World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:18, No:6, 2024

Reducing the Imbalance Penalty through Artificial Intelligence Methods Geothermal Production Forecasting: A Case Study for Turkey

H. Anıl, G. Kar

***Abstract*—**In addition to being rich in renewable energy resources, Turkey is one of the countries that promise potential in geothermal energy production with its high installed power, cheapness, and sustainability. Increasing imbalance penalties become an economic burden for organizations, since the geothermal generation plants cannot maintain the balance of supply and demand due to the inadequacy of the production forecasts given in the day-ahead market. A better production forecast reduces the imbalance penalties of market participants and provides a better imbalance in the day ahead market. In this study, using machine learning, deep learning and time series methods, the total generation of the power plants belonging to Zorlu Doğal Electricity Generation, which has a high installed capacity in terms of geothermal, was predicted for the first one-week and first two-weeks of March, then the imbalance penalties were calculated with these estimates and compared with the real values. These modeling operations were carried out on two datasets, the basic dataset and the dataset created by extracting new features from this dataset with the feature engineering method. According to the results, Support Vector Regression from traditional machine learning models outperformed other models and exhibited the best performance. In addition, the estimation results in the feature engineering dataset showed lower error rates than the basic dataset. It has been concluded that the estimated imbalance penalty calculated for the selected organization is lower than the actual imbalance penalty, optimum and profitable accounts.

***Keywords*—**Machine learning, deep learning, time series models, feature engineering, geothermal energy production forecasting.

I. Introduction

ARTIFICIAL intelligence is systems or machines that imitate human intelligence and can develop themselves with the information they collect, discover, make sense of, and learn from past information. Machine learning, which develops algorithms that learn from data for these operations, comes first in the field of artificial intelligence. As a subset of this, deep learning is a field of study that covers artificial neural networks and similar machine learning algorithms with one or more hidden layers. Artificial intelligence, which has become a transformative power in every field and sector today and continues to develop, is used for many purposes such as energy storage in the energy sector, production or consumption forecasting, energy cost, balancing energy supply and demand, and monitoring the potential of renewable energies. Energy use is important as a result of the increasing population, industrialization, and urbanization in Turkey. The importance of using renewable energy sources is increasing daily because most of the energy needs are met by fossil fuels. Natural gas and oil are two examples of fossil fuels that are expensive, finite, and bad for the environment. Thanks to renewable energy sources that can be used in many areas such as heating, electricity production, transportation, and industrial activities, the foreign dependency on the country's economy decreases, and the country has the opportunity to develop with its efforts. This situation is an important development opportunity, especially for Turkey, which has rich opportunities to maintain natural resources.

Geothermal energy is a domestic underground resource that is renewable, clean, inexpensive, and environmentally friendly. Due to its geological and geographical location, Turkey has a high geothermal potential spread all over the country with different temperature ranges, as it is located on an active tectonic belt [1]. Located in Aydın Denizli province, Büyük Menderes Graben has geothermal electricity installed capacity of approximately 1000 MWe, which constitutes the majority of Turkey's total geothermal electricity production [1]. Geothermal energy is used in district heating (city, residences), greenhouse heating, thermal and health facilities heating, and thermal water heating. According to the EMRA's December 2021 Sector Report, while the Geothermal Licensed Electricity Installed Power was 1,613.19 MW as of the end of December 2020, it increased to 1,676.17 MW in 2021. While its total production was 10,027,696.61 MWh in 2020, it increased to 10,770,879.81 MWh at the end of 2021. In summary installed power increased by 3.90% in 2021 [2]. It is important to estimate the geothermal energy production potential, which increases every year, to encourage the use of domestic energy resources and to investigate the effects of the estimated production on the energy market.

EXIST (Energy Exchange Istanbul), which is responsible for Turkey's energy market management, runs the Day-Ahead and Intraday Markets, settles transactions, and sends receivable-debt notifications to market participants [3]. DAM (Day-Ahead Market) is an organized market for electricity trading and balancing activities that is conducted by the market operator one day before the delivery day of electricity. DAM's goal is to assist market players in balancing their production or consumption demands with their day-ahead contractual

MSc. Hayriye Anıl is with Bahcesehir University, Department of Computer Engineering, 34353, Istanbul, Turkey (e-mail: hayriye.anil@bahcesehir.edu.tr).

Asst. Prof. Dr. Gorkem KAR, Bahcesehir University, Department of Computer Engineering, 34353, Istanbul, Turkey (e-mail: gorkem.kar@eng.bau.edu.tr).

obligations. Hourly, block, and flexible bids are the three types of bids that can be given. According to price levels, the supply side can alter how much it produces, and the demand side can adjust how much it consumes. All licensed legal entities that are market participants can participate in the DAM. DAM transactions are carried out on a daily, hourly basis. Each day consists of hourly periods that start at 00:00 and end at 00:00 the next day. Every day until 12:30, market participants participating in the DAM notify the Market Operator of their day-ahead market offers for the next day through the DAM system. Between 13:00 and 13:30, the optimization tool evaluates confirmed offers, and MCP (Market Clearing Price) and amounts are established by intersecting supply and demand for each hour of the day. Every day at 13:30, the relevant market participant is notified of commercial transaction approvals containing approved purchase-sale amounts. If the text of these messages contains an error, the market participant can object between 13:30 and 13:50. Between 13:50 and 14:00, objections are considered, and the outcome is communicated to the participant. Prices and matches for the next 24 hours are ultimately revealed at 14:00 [4].

Although a market with balanced production and consumption quantities is presented to the System Operator with the DAM, there are deviations in real time. If the promised sales and purchase quantities do not match the production or consumption quantities, an imbalance occurs. SMP (System Marginal Price) is defined as the bid price corresponding to the net order volume calculated by considering all bids in the balancing power market for load taking or load shedding, depending on the direction of the energy shortfall. The difference between the estimated purchase and the actual production or consumption creates a cost factor called imbalance penalty. If this equality is not achieved, an imbalance occurs due to these differences and the participant pays the cost of this imbalance. If the difference between the total production of the power plants belonging to the participant and the bid value (matched sales amount) estimated by the participant with mathematical programming is greater than zero, the positive imbalance formula is calculated, and if it is less than zero, the negative imbalance formula is calculated. Imbalance cost at h hour (I_h), calculated according to the positive (IP_h) and negative (IN_h) conditions is as follows:

$$IP_h = (A_h - D_h) * \min(MCP_h, SMP_h) * 0.97, if (A_h - D_h) > 0 \quad (1)$$

$$IN_h = (A_h - D_h) * \max(MCP_h, SMP_h) * 1.03, if\ (A_h - D_h) < 0 \quad (2)$$

D_h is the bid value given by the manufacturer as an estimate in h hour, A_h is the actual value produced by the plants in h hour, MCP_h, SMP_h are the MCP and SMP in h hour, respectively.

The price that the participant must earn in h hour:

$$A_h * MCP_h \quad (3)$$

Status at h hour when imbalance occurs:

$$(D_h * \ MCP_h) + I_h \quad (4)$$

Difference between the real profit and the imbalanced account is the price at which it loses.

A good production forecast reduces the imbalance penalty and provides a better imbalance with a better forecast in the DAM in EXIST. With the improvement of this estimation, it can be encouraged to increase the number, installed power, and capacity of generation plants, especially geothermal, in renewable energy sources. In this study, Zorlu Doğal Electricity Generation, which is among the top 5 companies with the highest installed geothermal power according to the EMRA license inquiry result, was chosen as a case study. Traditional machine learning, deep learning and time series models were used in the study. It is aimed to estimate the total production in the first one and first two weeks of March. With these estimated production results, the imbalance cost of the first 1 and 2 weeks of March was calculated and compared with the real values and its effect on the energy market was investigated.

The dataset was prepared by taking the production data of the power plants of the organization from the EXIST Transparency platform, and the temperature data from the StormGlass Weather API service.

The remaining parts of the study proceed as follows: In the second part, the relevant work is examined and briefly summarized. In the third chapter, the methodology of the study is mentioned, and general information about the preparation, collection, and analysis of the dataset and the methods used are given. In the fourth chapter, the experimental results are shared, and finally in the fifth and sixth chapter, the results are discussed and the article is summarized.

II. Related Work

A study by Dinler (2021), used a method consisting of a mixture of binary classification, LSTM autoencoder, and advanced classifiers to reduce the cost of imbalance caused by the wrong offers given to the market due to insufficient day-ahead forecasts of wind generators and the variable electricity price of the balancing market determined in the market. He made a predict the day-ahead or imbalance price would be higher at a certain time of the next day [5]. Unlike this study, wind energy was used in this article. Wind power producers are increasingly participating in the DAM, according to the report, as the potential of wind power grows at a global rate. Wind producers are the most vulnerable to imbalance costs compared to other renewable generators because of their variable generation and poor day-ahead forecasts [5]. In addition, the other factor affecting the imbalance cost is the constantly changing balancing electricity price determined in the market [5]. The present study here and the important difference is that it is possible to reduce the imbalance cost by using historical market data. As the dataset, market values from EXIST for the years 2017-2018 were used and tested for the production data of four wind power plants. The correlation between these market values was examined and it was observed that there was a strong relationship between MP (Marginal Market Price) and DP (Day-Ahead Price). The main purpose of the proposed method is to avoid excessive imbalances that occur when the difference between marginal and day-ahead prices is high and forecast accuracy is low. Initially, the binary classification method was applied, and it was determined as -1 if MP is greater than DP and 1 if MP is less than or equal to DP [5]. It then preprocesses the data using the LSTM autoencoder and combines the five binary classifiers to form a hybrid classifier [5]. The results show that the LSTM autoencoder improves the accuracy of all classifiers and the hybrid classifier gives the best accuracy of 61.08% [5]. Therefore, the method extracts information about whether the day-ahead or balancing market price will be higher at a certain time of the next day. Then, using this information, auxiliary algorithms modify existing production estimates and avoid spikes in imbalance cost [5].

III. Methodology

A. Research Design

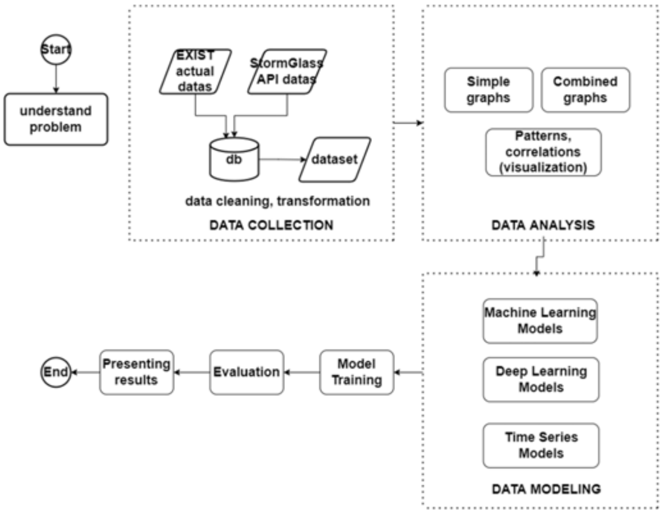

Fig. 1 System Overview

B. Data Collection

The generation data of the power plants belonging to Zorlu Doğal Electricity Generation have been obtained from the EXIST Web Service. Production data of the power plants were taken between January 2017 and March 2022. Information of the power plants belonging to the organization is given in Table I.

Temperature and humidity data of these power plants according to their latitude and longitude positions were obtained from StormGlass API Service. All power plants belonging to the organization are located at Denizli, Sarayköy.

TABLE I
PLANT INFORMATION

Plant ID	Plant Name
1	Kızıldere-3 JES
2	Kızıldere II JES
3	Kızıldere JES

1. Data Transformation

Since the installation dates of the power plants belonging to the organization are different from each other, the start date of the production data is also different. For this reason, the generation data of the power plants have been taken since January 2017 while collecting the data. Then, the date of installation of the power plants belonging to the organization with the latest date was selected, and the total generation of the power plants was taken. The establishment dates of the power plants are shown in Table II, the maximum start date available in the API service by organization is 2017-08-22.

TABLE II
ESTABLISHMENT DATES OF POWER PLANTS

Plant Start Date	Plant Name
27.04.2016	Kızıldere-3 JES
02.05.2013	Kızıldere II JES
21.08.2008	Kızıldere JES

2. Data Analysis

EDA (Exploratory Data Analysis) is the process of applying statistical measurements to examine an existing data collection in order to find patterns, detect anomalies, test hypotheses, and verify assumptions. The major responsibilities in this process are to summarize the data, discover hidden correlations and relationships between the data, create predictive models, evaluate the models, and calculate the accuracies. In this article, missing data and duplicate data were searched, the dataset was statistically evaluated, the correlation between the variables was observed, and then monthly and annual graphs were drawn. No missing or duplicate data were found in the dataset. The monthly and annual charts of organization and the correlation heatmap chart are Figs. 2 and 3, respectively.

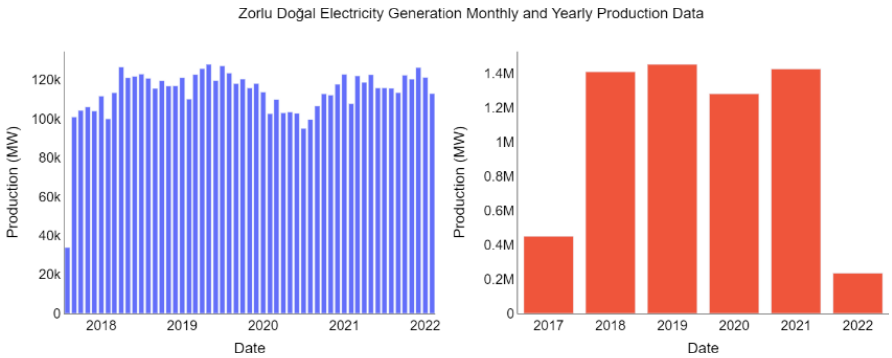

Fig. 2 Zorlu Doğal Electricity Generation Monthly and Yearly Production Data Bar Charts

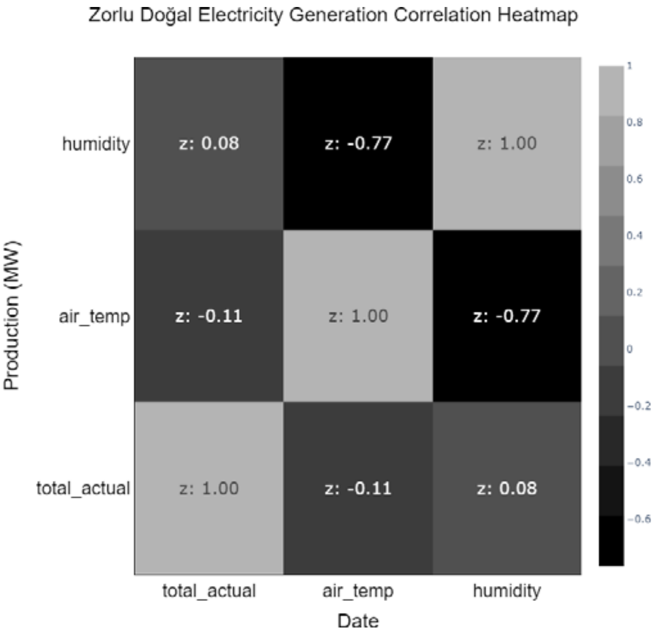

Fig. 3 Zorlu Doğal Electricity Generation Correlation Heatmap

As seen in Fig. 2, while geothermal production is low in summer, it is high in winter. As can be seen in Fig. 3, there is an inverse relationship between temperature and humidity and geothermal production. As a result of the data analysis, it was observed that as temperature and humidity increase, geothermal production decreases, and as temperature and humidity decrease, geothermal production increases. At the same time, looking at the monthly and annual charts of each organization, it has been seen that geothermal production has increased and its potential has increased over the years.

C. Methods

1) Train and Validation Set

After adjusting the dataset of the organization, forecasts for the first one and first two weeks of March were made for each organization. The date range to be estimated was chosen as the validation set, and the date range from the start of the organization to the beginning of the validation set was chosen as the training set. Train and validation set groups are shown in Table III.

TABLE III
TRAIN AND VALIDATION SET GROUPS

Train Set	Validation Set
Organization Start Date – February 2022	2022/03/01 – 2022/03/15
Organization Start Date – February 2022	2022/03/01 – 2022/03/08

2) Feature Engineering

Feature engineering is the practice of changing data into features that better describe the underlying problem in order to improve machine learning performance. Feature engineering's purpose is to get data from which learning algorithms may discover patterns and apply them to improve outcomes. Understanding features, improving features, selecting features, and constructing features are the four fundamental parts of feature engineering [6]. Feature understanding means knowing how to classify material based on its attributes and quantitative status. Feature improvement is the cleaning and assigning of missing data values to maximize the value of the dataset. Feature selection is a statistical selection and subset of feature set to reduce noise in data. Feature construction is creating new features to take advantage of feature interactions [6]. In this article, simply two datasets are used: The main dataset with date order (index), temperature, humidity, and total production variables, and feature engineered dataset with year, month, day and hour data extracted from date, temperature, humidity and total production. Without any feature engineering procedure for the main dataset, the performances of machine learning and deep learning models were obtained. Then, the performances of machine learning and deep learning models were applied for the feature engineered dataset obtained by applying a combination of feature engineering. Finally, the performance results in the two datasets are compared. In this article, feature engineering has been tested on all preferred machine learning and deep learning models and models such as Prophet and SARIMAX that accept multiple variables in time series models.

3) Machine Learning Models

The study of computer algorithms that evolve automatically as a result of experience and data is known as machine learning. According to the kind of issue, machine learning algorithms are separated into a variety of strategies, such as regression and classification. Regression is a supervised learning strategy in machine learning in which one or more regressions are used to try to predict target variables. By fitting a line to the observed data, regression models are used to characterize relationships between variables. The ability to forecast how the dependent variable will change when the independent factors change is provided by regression.

a) Random forest: RF is an ensemble learning approach that combines numerous weak models to find answers to complicated problems. A Random Forest, as the name implies, is made up of multiple decision trees. It takes the estimates from each tree and calculates the final output based on the majority votes of the estimates, rather than being linked to a tree. RF are commonly employed in regression and classification applications, and they frequently yield excellent results even when hyperparameter adjustment is skipped.
b) Support vector regression: SVR is used for both regression and classification. Its primary principle is to cover as much data as possible in the shortest period. It is an appropriate model for SVM regression since it generates lines that take the greatest data yet have the smallest range, with some outliers left out.

4) Time Series Models

A time series is a collection of data points that are arranged chronologically. Data are captured every hour, minute, month, or quarter, and is regularly spaced across time. The closing price of a stock, a home's power use, or the temperature outside are all instances of time series. Traditional regression approaches differ from time series estimation. This is because time series have an order that cannot be modified when modeling. Future values are stated as a function of previous values in time series estimation. As a result, data should be organized so as not to damage this link. Also, time series can contain only one variable; it is not uncommon for a basic dataset with only one time column to be assigned a value at that point in time. Similarly, time series having a single value in each time step are known as univariate time series, while multivariate time series have numerous values. Trend, seasonality, and residuals are the three components of a time series. A trend is a pattern that has been observed over time and indicates the average rate of change. In the long term, a trend typically demonstrates the pattern of data to increase/rise or reduce/decrease. Seasonality is a type of periodic fluctuation in which the same pattern repeats itself at regular intervals. It is a common aspect of time series data in economics, weather, and stock markets; it is less common in scientific data. The term ‘‘residuals” refers to behavior that is not explained by the trend or seasonality components. They are random mistakes, commonly known as white noise. The method of identifying these components is called decomposition. A statistical technique called decomposition can be used to separate a time series into its component elements. Geothermal output and temperature data are time-dependent series in the datasets utilized in this study. The trend and seasonality patterns of the production data of the organization are shown in Figs. 4-6, respectively.

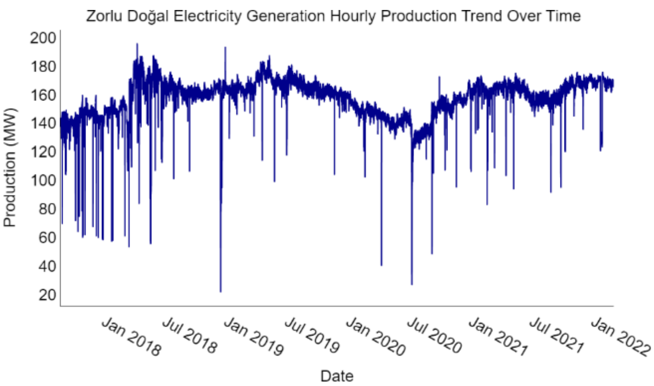

Fig. 4 Zorlu Doğal Electricity Generation Hourly Production Trend Over Time Line Chart

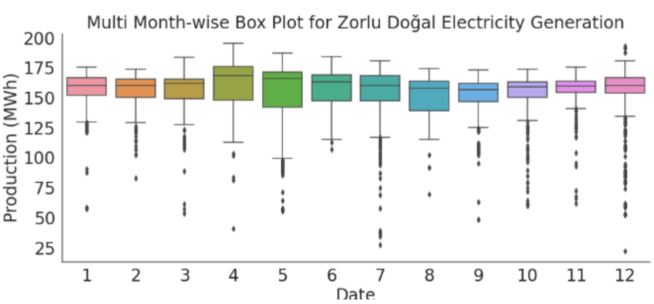

Fig. 5 Multi month-wise Box Plot for Zorlu Doğal Electricity Generation

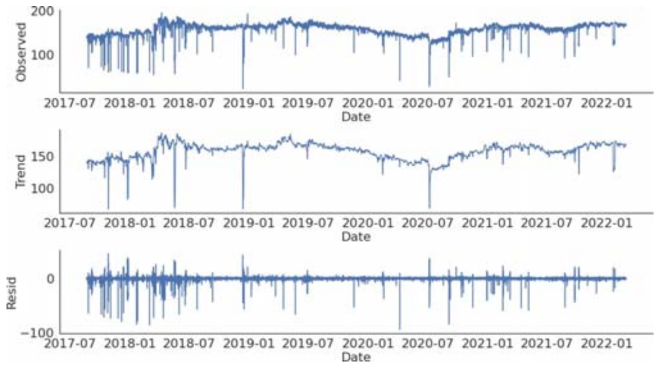

Fig. 6 Seasonal Decomposition of Zorlu Doğal Electricity Generation

TABLE IV
ZORLU DOĞAL ELECTRICITY GENERATION ADF TEST RESULT

Results of Dickey-Fuller Test	Zorlu Doğal Electricity Generation
Test Statistic	-9.861207e+00
p-value	4.228434e-17
No Lags Used	5.300000e+01
Number of Observations Used	3.959400e+04
Critical Value (1%)	-3.430515e+00
Critical Value (5%)	-2.861613e+00
Critical Value (10%)	-2.861613e+00
Conclusion	Data is stationary

The variances seen in Fig. 5 mean that there are production values lower than the lower boundary or greater than the upper boundary in that month, according to the normal distribution lower and upper boundary formula. There may be many reasons for these variances such as low production, production stoppage, power plant failure or maintenance [7]. The statistical parameters of a certain process do not vary over time, which is referred to as stationary data. When time series data are not stationary, it indicates that there are seasonality and trend patterns that must be eliminated. It is critical to transform data into a stationary format before training a time series forecasting model. As a result, whether the data are steady or not should be evaluated before modeling. There are several approaches for determining whether or not a series is stationary. In this article ADF (Augmented Dickey-Fuller) test was used. ADF is a statistical test that determines if a unit root exists in a series' domain and aids in determining whether the time series is stationary or not. The null hypothesis (H_0) in the ADF test is that the unit root exists in a time series. For the stationary form of the time series, the alternative hypothesis is utilized. The null hypothesis is rejected if the p-value is smaller than the critical value, indicating that the series is stationary. The null hypothesis cannot be rejected if the p value is bigger than the critical value, hence the series is not stationary. Non-stationary data must first be transformed to stationary series before time series forecasting methods may be used. Differencing is a technique for calculating the difference between two or more words in a series. Differentiation is commonly used to remove the variable mean. Mathematically, taking the difference can be written as:

$$y_{t0} = y_t - y_{(t-1)} \quad (5)$$

where, yt is the value at time t.

As shown in Table IV, ADF Test was applied to the datasets, each dataset was found to be stationary, and there was no need to apply an extra method to convert it to stationary.

a) Prophet: is Facebook's Core Data Science team's open-source software. It is a free and open-source framework for monitoring and forecasting time series. It relies on an additive model that aligns nonlinear patterns with daily, weekly, and annual seasonality, as well as additional holiday impacts. It is simple to use, and it is set up to find a good set of hyperparameters for the model to make clever predictions for trends and seasonally structured data by default.

b) SARIMAX: Seasonal ARIMA (SARIMA) is an ARIMA approach that can manage the seasonal component in univariate time series data. AR(P), I(D), and MA(Q) for the seasonality component of a time series are determined by combining three new hyperparameters. The letters p, d, and q stand for trend autoregressive order, trend difference order, and trend moving mean order, respectively. In a multiplicative model, the SARIMA model combines both non-seasonal and seasonal components. Equation (6) is a definition of the notation:

$$ARIMA(p, d, q)X(P, D, Q)_m \qquad (6)$$

SARIMAX (p, d, q) X(P, D, Q)m (X) is a SARIMA model with externally influencing variables, where X is the vector of exogenous variables. SARIMAX is a multivariate time series model.

5) Deep Learning Models

Deep learning is a field covered in the machine learning subcategory. Artificial neural networks and artificial neural network designs based on them make up the physical foundation of deep learning. The foundation of deep learning algorithms is the creation of a network of interconnected computer units. A tiny computational bundle known as an artificial neuron, though it is sometimes referred to simply as a neuron, is the fundamental unit of such networks. The nerve cells that make up our brain and central nervous system are the inspiration for artificial neurons. A single-layer perceptron, often known as an artificial nerve, is a mathematical model of the artificial nerve. It is designed to seem like an organic neuronal structure. It is only a part of the artificial neuron that detects and transmits signals to another nerve. In traditional machine learning, no matter how much the data are increased after a certain quantity, there will be no improvement in learning. In deep learning, the more data, the better the learning. Machine learning almost always requires structured data, while deep learning relies on levels of artificial neural networks.

a) LSTM: is a cell state that keeps a state for the duration of the training, allowing it to change from cell to cell and timestamp to timestamp while being better preserved. This suggests that RNN may have a higher influence on the total projection than earlier data in the window. The major difficulty with RNN is that it only remembers the previous state, which leads to the vanishing gradient problem. This challenge is handled with LSTM by establishing information tracking over several time series. Instead of having a single neural network layer, LSTM has a similar chaining structure. The key notion underlying LSTM is that these structures abstracted by gates may efficiently add and delete information from the inner cell state. A conventional neural network layer, such as a sigmoid, is used to create these gates. To process it for storage, LSTM performs this sort of operation, first forgetting unnecessary history, then keeping the most relevant new information, updating cell states, and finally providing an output.

b) Transformer: Vaswani et al. published Transformer, a state-of-the-art deep learning model that follows an encoder-decoder structure but does not rely on recurrent and convolutional to create an output [8]. Working process of the Transformer is as follows. A one-dimensional embedding vector is created for each word in the input string. Each embedding vector representing an input word is aggregated (per element) into a positional coding vector of the same length, giving the input positional information. The encoder block receives enhanced embedding vectors. All bidirectional words, whether they occur before or after that word, are dealt with by the encoder. In the time step, the decoder uses its expected output word as input. Positional encoding, as well as the encoder, enhances the input to the decoder. The decoder block is divided into three sublayers by the augmented decoder input. To prevent the decoder from connecting succeeding words, masking is used on the first bottom layer. The decoder also outputs the decoder at the second bottom layer, allowing it to connect all of the words in the decoder input string. To provide a forecast for the next word of the output sequence, the decoder output passes through a fully connected layer and then a softmax layer. In this article, in the Transformer implementation, besides the production output, air temperature and humidity information are added as historical covariates.

6) Evaluation Metrics

a) MAE (Mean Absolute Error): is calculated by averaging the absolute values of the difference between the predicted value and the actual value per line. The MAE formula is as follows:

$$MAE = \frac{1}{n} \sum | y_i - \acute{y}_i | \qquad (7)$$

b) MASE (Mean Absolute Scaled Error): is a metric for determining how accurate forecasts are. The output of the previous period is used to anticipate the next step in a naïve prediction. For naïve estimate across the whole time, the MASE is as follows:

$$MAE_{naive} = \frac{1}{N-1} \sum_{i=2}^{N} | y_i - y_{i-1} | \qquad (8)$$

The MASE is found by dividing the MAE by the naive estimate by the MAE ratio. A model with a lower MASE is better. The formula is as follows:

$$MASE = \frac{MAE}{MAE_{naive}} \quad (9)$$

c) MAPE (Mean Absolute Percent Error): For each prediction, MAPE is derived by dividing the error by the real value. This is done in order to obtain the mistakes as a percentage of the actual values. As a result, the error will be expressed as a percentage and will be standardized. A model with a lower MAPE is better. The MAPE formula looks like this:

$$MAPE = \frac{1}{n}\sum \mid \frac{y_i - \acute{y}_i}{y_i} \mid \quad (10)$$

d) MSE (Mean Squared Error): The difference between the estimated value per row and the actual value is squared, and then these errors are averaged. The smaller the error, the better the model. The MSE formula is as follows:

$$MSE = \frac{1}{n}\sum(y_i - \acute{y}_i)^2 \quad (11)$$

e) RMSE (Root Mean Squared Error): RMSE is the square root of MSE. The reason for taking the square root of the MSE is that the scale of the RMSE is the same as the scale of the actual values. A lower RMSE indicates a better pattern. The RMSE formula is as follows:

$$RMSE = \sqrt{MSE} \quad (12)$$

f) R2 (R squared): The R2 metric is very close to the 1 – MAPE metric. It is a performance metric rather than an error metric, which makes it great for communication. R2 is a value that tends to be between 0 and 1, with 0 being bad and 1 being excellent. The formula for R2 is as follows:

$$R^2 = 1 - \frac{\sum(y_i - \acute{y}_i)^2}{\sum(y_i - \acute{Y}_i)^2} \quad (13)$$

IV. Experimental Results

Two datasets were created, the main features and the features added with feature engineering. For these two datasets, the first 1-week prediction and the first 2-weeks prediction for March were made separately. The flow applied to each organization is shown in Fig. 7.

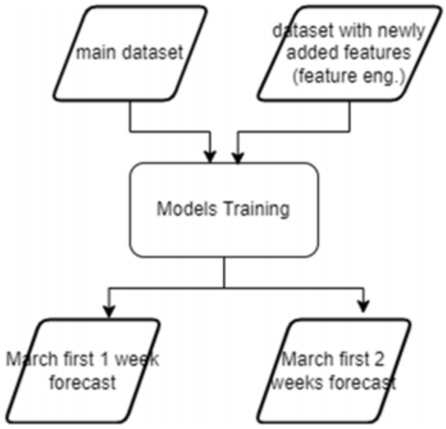

Fig. 7 Models forecasting system

In this section, the graph and model performance metric results of the models that give the best results in each of the machine learning, deep learning, and time series models are given in the tested models. The best model ordering was chosen considering the MAPE and MASE metric results. Finally, loss of the imbalance that occurred as a result of the real and forecasts for the first 1 and 2 weeks of March was calculated and the results were compared.

A. Forecast Results

1) *The first week of March (main dataset):* The results of the models that give the best 1-week forecast for March as a result of using the main dataset of the organization is as in Fig. 8 and Table V.

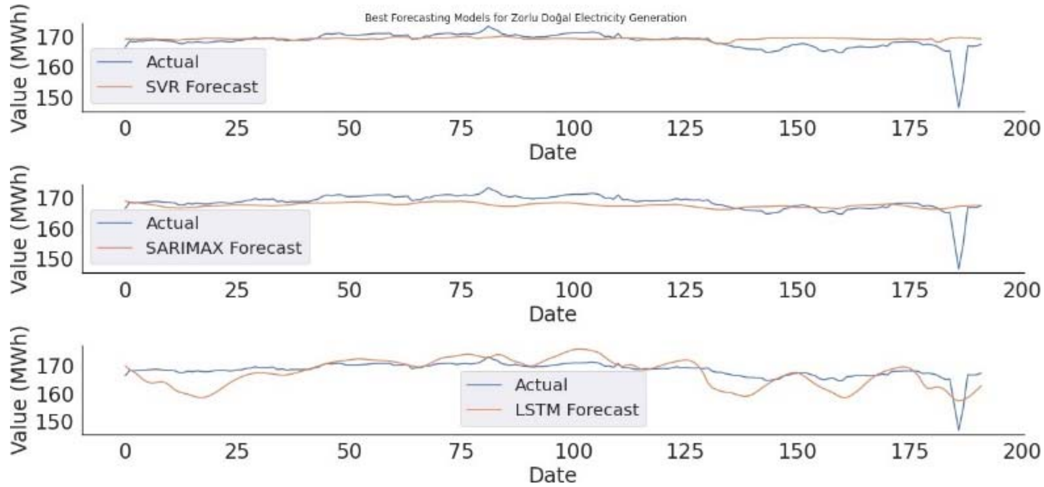

Fig. 8 First week of March comparison of best forecasting models and actual for Zorlu Doğal Electricity Generation (main dataset)

2) *The first two weeks of March (main dataset):* The results of the models that give the best 2-weeks forecast for March as a result of using the main dataset of the organization is as in Fig. 9 and Table VI.

TABLE VI
FIRST TWO WEEKS OF MARCH BEST MODELS PERFORMANCE METRIC RESULTS FOR ZORLU DOĞAL ELECTRICITY GENERATION (MAIN DATASET)

Models	MAE	MSE	RMSE	R^2	MAPE/MASE
SVR	1.83	7.40	2.72	-0.20	1.10
SARIMAX	1.97	6.54	2.55	-0.06	1.16
LSTM	4.61	33.96	5.82	4.60	4.60

3) *The first week of March (feature eng. dataset):* The results of the models that give the best 1-week forecast for March as a result of using the newly added features dataset with feature engineering of the organization is as in Fig. 10 and Table VII.

TABLE V
FIRST WEEK OF MARCH BEST MODELS PERFORMANCE METRIC RESULTS FOR ZORLU DOĞAL ELECTRICITY GENERATION (MAIN DATASET)

Models	MAE	MSE	RMSE	R^2	MAPE/MASE
SVR	1.62	8.10	2.84	-0.05	0.98
SARIMAX	1.87	7.35	2.71	0.04	1.12
LSTM	2.94	14.64	3.82	-0.90	2.94

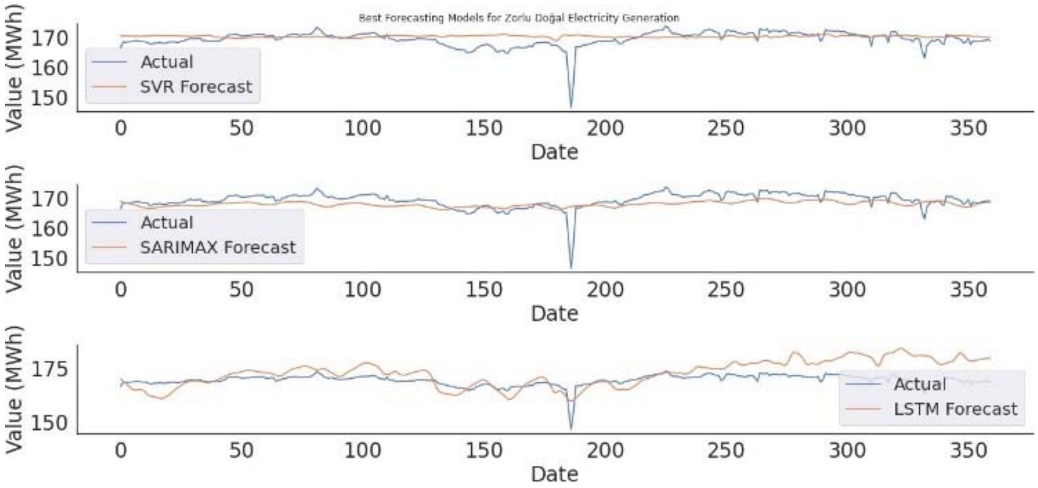

Fig. 9 First two weeks of March comparison of best forecasting models and actual for Zorlu Doğal Electricity Generation (main dataset)

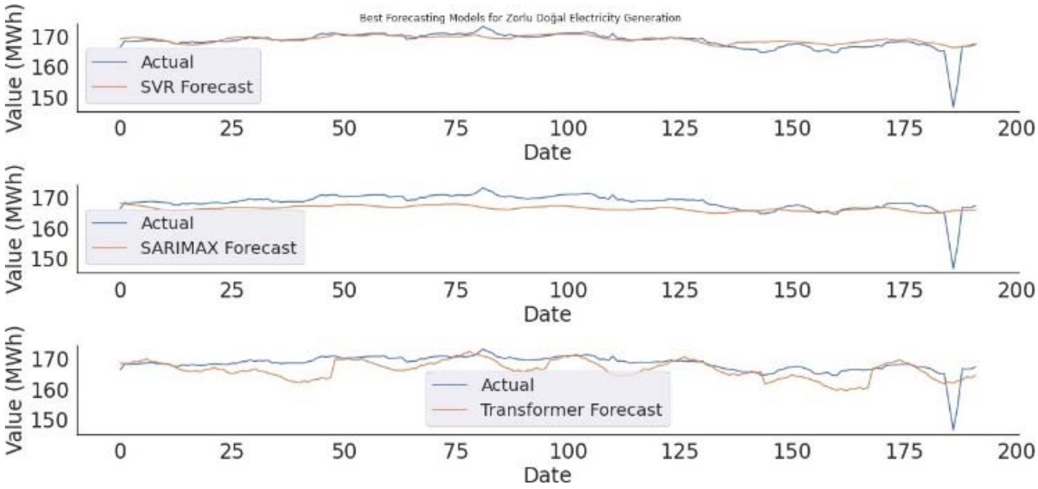

Fig. 10 First week of March comparison of best forecasting models and actual for Zorlu Doğal Electricity Generation (feature eng. dataset)

TABLE VII
FIRST WEEK OF MARCH BEST MODELS PERFORMANCE METRIC RESULTS FOR ZORLU DOĞAL ELECTRICITY GENERATION (FEATURE ENG. DATASET)

Models	MAE	MSE	RMSE	R^2	MAPE/MASE
SVR	1.08	4.69	2.16	0.38	0.66
SARIMAX	2.53	10.21	3.19	-0.33	1.50
Transformer	2.51	10.78	3.28	-0.40	2.51

4) *The first two weeks of March (feature eng. dataset):* The results of the models that give the best 2-weeks forecast for March as a result of using the newly added features dataset with feature engineering of the organization is as in Fig. 11 and Table VIII.

B. Imbalance Calculation Results

1) *The first week of March (main dataset):* Forecast and actual imbalance loss comparison results for the first week of March using the main dataset for organization is as in Fig. 12 and Table IX.

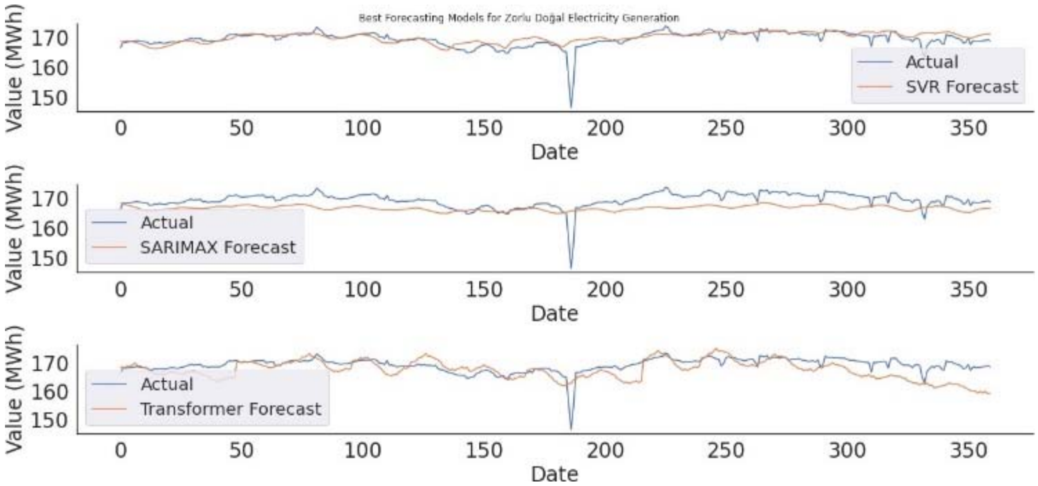

Fig. 11 First two weeks of March comparison of best forecasting models and actual for Zorlu Doğal Electricity Generation (feature eng. dataset)

TABLE VIII
FIRST TWO WEEKS OF MARCH BEST MODELS PERFORMANCE METRIC RESULTS FOR ZORLU DOĞAL ELECTRICITY GENERATION (FEATURE ENG. DATASET)

Models	MAE	MSE	RMSE	R^2	MAPE/MASE
SVR	1.19	4.44	2.10	0.27	0.72
SARIMAX	2.99	11.93	3.45	-0.93	1.76
Transformer	2.70	12.69	3.56	-1.06	2.70

TABLE IX
FIRST WEEK OF MARCH COMPARISON OF FORECAST AND REAL IMBALANCE LOSS FOR ZORLU DOĞAL ELECTRICITY GENERATION (MAIN DATASET)

Models	Forecast Imb. Loss	Real Imb. Loss	Saved Cost
SARIMAX	21,084.37 TL	123,363.87 TL	102,279.50 TL
SVR	21,895.38 TL	123,363.87 TL	101,468.49 TL
LSTM	41,950.22 TL	123,363.87 TL	81,413.65 TL

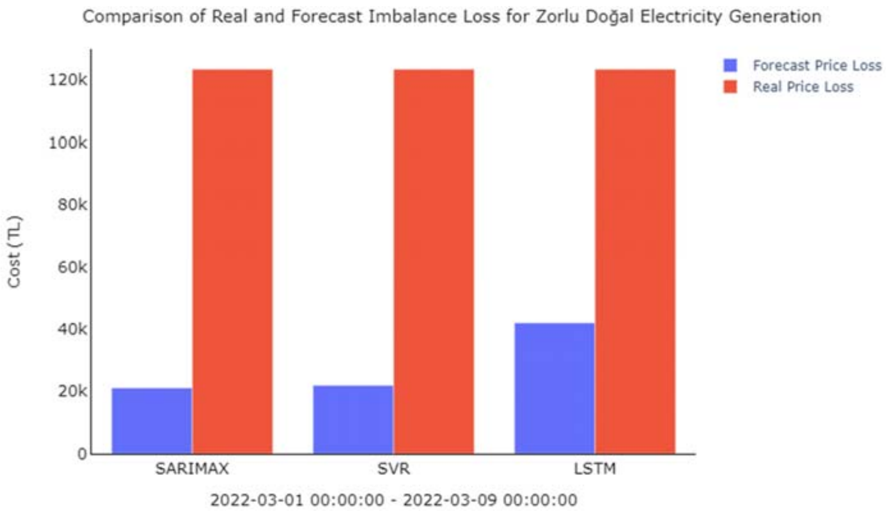

Fig. 12 First week of March comparison of best forecasting models and real imbalance loss on bar chart for Zorlu Doğal Electricity Generation (main dataset)

2) *The first two weeks of March (main dataset):* Forecast and actual imbalance loss comparison results for the first two weeks of March using the main dataset for organization is as in Fig. 13 and Table X.

TABLE X
FIRST TWO WEEKS OF MARCH COMPARISON OF FORECAST AND REAL IMBALANCE LOST FOR ZORLU DOĞAL ELECTRICITY GENERATION (MAIN DATASET)

Models	Forecast Imb. Loss	Real Imb. Loss	Saved Cost
SARIMAX	41,117.39 TL	228,977.16 TL	187,859.77 TL
SVR	52,647.23 TL	228,977.16 TL	176,329.93 TL
LSTM	145,526.65 TL	228,977.16 TL	83,450.51 TL

TABLE XI
FIRST WEEK OF MARCH COMPARISON OF FORECAST AND REAL IMBALANCE LOSS FOR ZORLU DOĞAL ELECTRICITY GENERATION (FEATURE ENG. DATASET)

Models	Forecast Imb. Loss	Real Imb. Loss	Saved Cost
SVR	15,719.74 TL	123,363.87 TL	107,644.13 TL
SVR	27,962.26 TL	123,363.87 TL	95,401.31 TL
LSTM	28,673.97 TL	123,363.87 TL	94,689.90 TL

3) *The first week of March (feature eng. dataset):* Forecast and actual imbalance loss comparison results for the first week of March using the newly added features dataset with feature engineering for all organization is as in Fig. 14 and

Table XI.

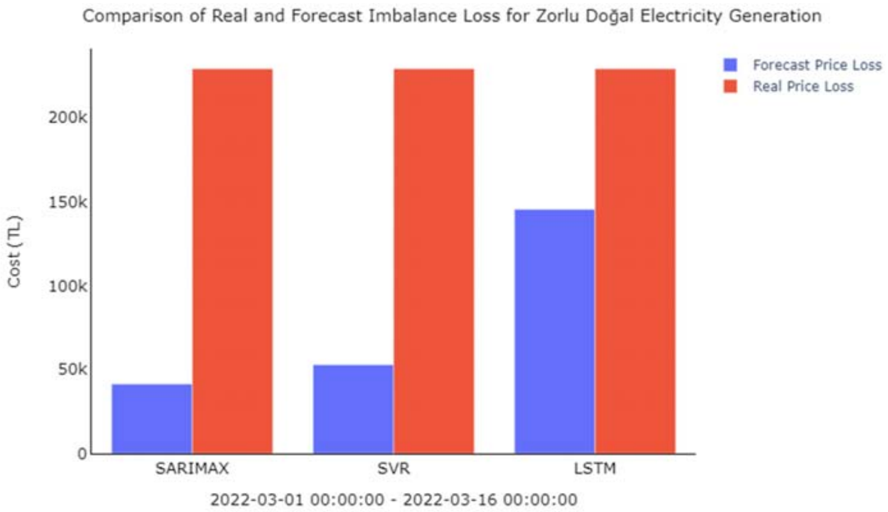

Fig. 13 First two weeks of March comparison of best forecasting models and real imbalance loss on bar chart for Zorlu Doğal Electricity Generation (main dataset)

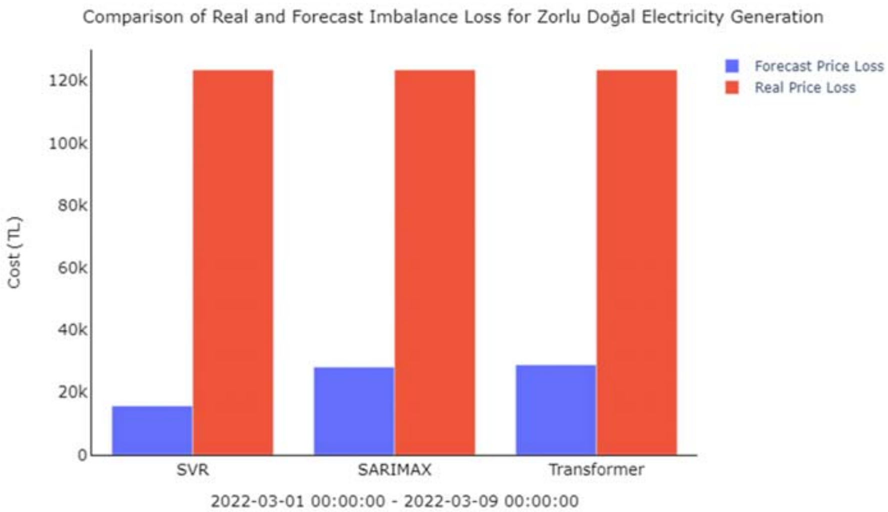

Fig. 14 First week of March comparison of best forecasting models and real imbalance loss on bar chart for Zorlu Doğal Electricity Generation (feature eng. dataset)

4) *The first two weeks of March (feature eng. dataset):* Forecast and actual imbalance loss comparison results for the first two weeks of March using the newly added features dataset with feature engineering for organization is as in Fig. 15 and Table XII.

TABLE XII
FIRST TWO WEEKS OF MARCH COMPARISON OF FORECAST AND REAL IMBALANCE LOSS FOR ZORLU DOĞAL ELECTRICITY GENERATION (FEATURE ENG. DATASET)

Models	Forecast Imb. Loss	Real Imb. Loss	Saved Cost
SVR	30,080.06 TL	228,977.16 TL	198,897.10 TL
Transformer	58,794.32 TL	228,977.16 TL	170,182.84 TL
SARIMAX	60,526.84 TL	228,977.16 TL	168,450.32 TL

C. Summary Results

TABLE XIII
BEST MODELS FOR THE FIRST 1 WEEK AND FIRST 2 WEEKS OF MARCH FORECAST

Date	Model	MAPE/MASE
2 weeks (main dataset)	SVR	1.10
2 weeks (feature eng.)	SVR	0.72
1 week (main dataset)	SVR	0.66
1 week (feature eng.)	SVR	0.66

There has not been much change in the best model rankings in March 1-week and 2-weeks forecasts. The model that gave the best predictions is SVR.

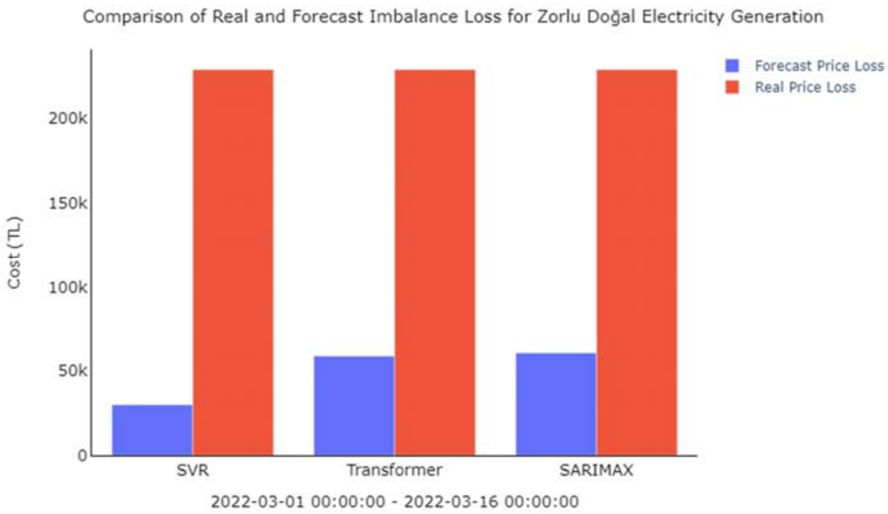

Fig. 15 First two weeks of March comparison of best forecasting models and real imbalance loss on bar chart for Zorlu Doğal Electricity Generation (feature eng. dataset)

V. Discussion

In this article, 2-weeks and 1-week forecasts were made for March, respectively, with the main dataset created with temperature and humidity variables to predict the total geothermal production of power plants belonging to Zorlu Doğal Electricity Generation organization using traditional machine learning, time series models and deep learning models. Then, 2-weeks and 1-week forecasts were made for March, respectively, using the dataset with newly added features with feature engineering. Using the prediction results of the models, loss of the imbalance penalties of the organizations were calculated for the first 1 week and first 2 weeks of March and compared with the actual imbalance loss. Finally, it was observed whether there would be a model change or improvement in the 2-weeks and 1-week forecasts within the two datasets, and the models and error rates were compared according to the use of the datasets and the forecasting times. According to the results of March; SVR model gave the best estimation results in 1-week and 2-weeks forecasts. In addition, the feature eng. applied dataset performed better than the main dataset. Table XIV shows the order of the date ranges and models that save the most money according to the estimated imbalance penalty, which took place in March for 2 weeks and 1 week.

TABLE XIV
March Comparison of Forecast and Real Imbalance Cost for Organization

Date	Models	Forecast – Real Imb. Loss	Saved Cost
2 weeks	FE SVR	30,080.60 TL-228,977.16 TL	198,897.10 TL
1 week	FE SVR	15,719.74 TL-123,363.87 TL	107,644.13 TL

SVR model with feature engineering was the model that saved the most money according to the difference between the estimated imbalance penalty in the 1-week and 2-weeks forecasts for organization.

VI. Conclusion

In this study, the total geothermal production of power plants belonging to Zorlu Doğal Electricity Generation organization with high geothermal installed capacity was estimated using machine learning, time series models, and deep learning models, and then the imbalance penalty paid by these organization for the difference between these estimates and the actual value about how much these organizations would produce in the energy market was calculated. These predictions were made for the first 1 week and 2 weeks of March, with the main dataset of the total geothermal production of the plants belonging to the organization and the temperature and humidity variables according to the cities where the plants are located, and with the feature engineering dataset from which new features such as year, month, day and hour were extracted from the main dataset. Geothermal production of power plants belonging to the organization, MCP and SMP data for imbalance calculation and matching sales amount for comparison were obtained from EXIST Web Service and temperature and humidity data were obtained from StormGlass Weather API service. All coding processes, including data collection, analysis, and modeling, are written in Python 3.8 version. Models were run on Google Colab. According to all tested models and all prediction processes, traditional machine learning models, SVR showed the best performances. In the use of the feature engineering dataset and the main dataset, the use of the dataset with the newly added features with the feature engineering produced better estimation results than the use of the main dataset. In addition, model changes were rarely observed in the 2-weeks and 1-week forecasts. It was observed that there was a high good difference between the production forecasts made in organization and the calculated imbalance penalty and the real imbalance penalty. In the future, this study can be continued to provide a better supply-demand balance in the market by testing the datasets prepared in power plants such as wind, hydroelectric, and especially geothermal production plants, and minimizing the imbalance in the energy market.

Acknowledgment

The authors would like to thank Ahmet Ozan Zeybek for his qualified knowledge and consultancy in the energy market. Also, they would like to thank Sudad Abdulrazzaq for his advice in the field of data science. Finally, authors would like to thank Bahcesehir University for opening the Artificial Intelligence Master Program and allowing there to carry out this thesis.

References

[1] Mertoglu, O., Simsek, S., Basarir, N. (2020). Geothermal Energy UseProjection, Country Update for Turkey. In Proceedings World Geothermal Congress (p. 1).

[2] EMRA (2021). Electricity Market Annual Sector Report List. Retrieved from https://www.epdk.gov.tr/Detay/Icerik/3-0-23/aylik-sektor-raporu at 2022

[3] EXIST (2022), About us. Retrieved from https://www.epias.com.tr/en/corporate/about-us/ at 2022

[4] EXIST (2016), Electricity Market Introduction. Retrieved from https://www.epias.com.tr/en/day-ahead-market/introduction/ at 2022

[5] Dinler, A. (2021). Reducing balancing cost of a wind power plant by deep learning in market data: A case study for Turkey. Applied Energy, 289, 116728.

[6] Ozdemir, S., Susarla, D. (2018). Feature Engineering Made Easy: Identify unique features from your dataset in order to build powerful machine learning systems. Packt Publishing Ltd.

[7] TE˙IAS¸ (2022), Power Plant Failure Maintenance Notices. Retrieved from https://tpys.teias.gov.tr/tpys/app/login.html at 2022

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.