1.

System Marginal Price Forecasting Project

Learning From Data

Supervised by:
Dr. Öğr. Üyesi Cumali TÜRKMENOĞLU

Made by: Asma Omar Ahmed Alrefaei – 2221251369
Ahmet BOZKIR – 2121251022

Introduction

The electricity market plays a crucial role in the energy sector, and accurate price forecasting is vital for stakeholders, including utilities, energy traders, and policy-makers. This project aims to develop robust forecasting models for electricity prices using different approaches, including **statistical**, **machine learning**, and **deep learning** models. These models aim to predict electricity prices based on historical data, weather conditions, and economic indicators. Through this analysis, we explore the performance of various models, identify the most effective ones, and demonstrate their potential for real-world applications.

Goal and Purpose

The goal of this project is to forecast electricity prices by utilizing machine learning (ML), deep learning (DL), and statistical models. The purpose is to:

- Predict the future electricity prices.
- Analyze the effect of weather and energy production data on price forecasting.
- Explore the viability of different models, such as **SARIMA**, **SVR**, **XGBoost**, **LSTM**, **GRU**, and **Transformer**.

By achieving this goal, we aim to provide a reliable and robust method for electricity price prediction that can assist market participants in optimizing their strategies.

Design

The approach involves data collection, preprocessing, feature engineering, model development, evaluation, and visualization. The design was divided into the following steps:

1. **Data Collection**: Collect historical data, including electricity prices and weather information from multiple cities, such as **Istanbul**, **Izmir**, **Adana**, and **Gaziantep**.
2. **Data Preprocessing**: Clean the data, handle missing values, create lag features, and normalize the data.
3. **Feature Engineering**: Generate time-related features (e.g., hourly, weekly, monthly) and weather-related features (e.g., temperature, humidity).
4. **Modeling**: Apply different models, including **SARIMA**, **SVR**, **XGBoost**, **LSTM**, **GRU**, and **Transformer**, to predict electricity prices.
5. **Evaluation**: Evaluate model performance using metrics like **MAE**, **SMAPE**, and **RMSE**.
6. **Visualization**: Create graphs and visualizations to compare the performance of different models and to show actual vs. predicted values.

Data Collection

For the Data Collection process, multiple datasets were used. These include:

1. Electricity Price Data: Collected from internal market data sources.
2. Weather Data: Collected using the Weatherbit API, providing hourly temperature and humidity data from cities such as Istanbul, Izmir, Adana, and Gaziantep.

Below is a sample of the data collection code for obtaining historical weather data using the Weatherbit API:

```
def get_historical_weather_data(api, city_name, city_id, start_date, end_date):
api = "e8ffcf15c38447f6bcc1cc39525ffb2c"
date_now = end_date url =
f"https://api.weatherbit.io/v2.0/history/hourly?city_id={city_id}&start_date={start_date}&end_date={date_now}&tz=local&key={api}
 response = requests.get(url)
data = json.loads(response.text)
datetime = [entry["datetime"] for entry in data["data"]]
humidity = [entry["rh"] for entry in data["data"]]
temp = [entry["temp"] for entry in data["data"]]
weather_data = pd.DataFrame({
 "Tarih": datetime,
f"{city_name}_temp": temp,
f"{city_name}_humidity": humidity
 })
```

```
return weather_data.iloc[2:]
```

After collecting the weather data, it was merged with energy market data (SMF, PTF) to create a final dataset with a comprehensive set of features for modeling.

Data Analysis & Feature Engineering

Exploratory Data Analysis (EDA)

Initial analysis of the data focused on understanding the distribution of features like electricity prices (SMF, PTF), weather data (temperature, humidity), and economic indicators. Key findings included:

- Price Distribution: The electricity prices showed some seasonal trends, with higher prices in winter months.
- Correlation: There was a strong correlation between weather variables (temperature and humidity) and electricity prices.
- Missing Values: Some weather and economic indicators had missing values, which were handled by forward and backward filling.

Feature Engineering

To improve model performance, we created several lag features (e.g., **Ptfdolar_Lag1**, **Ptfdolar_Lag7**, **Ptfdolar_Lag30**) and handled categorical variables like **Season** and **Is_Weekend** using **LabelEncoder**.

The features used in the models included:

- **Lag Features**: Past values of electricity prices.
- **Weather Features**: Temperature, humidity for cities such as **Istanbul**, **Izmir**, **Adana**, and **Gaziantep**.
- **Economic Indicators**: Variables like **Dolar**, **Euro**, **Ptfdolar**, **Smfdolar**.
- **Time-Related Features**: Hour, Day, Week, Month, and Season

Models (SARIMA, ML, DL)

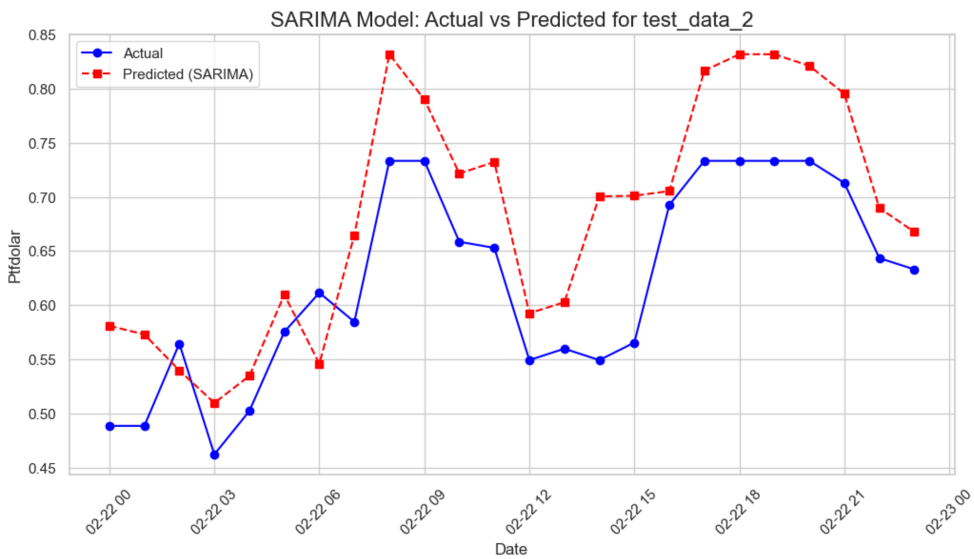
SARIMA (Seasonal Autoregressive Integrated Moving Average)

SARIMA is a statistical time-series forecasting model that combines autoregressive (AR), moving average (MA), and seasonal components to model and predict time-series data.

- **Order**: (1,1,0) for the AR, differencing, and MA components.
- **Seasonal Order**: (1,1,0,24) to capture seasonal effects over a 24-hour period.

Results:

- Train MAE: 0.0330 | Test MAE: 0.1041
- Train SMAPE: 7.82% | Test SMAPE: 8.74%
- Train RMSE: 0.1068 | Test RMSE: 0.1259



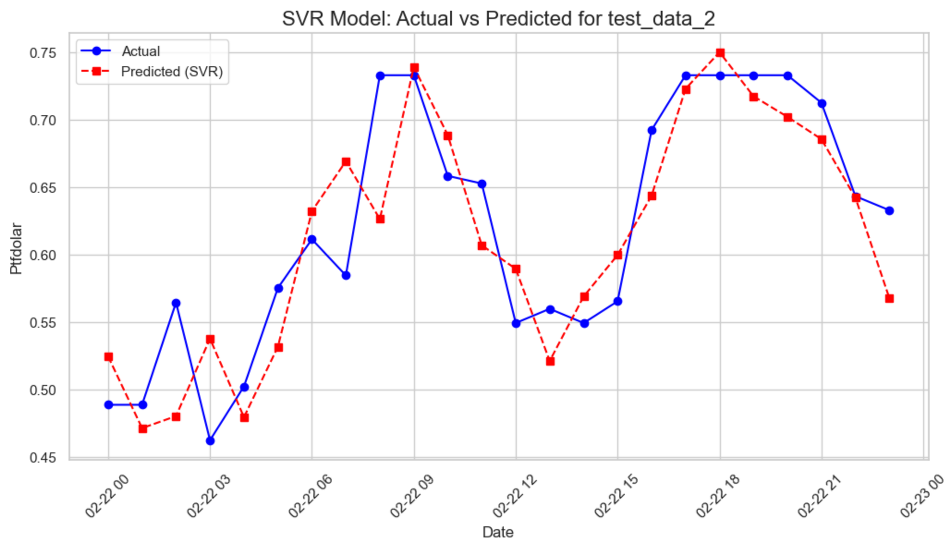
Machine Learning Models

SVR (Support Vector Regression)

SVR was trained on the lag features and other relevant variables like weather conditions and economic indicators.

Results:

- Train MAE: 0.0583 | Test MAE: 0.0483
- Train SMAPE: 6.45% | Test SMAPE: 4.21%
- Train RMSE: 0.0844 | Test RMSE: 0.0625

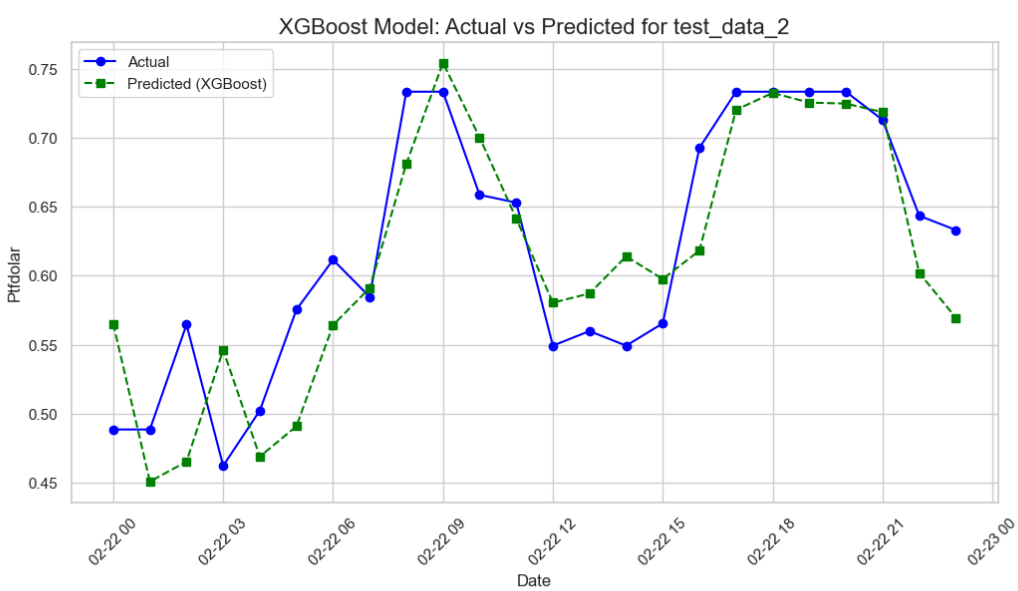


XGBoost

XGBoost is a gradient-boosted decision tree algorithm known for its efficiency and accuracy in regression tasks.

Results:

- Train MAE: 0.0471 | Test MAE: 0.0393
- Train SMAPE: 5.33% | Test SMAPE: 3.49%
- Train RMSE: 0.0682 | Test RMSE: 0.0542

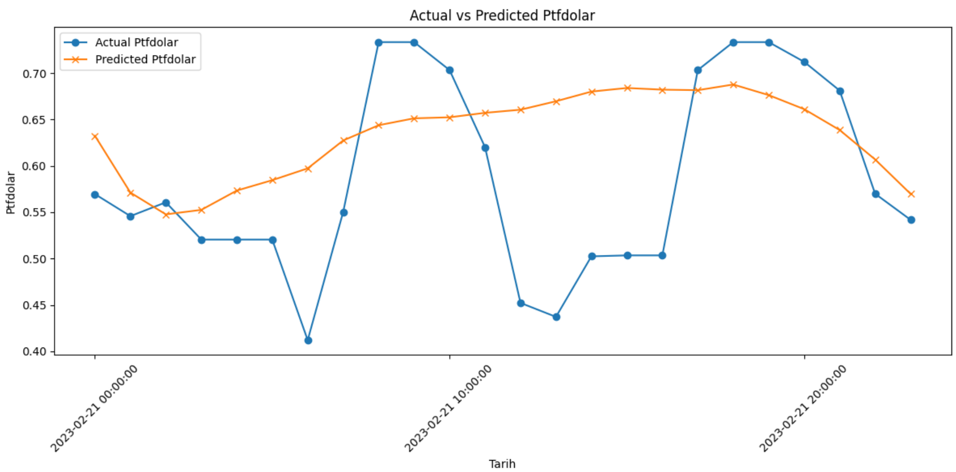


Deep Learning Models (LSTM, GRU, Transformer)

We used deep learning models to capture non-linear relationships in the data.

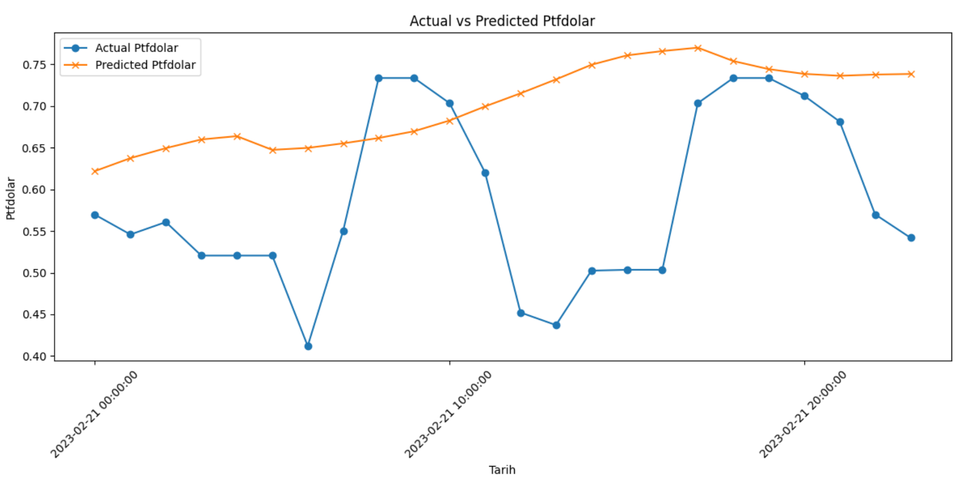
LSTM (Long Short-Term Memory)

LSTM is a type of recurrent neural network (RNN) designed to model sequences of data with long-range dependencies.



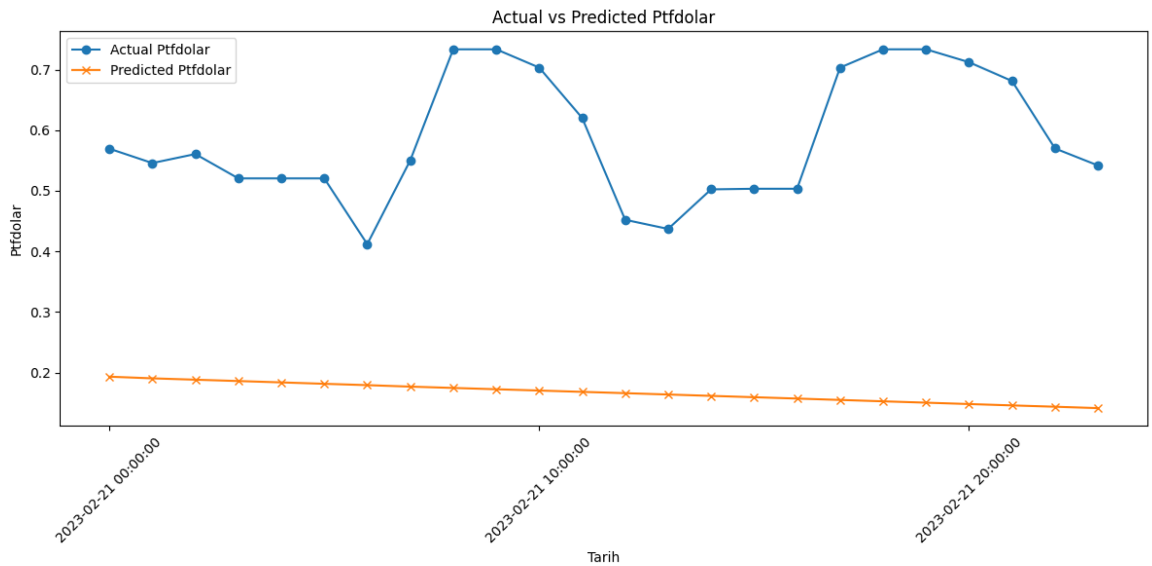
GRU (Gated Recurrent Unit)

GRU is a simplified version of LSTM, with fewer parameters and faster computation, while maintaining similar performance.



Transformer

The Transformer model leverages self-attention mechanisms and is especially useful in handling sequential data.



Results and Evaluations

The models were evaluated using the following metrics:

- **MAE (Mean Absolute Error)**
- **SMAPE (Symmetric Mean Absolute Percentage Error)**
- **RMSE (Root Mean Squared Error)**

Overall Model Performance

The results from all models were compared to evaluate their predictive power. SARIMA showed a solid performance, especially for short-term predictions, while machine learning models (SVR and XGBoost) outperformed SARIMA in terms of both MAE and SMAPE. The deep learning models (LSTM, GRU, Transformer) showed promise for more complex patterns but took longer to train.

Conclusion

In conclusion, the project demonstrated that both statistical models (SARIMA) and machine learning models (SVR, XGBoost) could predict electricity prices effectively, with machine learning models showing a slight edge in performance. The deep learning models showed promise for more complex, non-linear patterns but required longer training times.

The analysis indicates that including weather and economic features significantly enhances the predictive capability of the models. Future work may involve optimizing deep learning models and experimenting with additional features such as real-time market data.

Learning From Data

System Marginal Price Forecasting Project

Supervised by: Dr. Öğr. Üyesi Cumali TÜRKMENOĞLU

Done by: Asma Omar Ahmed Alrefaei – 2221251369
Ahmet BOZKIR – 2121251022