|  |  |
| --- | --- |
| **Power Consumption Model** | |
|  | |
|  |  |
| **Keywords:**  *Power consumption, energy, energy consumption estimation, machine learning, artificial intelligence, random forest model.* | **Abstract:** Our study was conducted on the data set of the city of Tétouan, which includes power consumption values for zone one, zone two, zone three according to temperature, humidity, wind speed, general diffusion rate and diffusion rate measured every 10 minutes or every hour. With this data set, research has been carried out on machine learning, model training, training of different models, and examining and deciding which model has a better performance and score value among the trained models. As a result of the evaluations, one-week and one-month power consumption estimates were made for zone one, zone two and zone three depending on time, humidity, wind speed, general diffusion rate, diffusion rate, using the algorithm (model) with the best performance. In this study, five different models were used for the estimation of the energy consumption of Tétouan City, taking into account the important role of power consumption in the world economy: Linear Regression, Decision Trees, K-Nearest Neighbor, Support Vector Model, and Random Forest. The functionality of the models was validated with an open-source dataset and evaluated using different performance metrics. The test results show that the Random Forest model has the lowest error rate between measured and estimated energy consumption. In addition, the Random Forest model requires a shorter training and testing time compared to other models. Therefore, the Random Forest model provides an effective and accurate solution for power consumption estimation. |
|  |  |
|  |  |
| **Power Consumption Prediction using Random Forest Model** | |
|  | |
|  | |
| **Keywords:**  Power consumption, energy, energy estimation, machine learning, artificial intelligence, random forest model. | **Abstract:** Our study was conducted using a dataset containing power consumption values for region one, region two, and region three based on temperature, humidity, wind speed, general diffusion rate, and diffusion rate measured every 10 minutes or hourly in the city of Tétouan. Machine learning techniques, model training, and evaluation of different models were performed to determine which model performed better based on their score values. The best-performing algorithm (model) was used to predict weekly and monthly power consumption for region one, region two, and region three, considering time, humidity, wind speed, general diffusion rate, and diffusion rate. In this study, considering the significant role of power consumption in the global economy, five different models were used for predicting energy consumption in Tétouan City: Linear Regression, Decision Trees, K-Nearest Neighbors, Support Vector Model, and Random Forest. The functionality of the models was verified using an open-source dataset and evaluated using different performance metrics. The test results indicate that the Random Forest model has the lowest error rate between the measured and predicted energy consumption. Additionally, the Random Forest model requires shorter training and testing time compared to other models. Therefore, the Random Forest model provides an effective and accurate solution for power consumption prediction. |
|  |  |

# 1. Introduction

According to the 2014 Census, Tétouan is a city located in Morocco with an area of 11,570 km², with a population of 380,000. The city is located in the northern part of the country and overlooks the Mediterranean Sea. The weather is especially hot and humid in summer.

According to the state-owned ONEE (National Office for Electricity and Drinking Water), the majority of Morocco's energy production in 2019 consisted of coal (38%). It is followed by hydropower (16%), fuel oil (8%), natural gas (18%), wind (11%), solar (7%) and others (2%). [1]

Given the existence of a strong reliance on non-renewable sources (64%), forecasting energy consumption can help stakeholders better manage their purchases and stocks. On top of that, Morocco's plan is to reduce energy imports by increasing production from renewable sources. It is known that there is a risk that resources such as wind and solar may not be available throughout the year.

Starting from a medium-sized city to understand the country's energy needs can be a big step in planning for these resources.

# 2. materıals and methods

## 

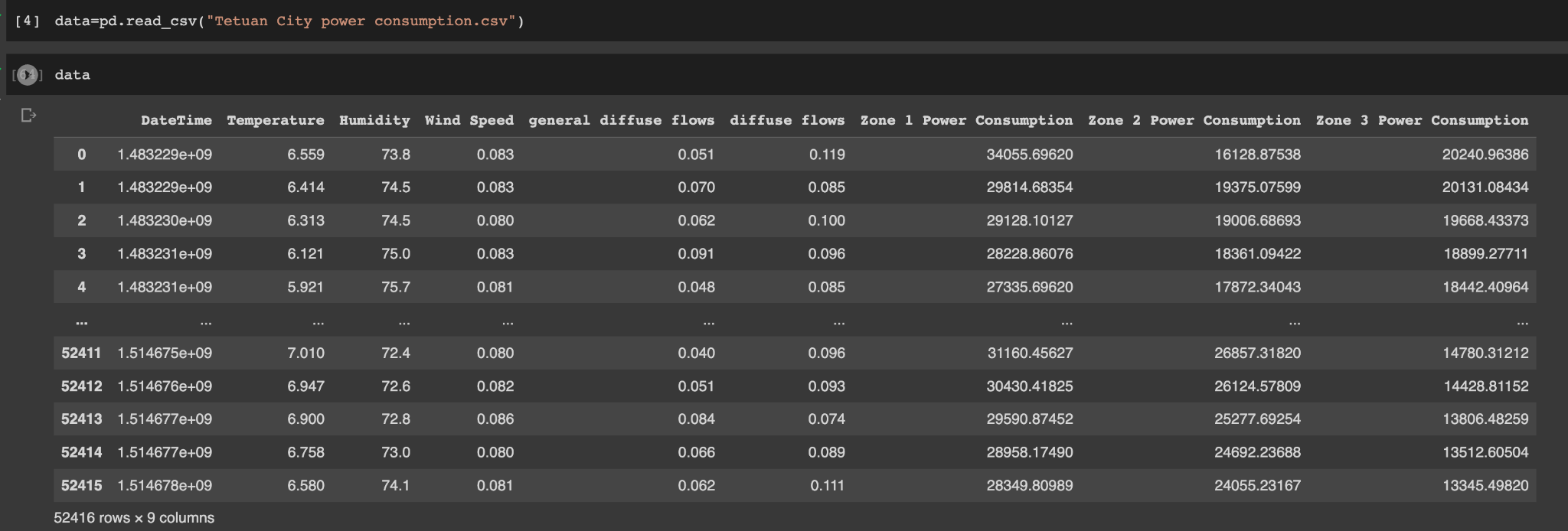
## 2.1. Data Set

Power consumption of Tetouan city Veri Seti (UCI)

# Since the beginning of 2017, our dataset has been measured at ten-minute intervals through sensors:

It consists of Temperature, Humidity, Wind Speed, General Scattered Flows, Scattered Flows, Zone 1 Power Consumption, Zone 2 Power Consumption, Zone 3 Power Consumption values.

**Figure 1.** A portion of the data set.



20% of the data was used for testing and 80% for training.

## 2.2. Models Used

**Random Forest model:**

* Data Collection:

First, we need to collect the power consumption data in Tetouan City. This data can often be recorded over a specific period of time (at intervals of ten minutes and one hour). In addition to power consumption data, other potential arguments (e.g., air temperature, time zone, wind) should also be recorded.

* Data Editing:

We must organize the data collected. We must clean up the data with incomplete or anomalies and present it in an appropriate format. For example, we need to convert the date and time information into a convenient time format.

* Determination of Dependent and Independent Variables:

Power consumption will be our dependent variable. We must identify other potential influencing factors and consider them as independent variables. For example, factors such as air temperature, time zone, wind can affect the power consumption of Tetouan City.

* Data Analysis and Model Building:

Using our dataset, we can create the Random Forest model. The Random Forest is an ensemble model and is a structure formed by the combination of many decision trees.

During model training, the Random Forest generates multiple decision trees with random samples (bootstrap) and feature sampling. Each tree is trained with a different subset of the dataset and uses a random subset of the independent variables. [3]

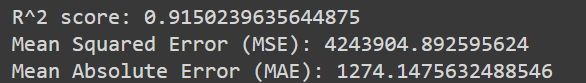
The model makes the final prediction by combining the predictions of each tree. In this way, the Random Forest model identifies complex relationships in the dataset to estimate the power consumption of Tetouan City.

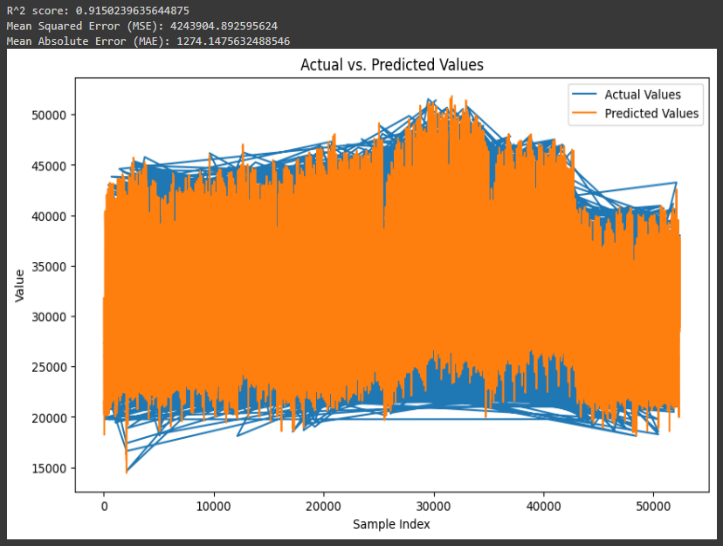
When a new data point arrives, the Random Forest model feeds that data point to each tree and combines the predictions from each tree to estimate power consumption.

The Random Forest model also provides variable importance ordering in the dataset. In this way, we can determine which independent variables affect power consumption more. For example, the air temperature variable may be the most important factor in power consumption, while other variables such as time zone or wind may be less important.

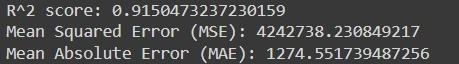
In conclusion, the Random Forest model is an effective method used to estimate the power consumption of Tetouan City. This model can capture complex relationships in the data set, provides variable importance ranking, and is resistant to overfitting. When new data comes in, we can use the model to make an effective estimate of power consumption.

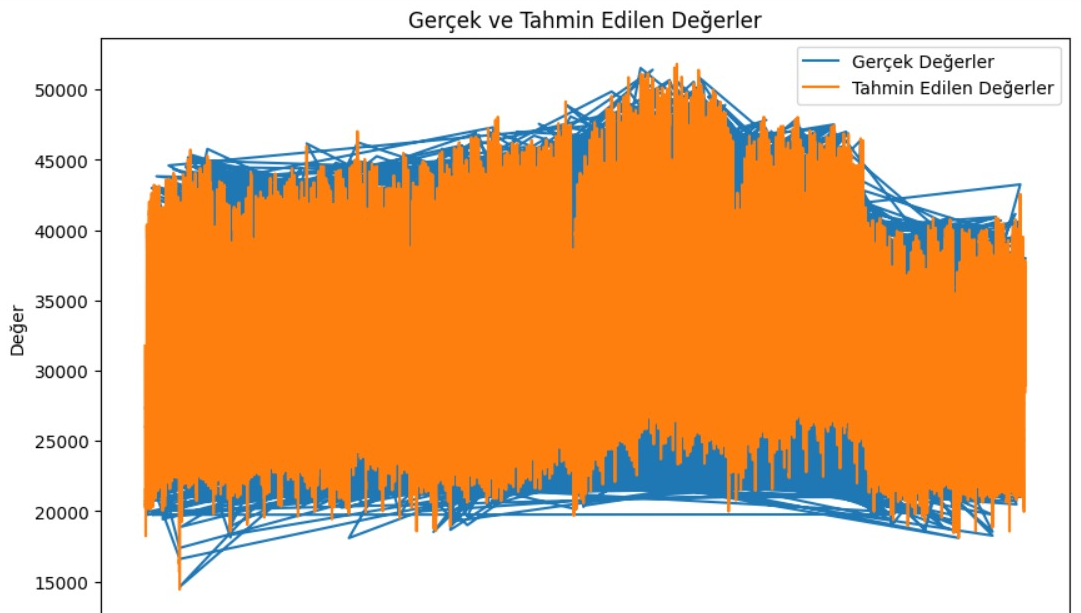
* The Mean Squared Error and Mean Absolute Error values of the models used are indicated in the visuals:



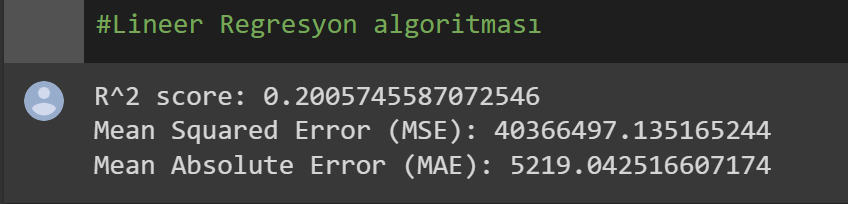


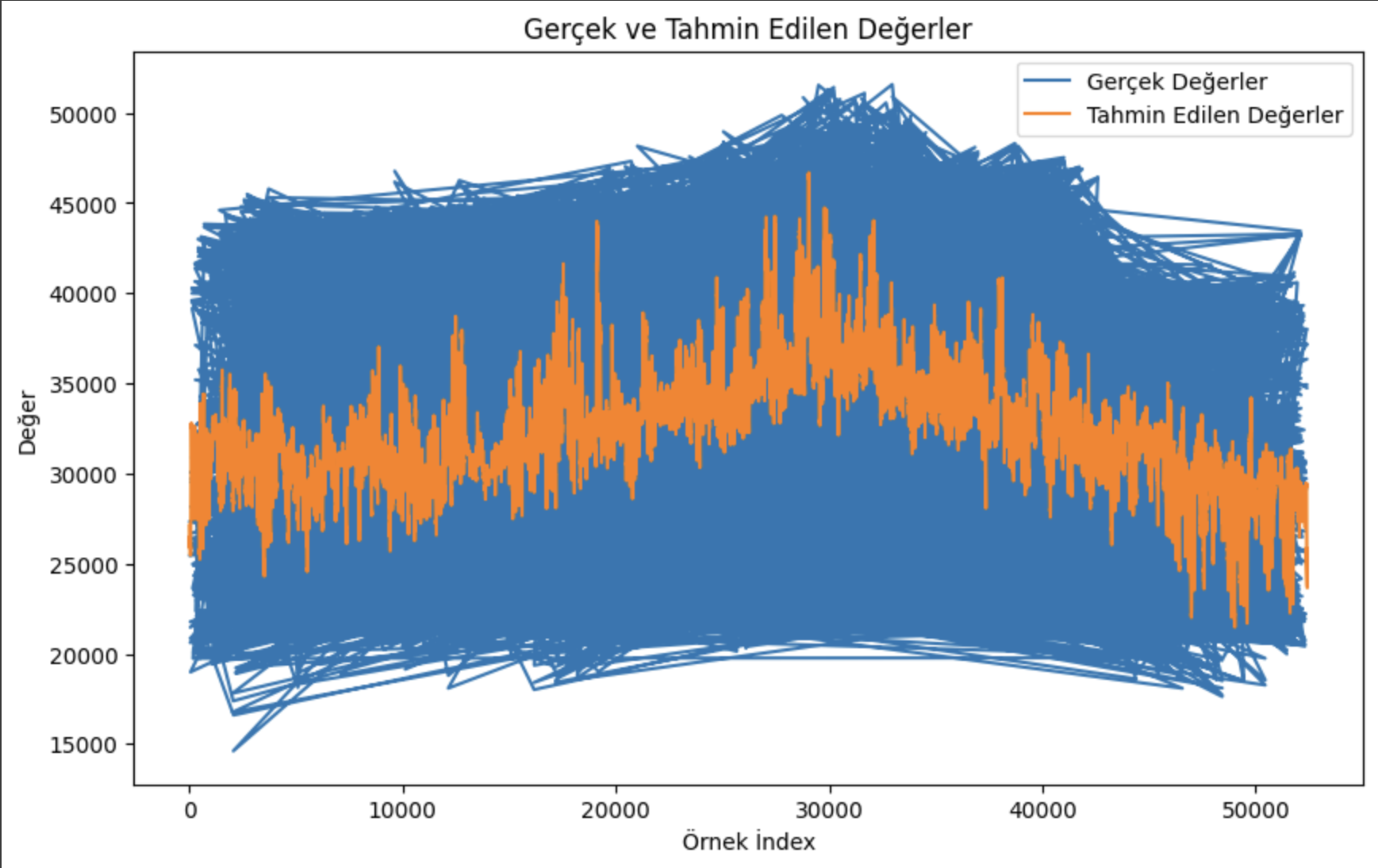
**The result of the 4% improvement achieved by the studies in the Random Forest Algorithm is given below.**

****

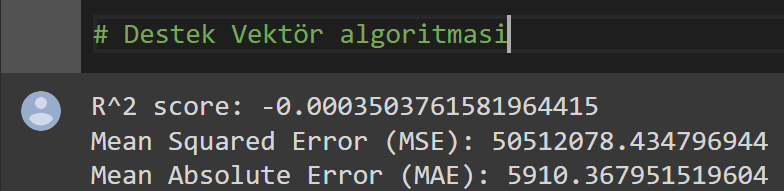
****

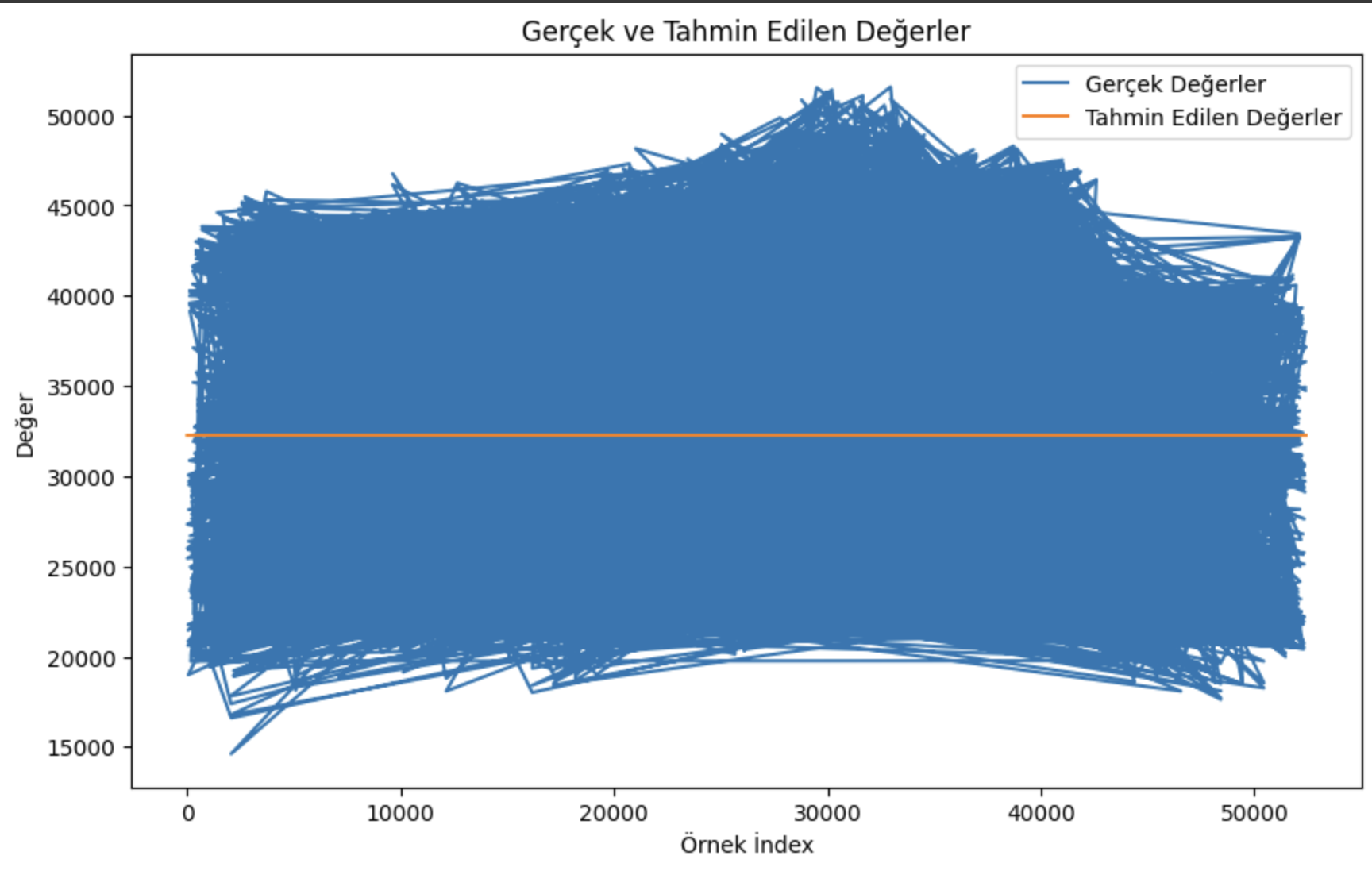
**Linear Regression:** Linear regression is used to model a linear relationship of a dependent variable (energy consumption) to one or more independent variables (e.g., air temperature, wind, time zone). Linear regression analyzes data to find the line (or hyperplane) that fits best. This model can be used to estimate the energy consumption we want to estimate.



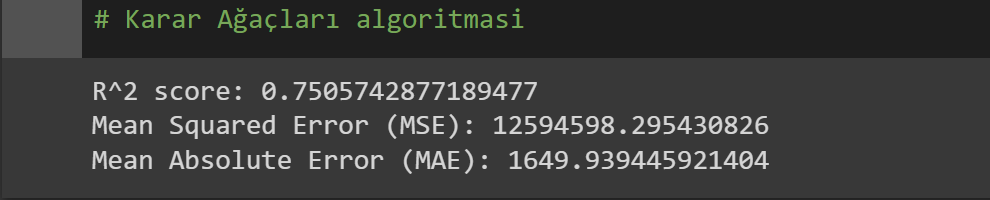


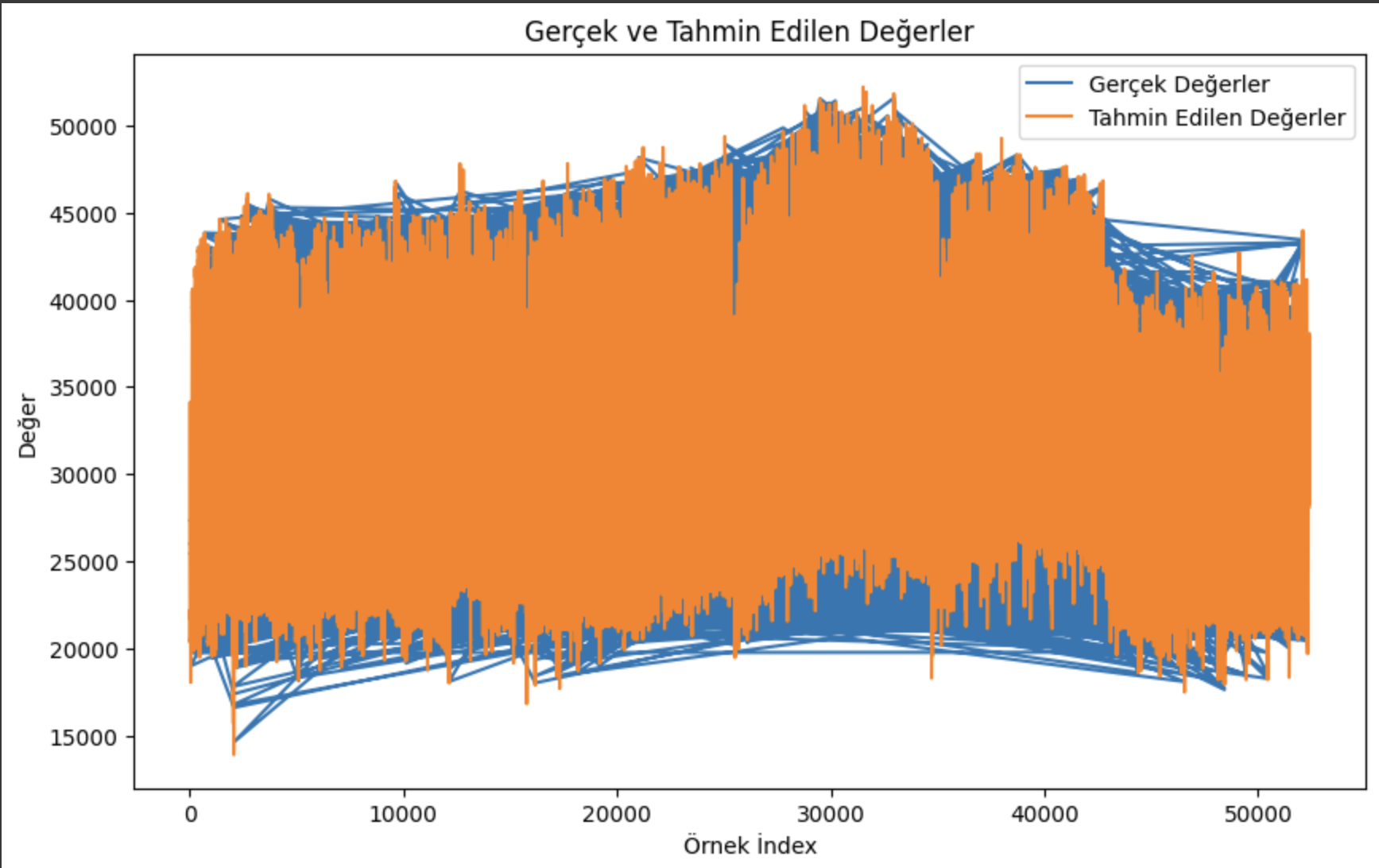
**Support Vector Machine (SVM):** The support vector machine is a learning algorithm used for classification and regression problems. SVM aims to optimally separate a dataset in such a way that labeled samples are hyperplaned between the two classes. To estimate energy consumption, SVM analyzes data by modeling the relationship between energy consumption and other independent variables.



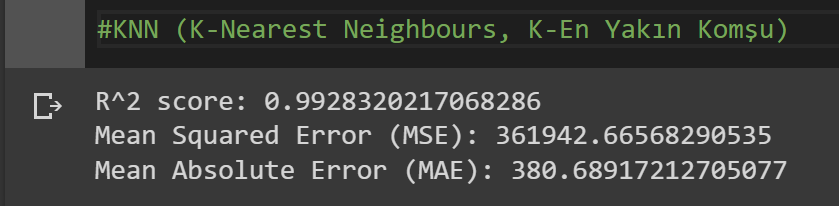


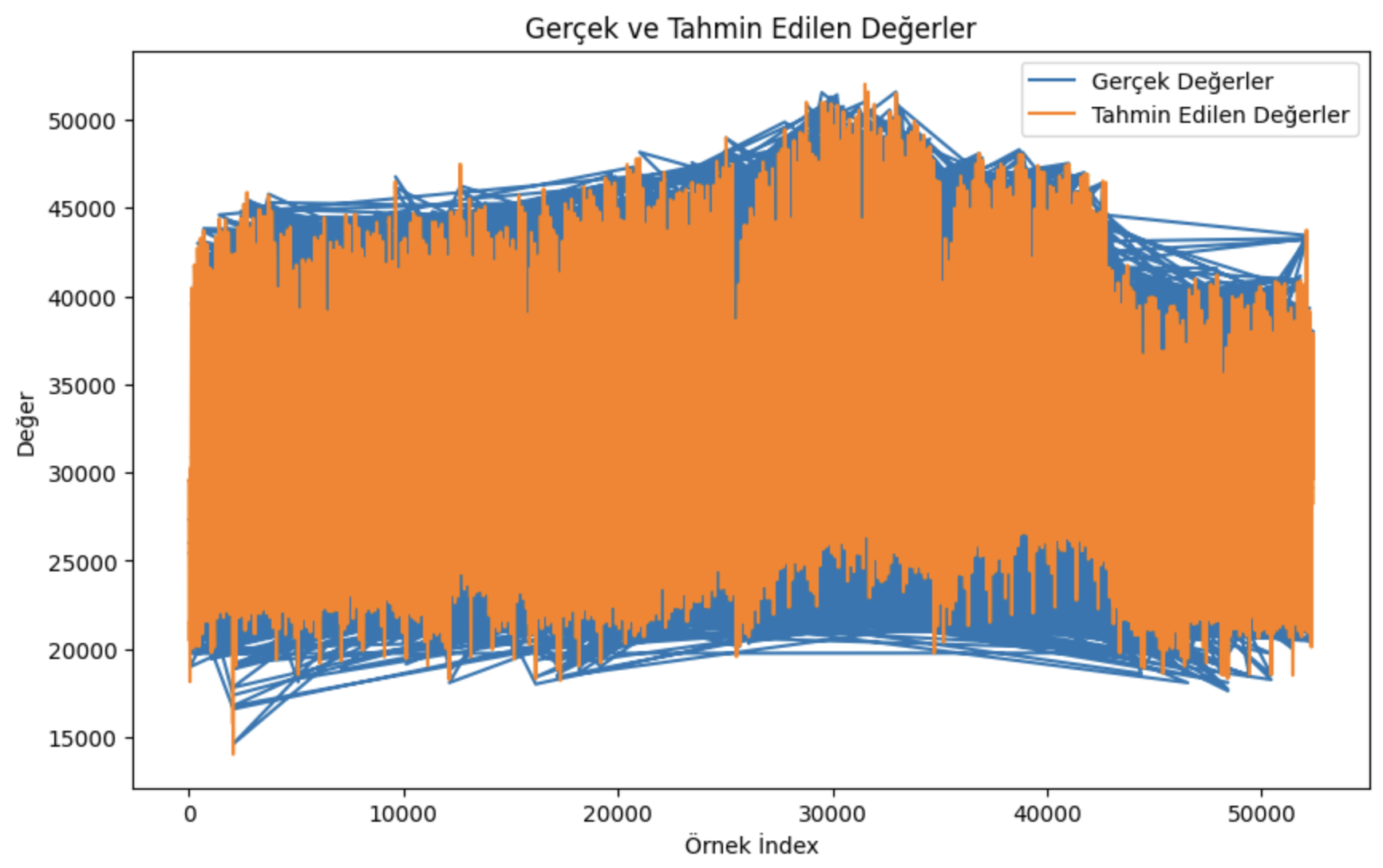
**Decision Trees:** Decision trees are a method used to classify data or perform regression analysis using a set of decision rules. In each decision node, a branch is selected based on the value of an argument, and as a result, an estimate of energy consumption is obtained. Decision trees can identify important patterns in the data and make predictions of energy consumption based on these patterns.





**K-Nearest Neighbor:** K-Nearest Neighbor (k-NN) is an algorithm used to perform classification or regression based on the majority of neighboring points around a new data point. To estimate energy consumption, we look at the energy consumption values between the neighbors of a new data point and make the prediction based on the values of these neighbors. The K-nearest neighbor algorithm can make estimates of energy consumption based on similarities in the data.





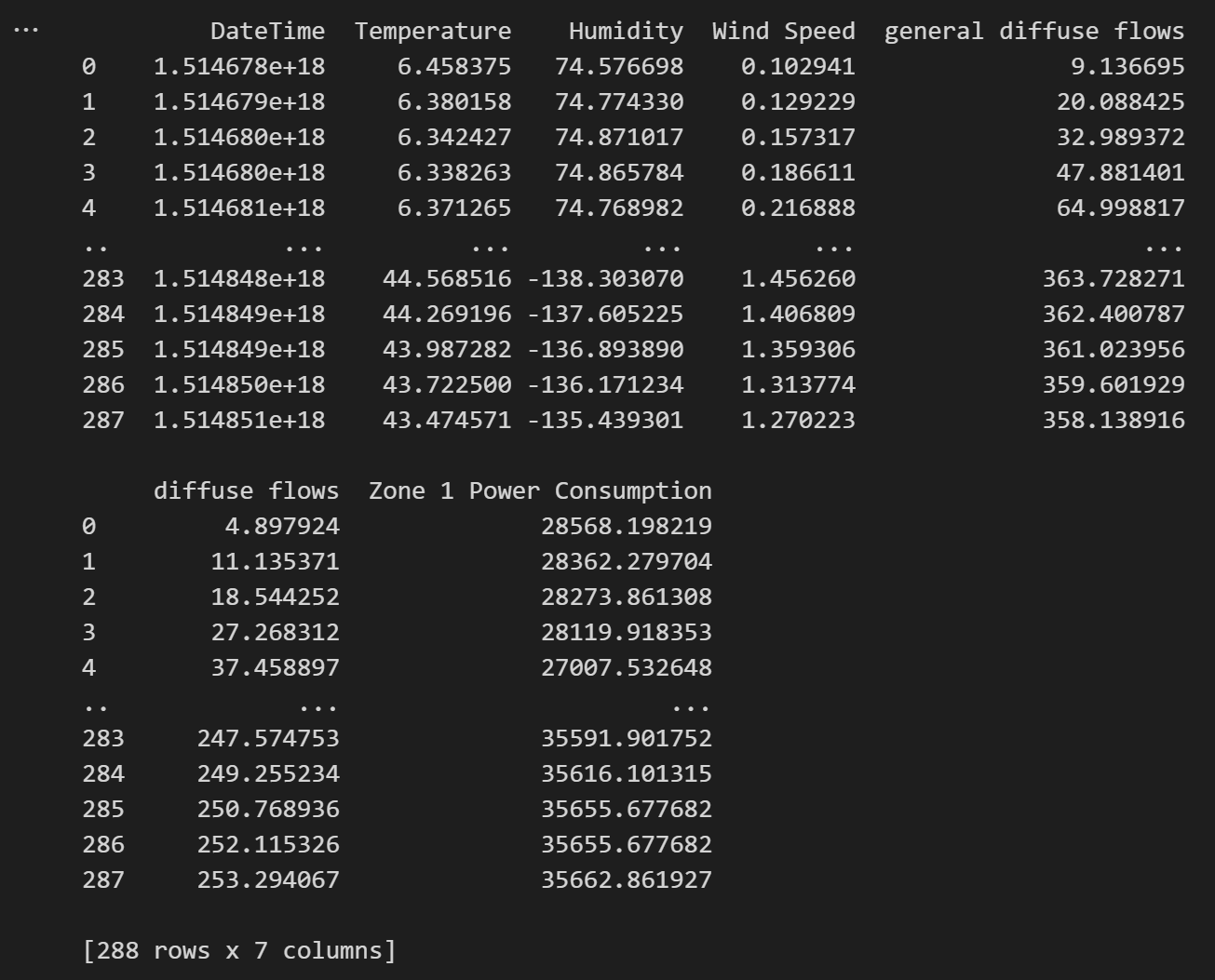
# 3. Bulgular (results)

**R^2 Scores, Mean Square Error and Mean Absolute Error values of five different models used in the study were obtained. When these values were compared, it was observed that the K-Nearest Neighbor and Random Forest models gave better results. The best score was obtained in the Random Forest model. The Random Forest model was determined to be the best predictive model in this study.**

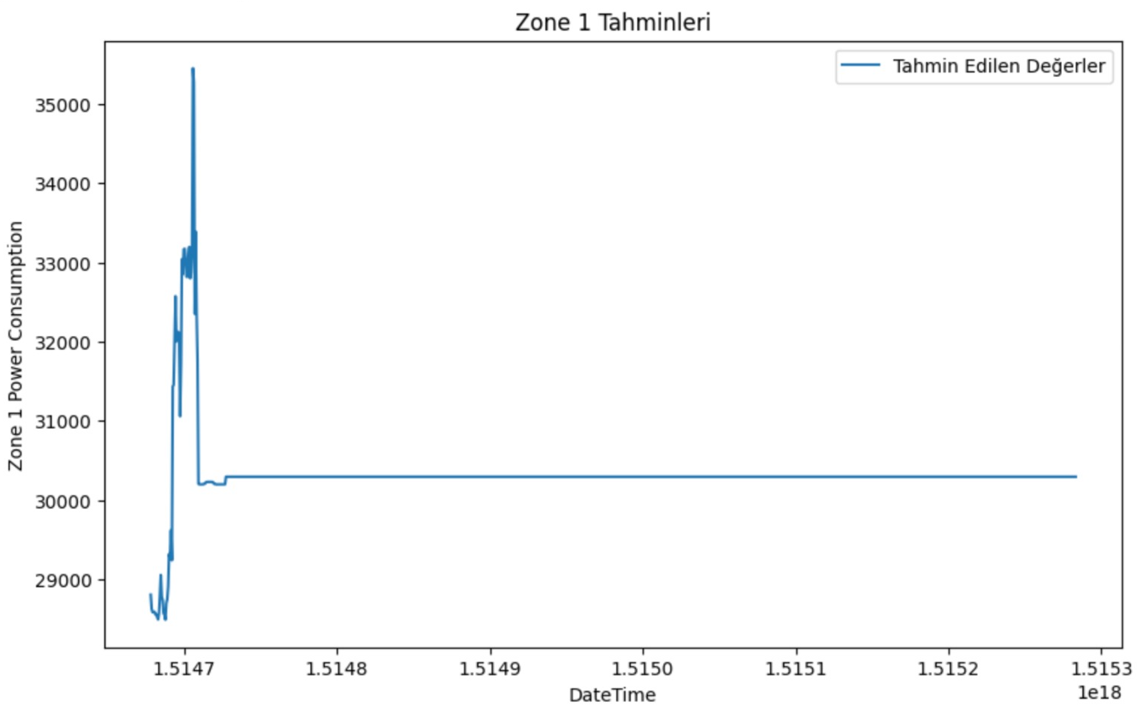
**The results show that the Random Forest model has the highest R^2 Score and the lowest Mean Square Error and Mean Absolute Error values. This model has been identified as the most effective method for estimating the power consumption of Tetouan City.**

.

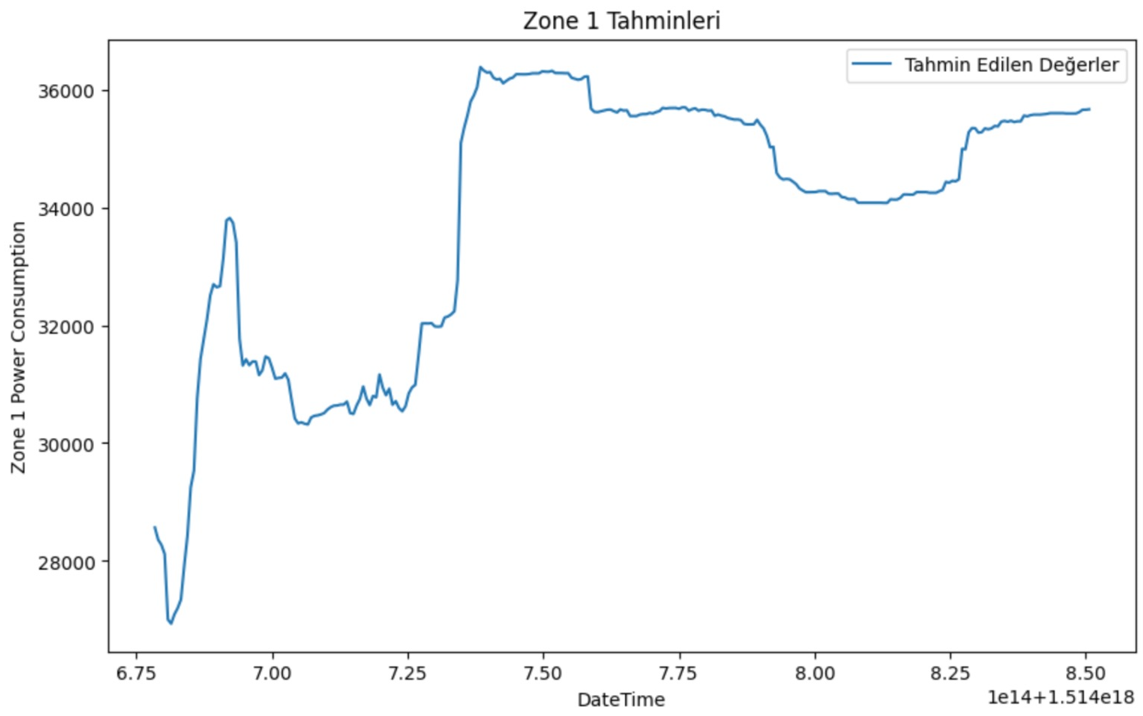
**Figure 1.** Some values of the dataset and predicted values for Zone-1.



**Figure 2.** One-week energy consumption forecast values for Zone-1 are shown in Figure 2.



**Figure 3.** 2-day energy consumption forecast values for Zone 1.



The following table shows the R^2 Score, Mean Square Error,

and Mean Absolute Error values for the five models used:

**Table 3**. R^2 Score, Mean Square Error, Mean Absolute Error values for the five models used.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lineer Regression | Decision Trees | K-Nearest Neighbor | Support Vector | Random Forest |
| 0.22780164454656449 | 0.7514442513611623 | 0.993430868345864 | -0.0003503761581964415 | 0.9150239635644875 |
| 38991682.142094575 | 12550670.01512195 | 331704.27209370903 | 50512078.434796944 | 4243904.892595624 |
| 5130.741562235543 | 1644.475274965662 | 370.4063455034337 | 5910.367951519604 | 1274.1475632488546 |

# 4. Discussion and Conclusion

Power consumption estimation is an important issue for energy management and improving energy use. In these studies, power consumption was estimated using different machine learning models and the results were evaluated.

A study of three different power distribution networks in the city of Tétouan used historical power consumption data from 2017. Different machine learning models such as support vector machine (SVM), artificial neural network (ANN), and random forest (RF) models are compared. Emphasizing the importance of power consumption during the energy crisis, it is concluded that the RF model makes more reliable and accurate predictions than other models. In addition, it is noted that the RF model completes the training and testing phases in a shorter time. [2]

In another study, linear regression, decision tree, feed forward neural network, support vector machine, and random forest models were compared. The results showed that the random forest model achieved lower prediction errors than other models. Therefore, the random forest model was preferred. [4]

Both studies show that the random forest model is an effective option for power consumption estimation. The random forest model provides lower prediction errors and higher accuracy rates compared to other models. Therefore, it may be recommended to prefer the random forest model in estimating power consumption.

In this study, which also took into account the studies in the literature, it was determined that the random forest model was the model with the lowest error rate in predicting future energy consumption. Comparisons with the results in the literature also show that the Random Forest model can obtain approximately 4% better prediction results than similar studies in this study.

In future studies, it is planned to further develop the random forest model and test it with larger data sets.

**Ethical Considerations**

**Machine learning technologies are evolving rapidly, and it is extremely important to use these technologies in an ethically correct manner. When using a machine learning algorithm such as the Random Forest model, the ethical considerations we approach with high precision are:**

**1. Data Privacy and Privacy:**

* The Random Forest model makes predictions by analyzing training data. Therefore, the dataset used should not contain personal information or sensitive data.
* If your data set contains personal information, it is important to take appropriate precautions regarding confidentiality and confidentiality. You can use methods such as data anonymization and encryption.

**2. Prejudice and Fairness:**

* The Random Forest model is formed by the combination of multiple decision trees. Biases in the training data can also be reflected in the results of the model.
* The diversity and representativeness of the training data can reduce the risk of the model producing biased results. It is important to make sure that the dataset is balanced and contains samples belonging to various groups.

**Transparency and Explainability:**

* The Random Forest model has a complex structure and makes predictions as a result of the combination of many decision trees it contains. Therefore, it can be difficult to explain and make the results of the model understandable.
* You can use explainability techniques to determine the model's decision process and important features. Visualizing feature weights or the outputs of decision trees can help with this.

**Unfair Competition:**

* When the Random Forest model is used in businesses and competitive environments, unfair competition and protection of trade secrets become important.
* You are required to comply with privacy policies and legal regulations regarding the training of the model and the use of its results.

**Responsibility and Human Control:**

* Although machine learning models produce results based on human decisions, it's important to leave the final decision entirely to humans.

**Compliance with ethical guidelines**

This study did not involve any procedures involving human volunteers or experimental animals. Our research was conducted in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Declaration of Helsinki and its subsequent amendments. Since there is no need for ethical permission, "Ethical Permission" document numbers are not submitted.

**Funding**

Authors should disclose whether their research has received a special grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Conflict of interest**

The author should report whether there is any conflict of interest.

**Acknowledgment**

In this section, the person or people who helped in the study should be thanked. At the same time, the institutions and organizations that received financial support should be thanked in this section for their help and support. If the article is produced from a project, project support should be specified. The title of the section should be written in "Times New Roman" font, 10 font size and bold. If you want to present your work in English, change the title to "**Acknowledgment**".

# BIblIography (REFERENCES)

**[1]** International Trade Administration | Trade.gov. (2019). Morocco — Energy. Retrieved September 25, 2022, from

https://www.trade.gov/country-commercial-guides/morocco-energy#:~:text=Per%20the%20state%2Downed%20power,%2C%20others%20(2%20percent).

**[2]**  Zogaan, W. A. (2022). Power Consumption prediction using Random Forest model. International Journal of Mechanical Engineering, 7(Special Issue 5), 329-341. ISSN: 0974-5823

**[3]** Beheshti, N. (2022, March 2). Random Forest Regression. Towards Data Science.

**[4]** Salam, A., & El Hibaoui, A. (2019). Comparison of Machine Learning Algorithms for the Power Consumption Prediction - Case Study of Tetouan city. In International Renewable and Sustainable Energy Conference (IRSEC) Proceedings, Tetouan, Morocco.

**[5]** Ouahabi, M. H., Elkhachine, H., Benabdelouahab, F., & Khamlichi, A. (2020). Comparative study of five different methods of adjustment by the Weibull model to determine the most accurate method of analyzing annual variations of wind energy in Tetouan - Morocco. Procedia Manufacturing, 46, 698-707.