no. 2, p. 416, 2024.

[33] A. Hamdan, K. I. Ibekwe, V. I. Ilojianya, S. Sonko, and E. A. Etukudoh, "AI in renewable energy: A review of predictive maintenance and energy optimization," *International Journal of Science and Research Archive*, vol. 11, no. 1, pp. 718–729, 2024.

[34] D. Song *et al.*, "Review on the application of artificial intelligence methods in the control and design of offshore wind power systems," *Journal of marine science and engineering*, vol. 12, no. 3, p. 424, 2024.

[35] M. D. Chaka *et al.*, "Improving wind speed forecasting at Adama wind farm II in Ethiopia through deep learning algorithms," *Case Studies in Chemical and Environmental Engineering*, vol. 9, p. 100594, 2024.

[36] S. Hanifi, A. Cammarono, and H. Zare-Behtash, "Advanced hyperparameter optimization of deep learning models for wind power prediction," *Renewable Energy*, vol. 221, p. 119700, 2024.

[37] I. Yilmaz, A. Adem, and M. Dağdeviren, "A machine learning-integrated multi-criteria decision-making approach based on consensus for selection of energy storage locations," *Journal of Energy Storage*, vol. 69, p. 107941, 2023.

[38] M. Asadi, M. Ramezanzade, and K. Pourhossein, "A global evaluation model applied to wind power plant site selection," *Applied Energy*, vol. 336, p. 120840, 2023.

[39] M. A. Raza, M. Yousif, M. Hassan, M. Numan, and S. A. A. Kazmi, "Site suitability for solar and wind energy in developing countries using combination of GIS-AHP; a case study of Pakistan," *Renewable Energy*, vol. 206, pp. 180–191, 2023.

[40] C. V. Ekeanyanwu, I. F. Obisakin, P. Aduwenye, and N. Dede-Bamfo, "Merging GIS and machine learning techniques: a paper review," *Journal of Geoscience and Environment Protection,* vol. 10, no. 9, pp. 61–83, 2022.

[41] L. H. Hassan, M. Moghavvemi, H. A. Almurib, and O. Steinmayer, "Current state of neural networks applications in power system monitoring and control," *International Journal of Electrical Power & Energy Systems,* vol. 51, pp. 134–144, 2013.

[42] K. Rahul, R. K. Banyal, P. Goswami, and V. Kumar, "Machine learning algorithms for big data analytics," in *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*, 2021: Springer, pp. 359–367.

[43] G. Villacreses, G. Gaona, J. Martínez-Gómez, and D. J. Jijón, "Wind farms suitability location using geographical information system (GIS), based on multi-criteria decision making (MCDM) methods: The case of continental Ecuador," *Renewable energy,* vol. 109, pp. 275–286, 2017.

[44] T. R. Ayodele, A. Ogunjuyigbe, O. Odigie, and J. L. Munda, "A multi-criteria GIS based model for wind farm site selection using interval type-2 fuzzy analytic hierarchy process: The case study of Nigeria," *Applied energy,* vol. 228, pp. 1853–1869, 2018.

[45] J. J. Diez-Rodríguez, T. B. Fischer, and S. Di Zio, "Introducing a group spatial decision support system for use in strategic environmental assessment of onshore wind farm development in Mexico," *Journal of Cleaner Production,* vol. 220, pp. 1239–1254, 2019.

[46] S. Rehman, M. Baseer, and L. Alhems, "GIS-based multi-criteria wind farm site selection methodology," *FME Trans,* vol. 48, no. 4, pp. 855–867, 2020.

[47] A. Mohamed El-hosiny, H. El-Ghareeb, B. T. Shabana, and A. AbouElfetouh, "Sustainable energy production’s spatial determination framework, based on multi criteria decision making and geographic information system under neutrosophic environment: a case study in Egypt," *Neutrosophic Sets and Systems,* vol. 42, no. 1, p. 9, 2021.

[48] F. M. I. Flora, N. Donatien, R. Tchinda, and O. Hamandjoda, "Selection wind farm sites based on GIS using a boolean method: evaluation of the case of Cameroon," *Journal of Power and Energy Engineering*, vol. 9, no. 1, pp. 1–24, 2021.

[49] L. Albraheem and L. AlAwlaqi, "Geospatial analysis of wind energy plant in Saudi Arabia using a GIS-AHP technique," *Energy Reports,* vol. 9, pp. 5878–5898, 2023.

[50] S. Karamountzou and D. G. Vagiona, "Suitability and sustainability assessment of existing onshore wind farms in Greece," *Sustainability*, vol. 15, no. 3, p. 2095, 2023.

[51] M. Khan, M. Rashid, and I. Ahmad, "Hybrid decision support system for wind farm site selection using fuzzy AHP and F-TOPSIS," *Renewable Energy*, vol. 147, pp. 1255–1271, 2020.

[52] P. D. Cerna *et al.*, "Wind power plant site selection using integrated machine learning and multiple-criteria decision making technique," in *E3S Web of Conferences*, 2023, vol. 405: EDP Sciences, p. 02030.

[53] K. F. Sotiropoulou, A. P. Vavatsikos, and P. N. Botsaris, "A hybrid AHP-PROMETHEE II onshore wind farms multicriteria suitability analysis using kNN and SVM regression models in northeastern Greece," *Renewable Energy*, vol. 221, p. 119795, 2024.

[54] P. V. Gorsevski, S. C. Cathcart, G. Mirzaei, M. M. Jamali, X. Ye, and E. Gomezdelcampo, "A group-based spatial decision support system for wind farm site selection in Northwest Ohio," *Energy Policy,* vol. 55, pp. 374–385, 2013.

[55] K. Xenitidis, K. Ioannou, G. Tsantopoulos, and D. Myronidis, "On the usage of artificial neural networks for the determination of optimal wind farms allocation," *Sustainability,* vol. 15, no. 24, p. 16938, 2023.

[56] G. Demir, M. Riaz, and M. Deveci, "Wind farm site selection using geographic information system and fuzzy decision making model," *Expert systems with applications,* vol. 255, p. 124772, 2024.

[57] A. Demir, A. E. Dinçer, C. Çiftçi, S. Gülçimen, N. Uzal, and K. Yılmaz, "Wind farm site selection using GIS-based multicriteria analysis with Life cycle assessment integration," *Earth Science Informatics*, vol. 17, no. 2, pp. 1591–1608, 2024.

[58] H. Zhao, S. Wang, and C. Lu, "A study on site selection of wind power plant based on prospect theory and VIKOR: a case study in China," *Kybernetes*, 2024.

[59] S. Gündüz, "Determination Of Last Glacial Equilibrium Line Altitudes Of Turkey By Glacial Mass Balance Model," 2018.

[60] [Online]. Available: https://web.uflib.ufl.edu/fhmaps/basic_result.asp

[61] H. Yousefi, S. Moradi, R. Zahedi, and Z. Ranjbar, "Developed analytic hierarchy process and multi criteria decision support system for wind farm site selection using GIS: A regional-scale application with environmental responsibility," *Energy Conversion and Management: X*, vol. 22, p. 100594, 2024.

[62] R. Van Haaren and V. Fthenakis, "GIS-based wind farm site selection using spatial multi-criteria analysis (SMCA): Evaluating the case for New York State," *Renewable and sustainable energy reviews*, vol. 15, no. 7, pp. 3332–3340, 2011.

[63] M. Van Grieken and B. Dower, "Wind turbines and landscape," in *Wind energy engineering*: Elsevier, 2023, pp. 443–462.

[64] M. Honarkhah and J. Caers, "Stochastic simulation of patterns using distance-based pattern modeling," *Mathematical Geosciences*, vol. 42, pp. 487–517, 2010.

[65] X. Zheng, Q. Lei, R. Yao, Y. Gong, and Q. Yin, "Image segmentation based on adaptive K-means algorithm," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 1–10, 2018.