UG Project

Load Forecasting using Machine Learning and Statistical Techniques

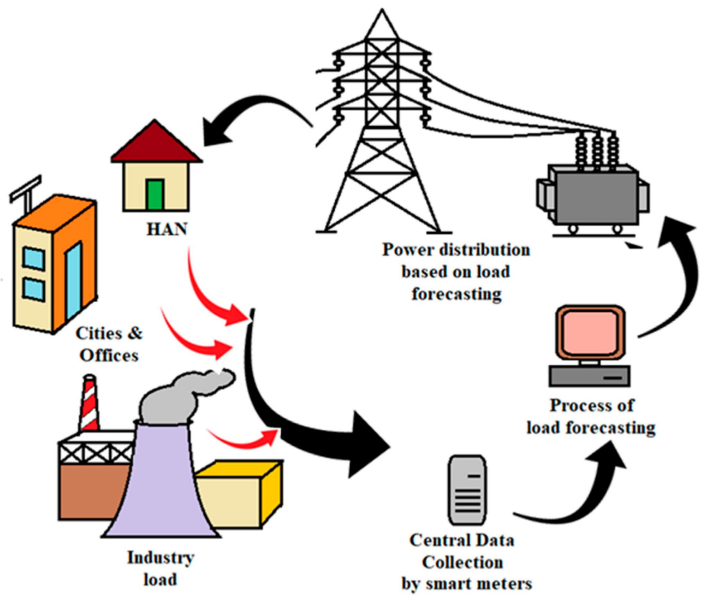
Under the supervision of Prof. Sobhita Mehar

BY :
Eshaan Agarwal 20085027 (EEE Part-3 B.Tech)
Aadish Jain 20085001 (EEE Part-3 B.Tech)
Tanvi Sharma 20085102 (EEE Part-3 B.Tech)

Content

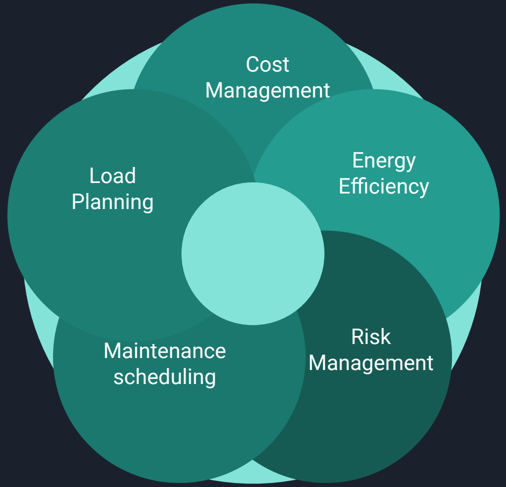
Load Forecasting

- Process of predicting the future demand or consumption of electricity or energy in a particular area or system.
- Use of statistical and mathematical models to analyze historical consumption data and identify trends, patterns, and factors that may affect energy consumption in the future.
- Essential tool for energy companies, utilities, and grid operators to plan and optimize their generation and transmission resources, as well as to ensure that they can meet the future energy demands of their customers.
- Timely implementations of such decisions lead to improvement of network reliability and to reduced power failures.

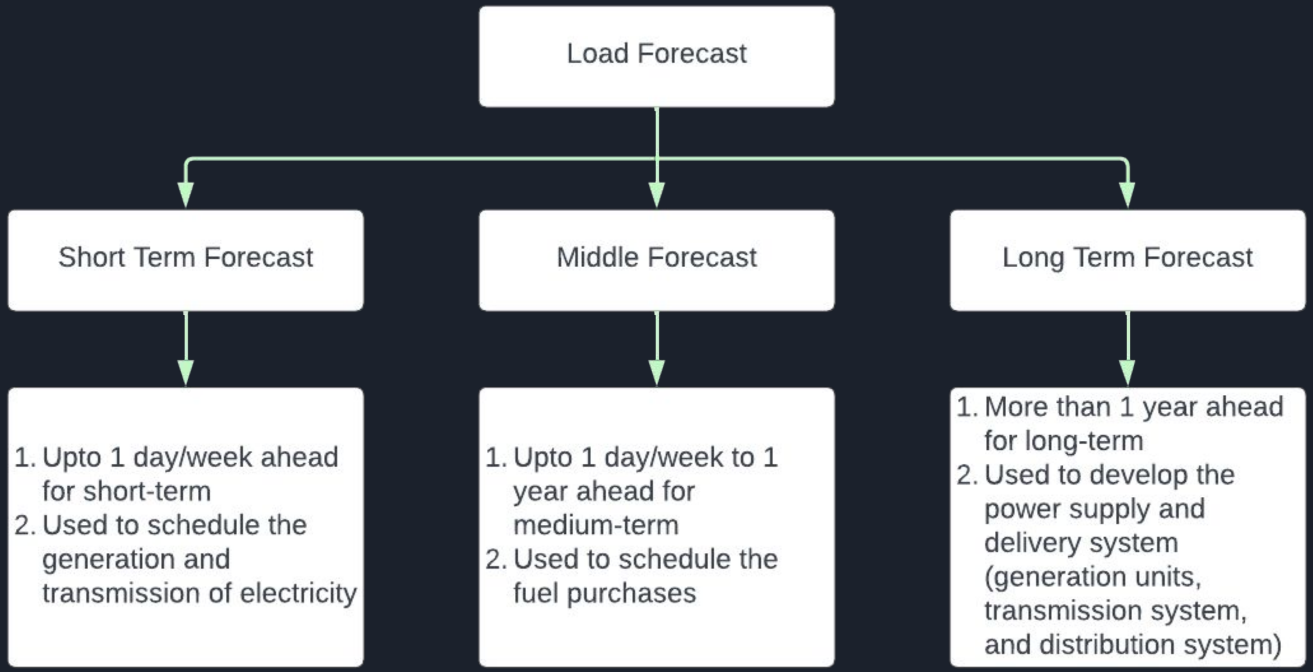


Purpose of Load forecasting

- **Planning**: Plan for future energy demand and ensure that there is enough capacity to meet that demand. Accurate load forecasting helps utilities and energy providers determine the optimal amount of resources to invest in power generation and transmission infrastructure.
- **Cost management**: Helps energy companies manage costs by reducing the need for expensive peak load generators or purchasing power from external sources during periods of high demand.
- **Energy efficiency**: Helps in identifying periods of low demand when energy usage can be reduced, helping to promote energy efficiency and reduce costs.
- **Maintenance scheduling**: Help utilities and energy providers schedule maintenance and repairs to power generation and transmission infrastructure during periods of low demand, minimizing the impact on energy supply.
- **Risk management**: Energy companies manage risks associated with energy price volatility, capacity shortages, and unexpected events such as extreme weather conditions.

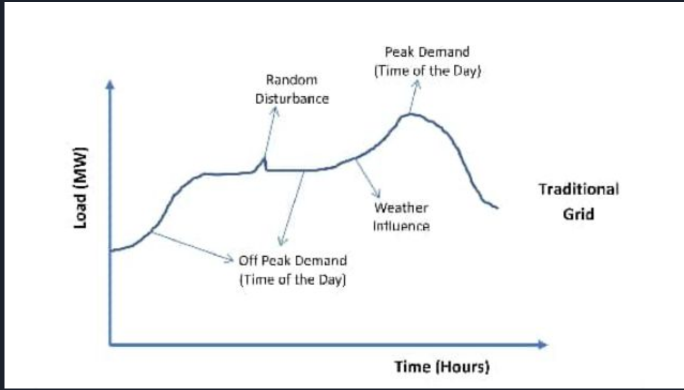


Types of forecasting

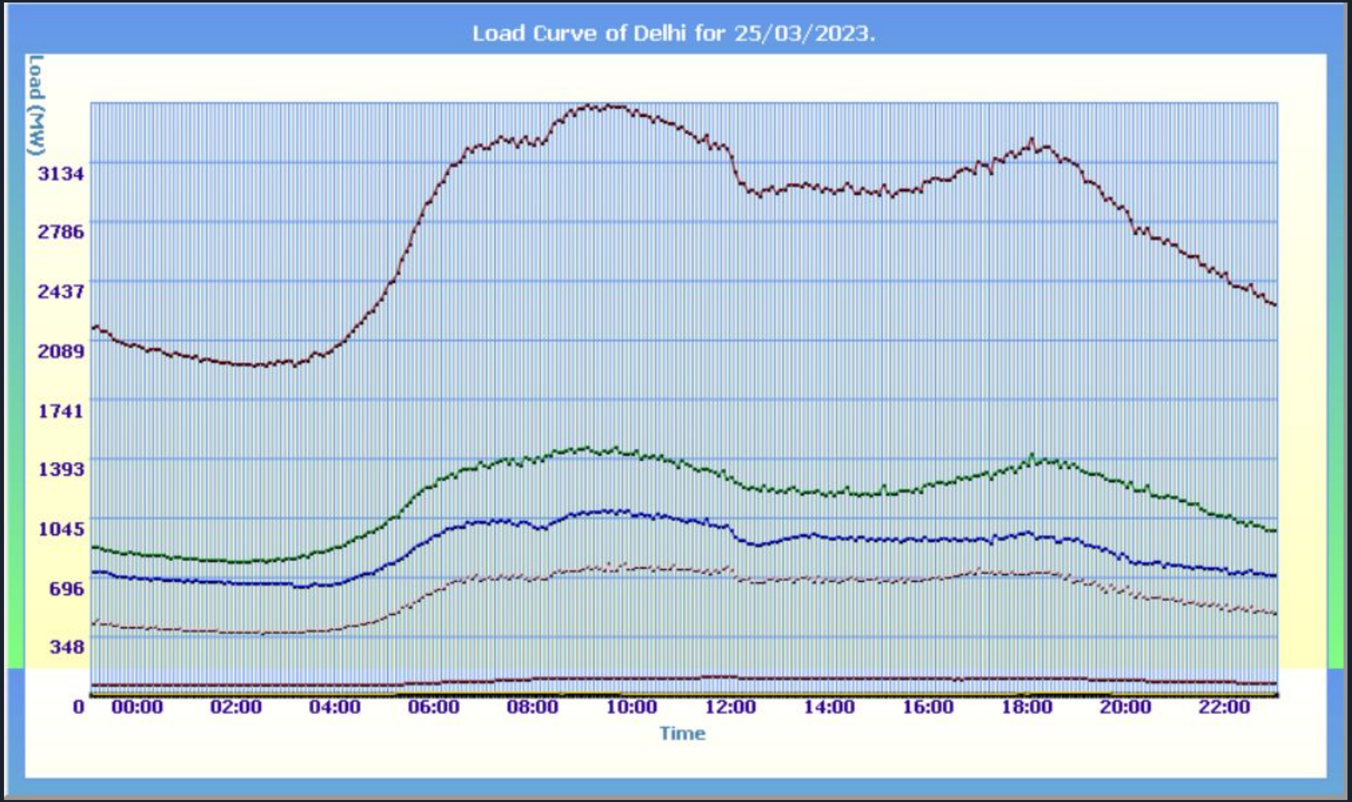


Factors affecting Load forecasting

- **Weather conditions**: Extreme temperatures, such as heatwaves and cold snaps, can significantly impact energy demand for heating and cooling.
- **Time of day and day of the week**: Energy demand varies depending on the time of day and day of the week, with peak demand typically occurring during weekday afternoons and evenings.
- **Seasonal variations**: Seasonal variations in energy demand, such as higher demand for heating during the winter months, can impact load forecasting.
- **Policy and regulatory changes**: Changes in energy policies and regulations, such as the introduction of renewable energy incentives or carbon taxes impact energy demand and load forecasting.



Load Curve of Delhi at 25/03/2023



Dataset

- Dataset - Time series data of Delhi Electricity Board's Data which provides demand of the state in load values (MW) in time step of 5 minutes using Data Scraping Script
  - Can be used to scrape data for real-time prediction
  - 288 values (24*60/5) of time-series data for each day
  - Scraped Data for 5 months using Beautiful Soup from Oct 2022 to Feb 2023
  - Total data points - approx 43200 data points (288*5*30)

- Humidity and Temperature data in time step of 5 minutes for corresponding period scraped from *Wunderground* weather site for NewDelhi (INEWDE9)

- Features Used -
  - Time of the day.
  - Day of the week.
  - Temperature at the corresponding time.
  - Humidity at the corresponding time.

Dataset Preprocessing and Scrapping

	datetime	load	humidity	temperature
0	01/10/2022 00:00	4581.220	46	24
1	01/10/2022 00:05	4553.49	49	22
2	01/10/2022 00:10	4533.33	44	24
3	01/10/2022 00:15	4507.54	60	27
4	01/10/2022 00:20	4489.97	43	20
...	...	...	...	...
35026	30/03/2023 23:20	2353.0	42	23
35027	30/03/2023 23:25	2342.45	45	30
35028	30/03/2023 23:30	2347.89	60	30
35029	30/03/2023 23:35	2333.57	65	23
35030	30/03/2023 23:40	2308.27	53	28

This data is being extracted from SLDC website of Delhi and Wunderground by an automated script.

The mentioned website updates the data for every 5 minutes. Thus we have data in the time steps of 5 minutes.

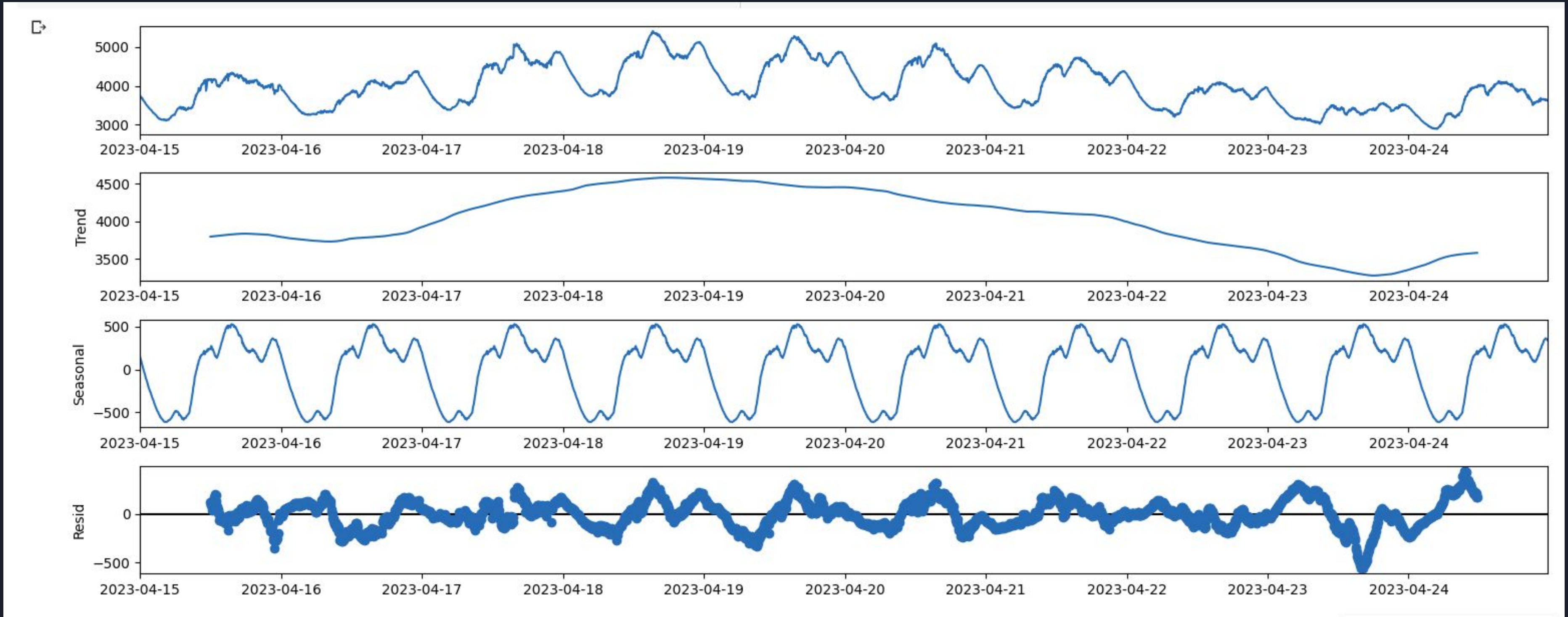
Problem - However, there are certain days when the SLDC website was not functional due to maintenance or any other factors, which implies the data collected might have some missing values.

Solution used - Used standard techniques for imputing like forward fill and back fill to make data more uniform

```
data_new = data.asfreq(freq='5T', method='bfill')
```

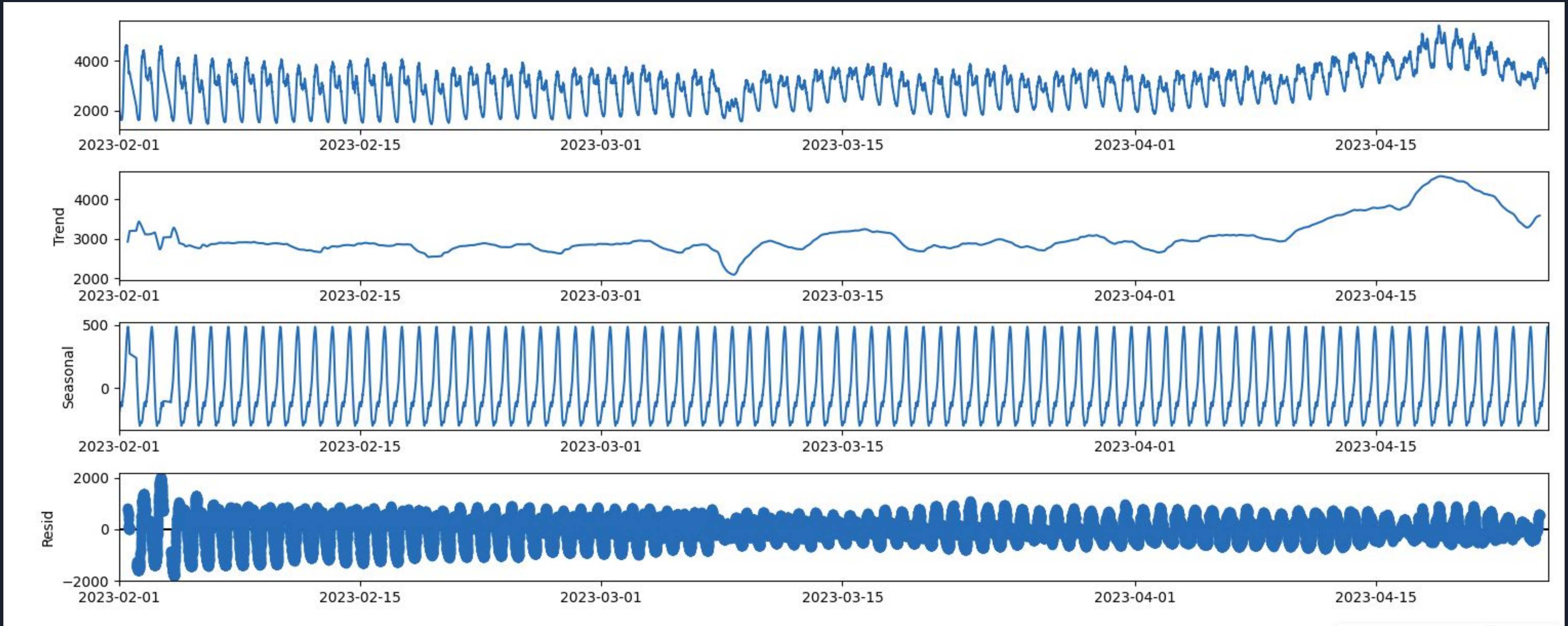
Dataset Analysis (Trends and Seasonality)

Trends and Seasonality in Load data from 04-15-23 to 24-04-23



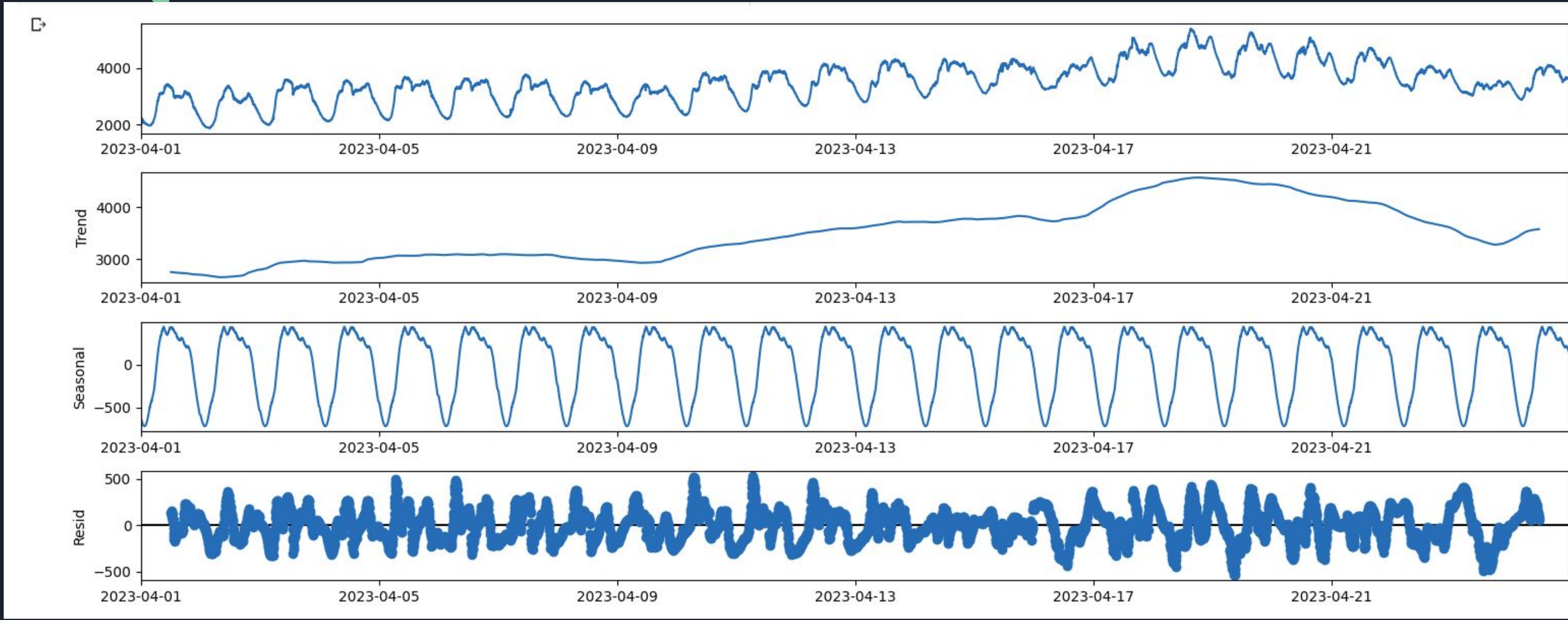
Dataset Analysis (Trends and Seasonality)

Trends and Seasonality in Load data from 01-02-23 to 24-04-23



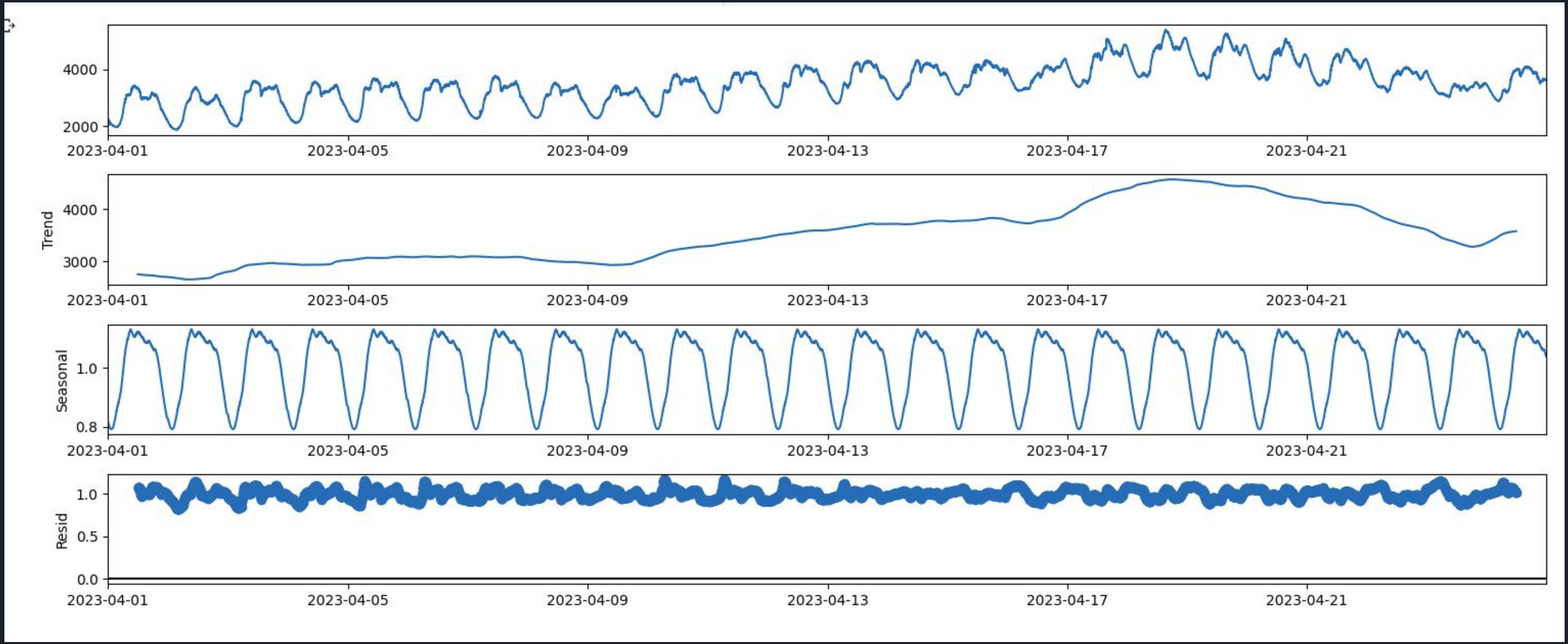
Dataset Analysis (Trends and Seasonality)

Weekly Trends and Seasonality in Load data from 01-04-23 to 24-04-23



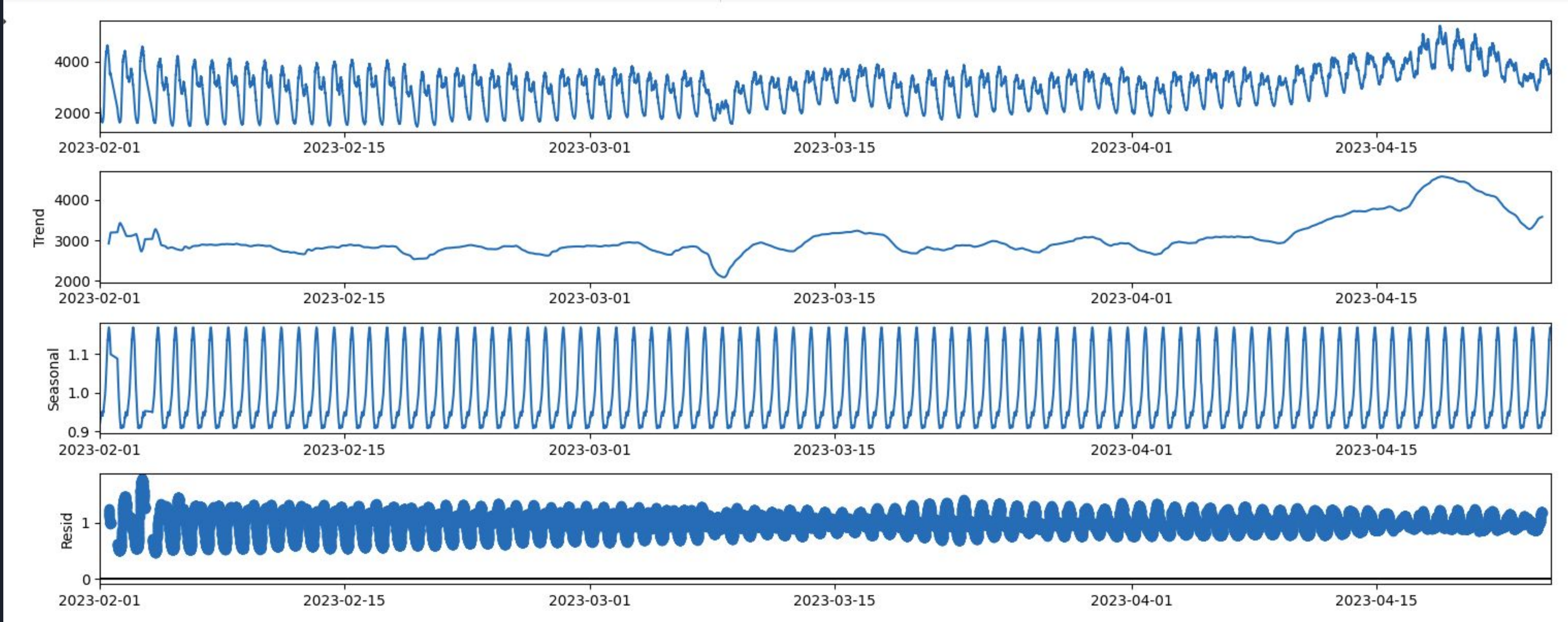
Dataset Analysis (Trends and Seasonality)

Multiplicative Trends and Seasonality in Load data from 01-04-23 to 24-04-23



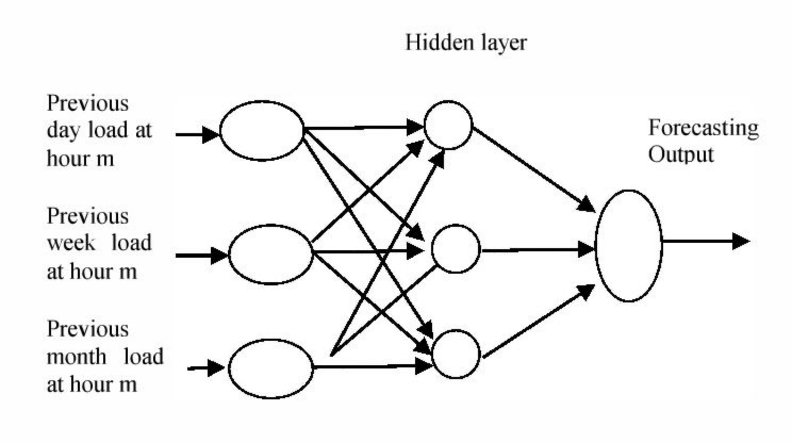
Dataset Analysis (Trends and Seasonality)

Multiplicative Trends and Seasonality in Load data from 01-02-23 to 24-04-23



Algorithms Implemented for Load forecasting

1. Simple Moving Average (SMA) - link
2. Weighted Moving Average (WMA) - link
3. Simple Exponential Smoothing (SES) - link
4. AutoRegressive Integrated Moving Average (ARIMA) - link
5. Seasonal AutoRegressive Integrated Moving Average (SARIMA) - link
6. XGBoost Gradient Boosting Algorithm - link
7. Simple Feedforward Neural Network (FFNN) - link



Simple Moving Average (SMA)

- This method involves calculating the average of the load data for a specific time period and then using this value to forecast the load for the next time period.
- The formula for a simple moving average is

$$F_t = \frac{\sum_{i=1}^{n} A_{t-i}}{n}$$

where F(t) = Forecast for the coming period, n = Forecast for the coming period, and A(t-1), A(t-2), A(t-3) and so on are the actual occurrences in the in the past period, two periods ago, three periods ago and so on respectively.

- There is a closer following of the trend in a shorter time span but it produces more oscillation. On other hand, a longer time span gives a smoother response but lags the trend.

Drawbacks of Simple Moving Average (SMA)

- **Limited accuracy:**
  - Good for short-term forecasts, but its accuracy tends to decrease for longer-term forecasts.
  - Method assumes that the load pattern is stable over the forecast horizon and does not account for other factors that may affect the load.
- **Ignores trends and seasonality:**
  - Assumes that the load demand is constant over time and does not consider trends or seasonality in the data which Lead to inaccurate forecasts.
- **Slow to respond to changes:** Based on a fixed window of historical data, which means that it can be slow to respond to sudden changes in load demand. This can be a problem in situations where load demand changes rapidly, such as during **extreme weather events.**

Weighted Moving Average (WMA)

- The simple moving average gives equal weight to each component of the moving average database, a weighted moving average allows any weights to be placed on each element, providing, of course, that the sum of all weights equals 1.
- The formula for the weighted moving average is

$$F_t = \sum_{i=1}^{n} w_i A_{t-i} \qquad \sum_{i=1}^{n} w_i = 1$$

Where F(t)= Forecast for the coming period, n = the total number of periods in the forecast, w(i) = the weight to be given to the actual occurrence for the period t-i, A(i) = the actual occurrence for the period t-i.

Choosing weights for Weighted Moving Average (WMA)

- Experience and trial and error are the simplest ways to choose weights.
- As a general rule, the most recent past is the most important indicator of what to expect in the future, and, therefore, it should get higher weighting.
- However, it can be more complex to implement and requires more frequent adjustments to the weights as the load patterns change over time.
- The WMA method is suitable for both short-term and medium-term load forecasting.

Simple Exponential Smoothing (SES)

- The Simple Exponential Smoothing (SES) method is a widely used statistical technique for load forecasting. It is based on the assumption that the future load will be a function of the past load values and an exponentially weighted average of the past forecast errors.
- The simple exponential smoothing forecast is given by

$$F_t = \alpha A_{t-1} + (1 - \alpha)F_{t-i}$$

where F(t) is the exponentially smoothed forecast for ith interval, α is the desired response rate, or smoothing constant.

AutoRegressive Integrated Moving Average (ARIMA)

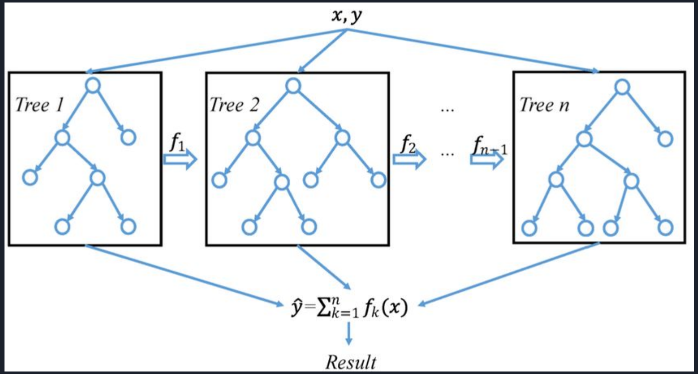
- ARIMA models are widely used because they can capture the patterns and trends in historical load data and use them to make predictions about future load.
- It can capture both short-term and long-term patterns in the data and adjust to changes in those patterns over time and trends in the data, as well as are able to handle non-stationary data.
- The ARIMA model is a combination of three components:
  - **Autoregression (AR) -** this component models the linear relationship between the previous values of the series and the current value.
  - **Moving Average (MA) -** this component models the linear relationship between the errors (the difference between the actual values and the predicted values) of the series and the lagged values of the errors.
  - **Integration (I) -** this component models the non-stationary behavior of the series by taking the differences between the values of the series at different time points.

Comparing SMA, WMA, SES and ARIMA

1. **Accuracy**: ARIMA is generally considered to be the most accurate method for load forecasting. SES can also be reasonably accurate for short-term forecasts, while SMA and WMA are generally less accurate.
2. **Flexibility**: ARIMA and SES are more flexible than SMA and WMA. They can adapt to changes in load demand patterns over time and account for trends, seasonality, and other factors that can affect load demand. SMA and WMA, on the other hand, are based on a fixed window of historical data and cannot adapt to changes in demand patterns.
3. **Computation**: SMA and WMA are relatively simple and computationally efficient methods. SES is slightly more complex than SMA and WMA, while ARIMA is the most computationally intensive method.
4. **Data requirements**: SMA and WMA only require historical load demand data, while SES and ARIMA may require additional data such as weather data, economic data, or other external factors that can influence load demand.

XGBoost Gradient Boosting Algorithm

- Gradient boosting is an ensemble learning method that combines the predictions of multiple weak learners (typically decision trees) to create a more accurate and robust model.
- The key idea behind gradient boosting is to sequentially fit new base learners to the residuals (or errors) of the previous learners, with the aim of reducing the residuals at each step.
- One of the main features of XGBoost is its optimized algorithm for finding the best splits in the decision trees. Instead of exhaustively searching all possible splits, XGBoost uses a technique called "approximate tree learning" to efficiently find the optimal splits, which greatly speeds up the training process.



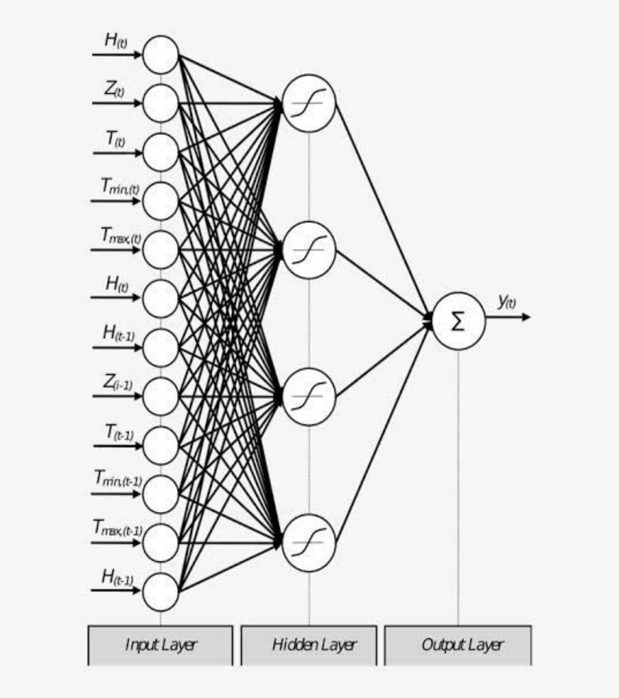
Benefits of XGBoost Gradient Boosting Algorithm

- **Accuracy**: Known for its improved accuracy compared to traditional methods.
- **Efficiency**: Optimized for speed and efficiency, making it well-suited for large-scale load forecasting tasks. It can handle large datasets and non linear relationships.
- **Flexibility:** XGBoost is flexible and can handle various types of data, including numerical, categorical, and ordinal data. This makes it well-suited for load forecasting tasks that involve different types of features.
- **Regularization:** Handles missing data values effectively by assigning a default direction for them when splitting a tree node. This means that the missing values are treated as either left or right children, depending on which direction gives the best split.

```
model = XGBRegressor(n_estimators=1000, max_depth=7, eta=0.1, subsample=0.7, colsample_bytree=0.8)
```

Simple Feedforward Neural Network (FFNN)

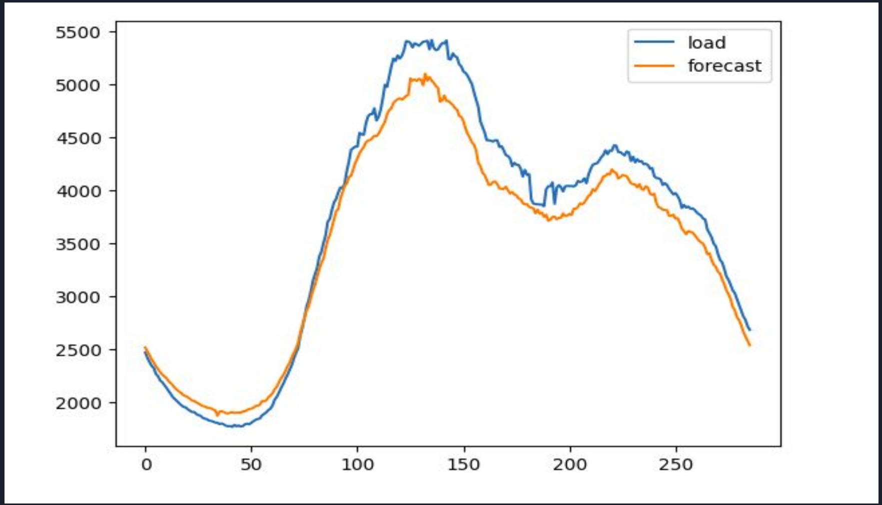
- Type of artificial neural network.
- Called **"feedforward"** because the data flows in one direction through the network, from input to output, without any feedback connections or loops.
- Neurons in each layer are fully connected to the neurons in the next layer, and non linear activation function introduces non-linearity into the network and **allows it to learn complex patterns in the data.**
- FFNNs are particularly well-suited for load forecasting because they can learn complex non-linear relationships between the load demand and various input factors.
- Suitable for Short term, mid-term and long term forecasting based on the complexity and volume of data fed to the network.



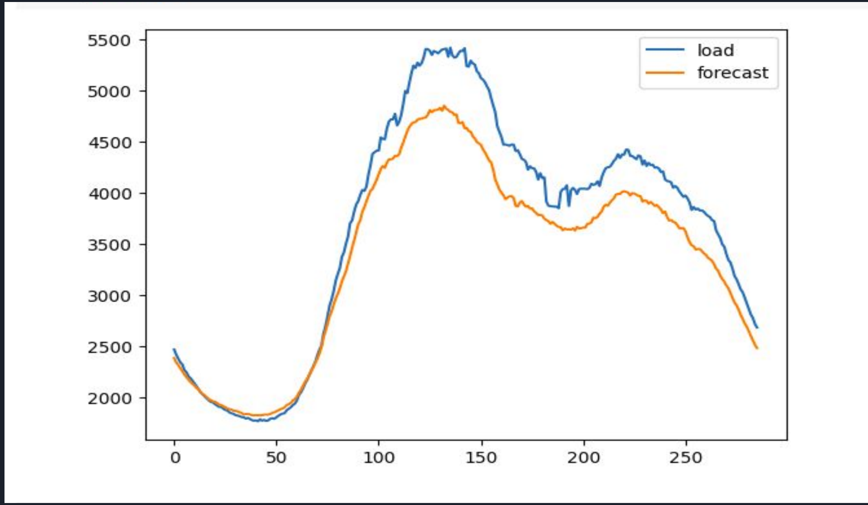
Results of various algorithms implemented

Algorithm	MAPE	MAE
SMA	128.61	3.16
WMA	82.45	1.83
SES	102.47	2.58
SARIMA	178.94	3.73
XGboost	145.13	3.42
FFNN	156.72	3.66

Results

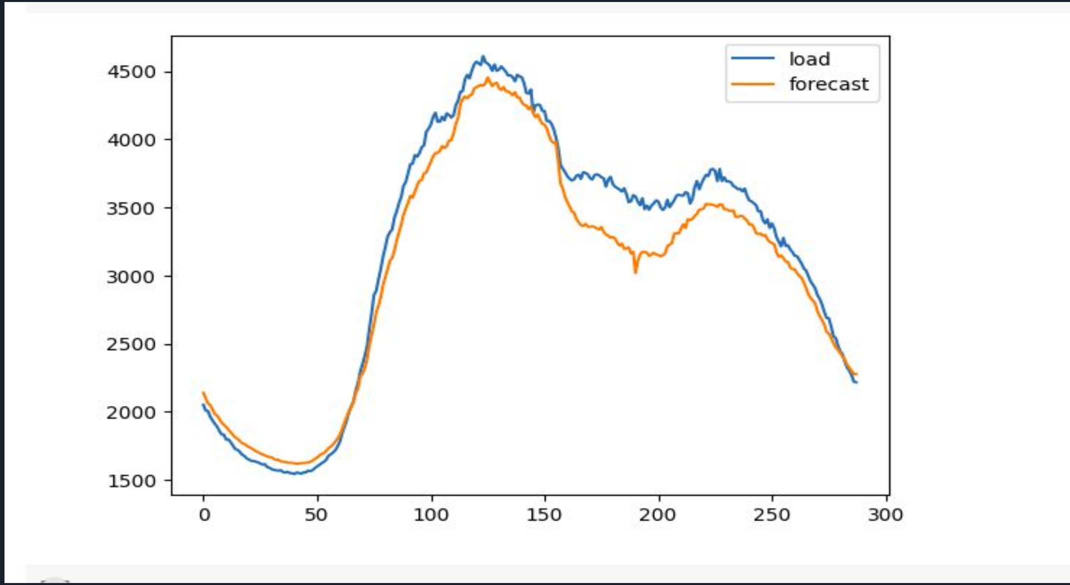


Actual Load vs forecasted load for SMA at 14-04-23



Actual Load vs forecasted load for WMA at 14-04-23

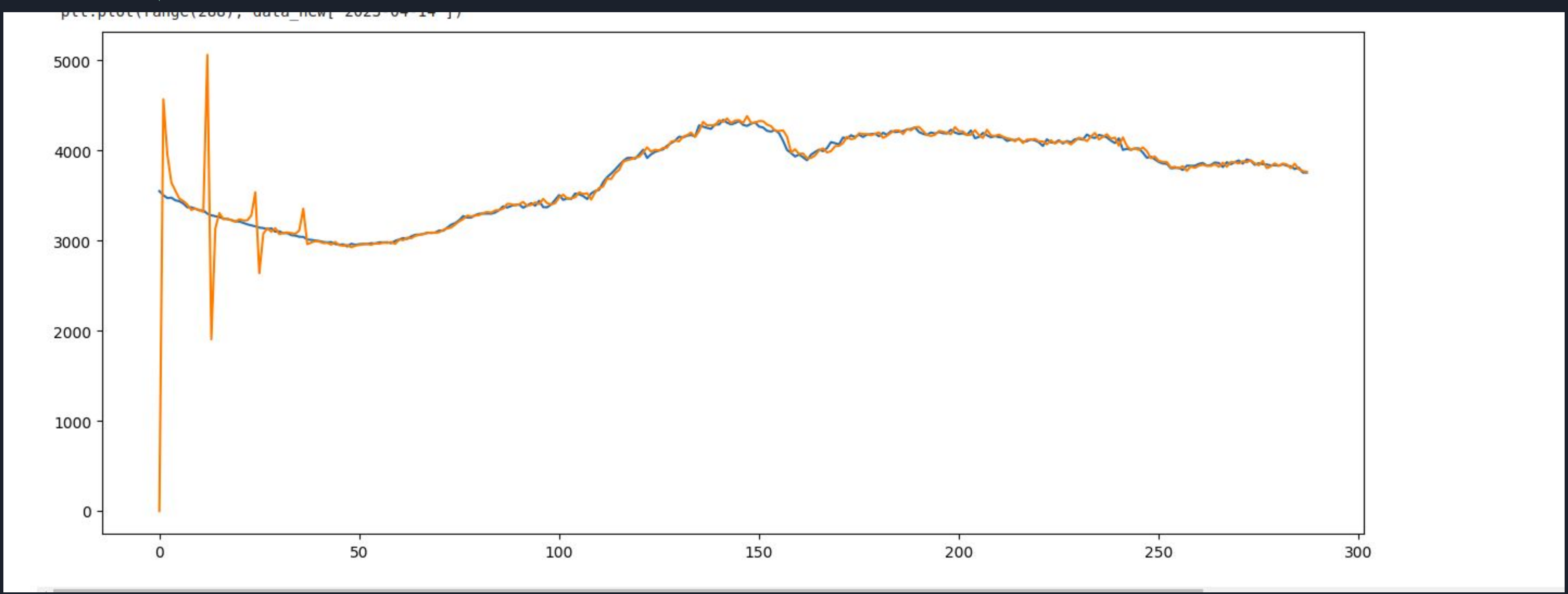
Results



Actual Load vs forecasted load for SES at 14-04-23

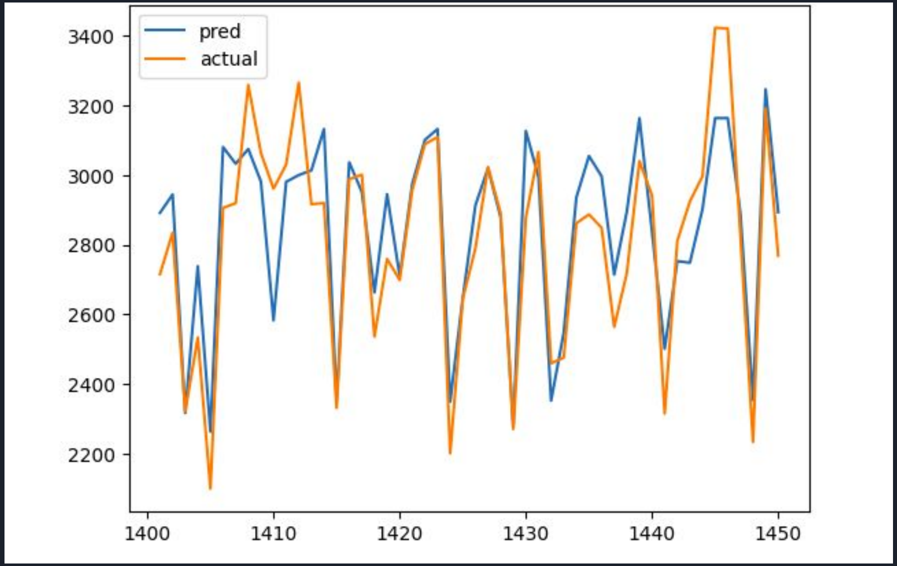
Results

Actual Load vs Forecasted Load using ARIMA

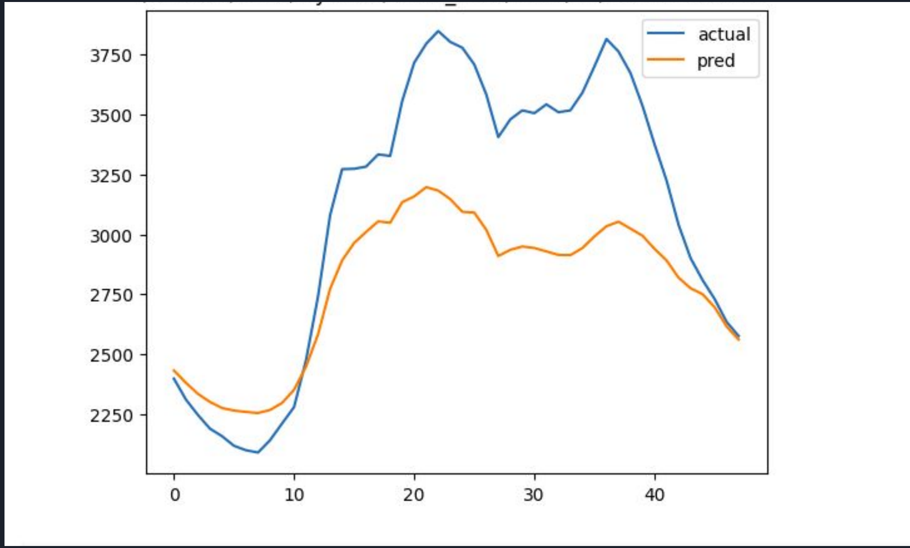


Results

Feed Forward Neural Network



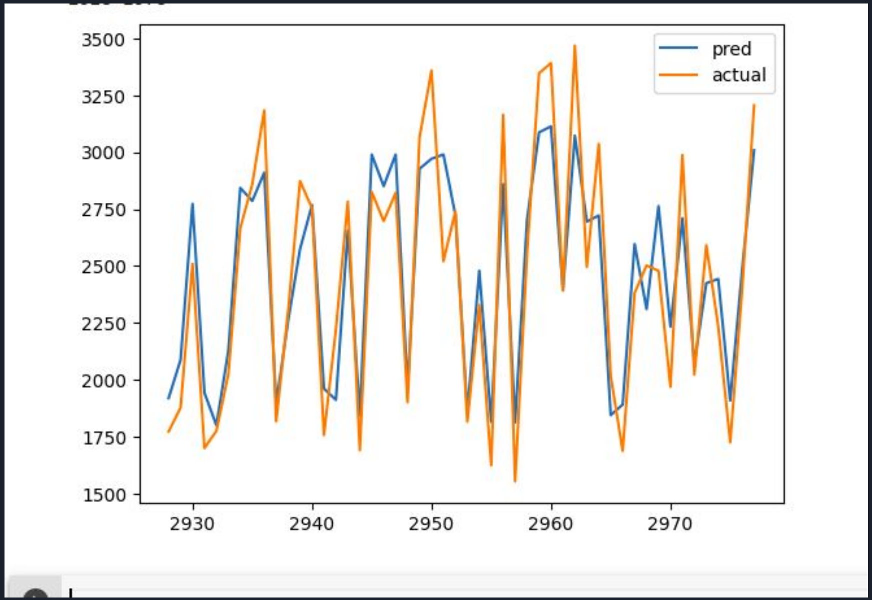
Actual Load vs Predicted Load across a period of 50 days



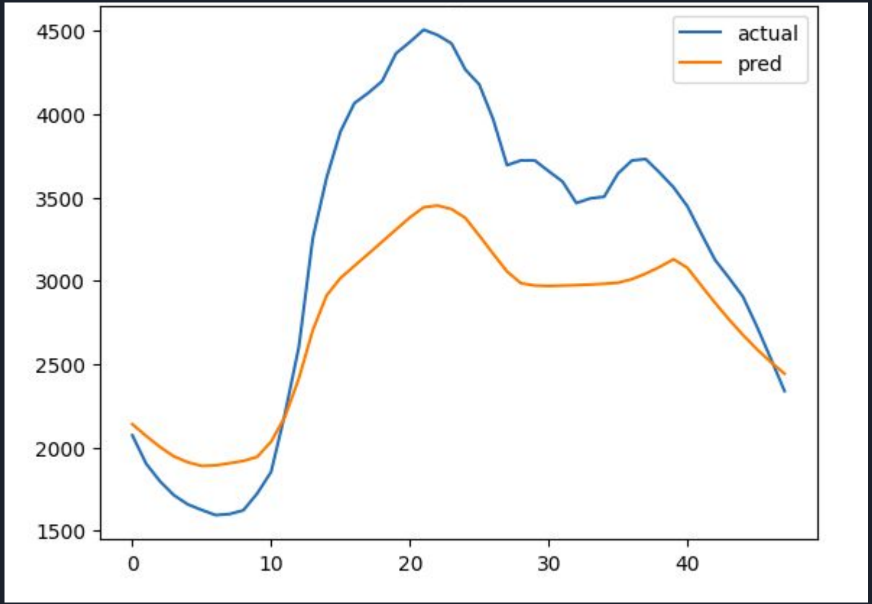
Actual Load vs Predicted Load on 13-03-23

Results

XGBoost



Actual Load vs Predicted Load across a period of 50 days



Actual Load vs Predicted Load on 15-03-23

Future Plans

1. Use of **Recurrent Neural Networks (RNN)**, **LSTMS and GRU** in Load Forecasting is highly useful as it maintains context across a series. In particular, these features of sequence models allow you to carry information across a larger time window than simple deep neural networks.
2. Use of **Convolutional Neural Network (CNNS)** for load forecasting has not been done properly and it can provide us interesting results in terms of time series analysis.
3. Combining ESE's exponential smoothing with RNN to make **ES-RNN** for time series forecasting
4. Use of **Graph Convolutional Network (GCN)** for time series analysis by extracting temporal and spatial insights from load demands curve for time series.
5. Combining **Auto regressive features of ARIMA with self attention mechanism of Transformers** for huge data of univariate Time Series Forecasting.
6. **Weighted Average Ensemble** of various techniques to provide more robust and stable load forecasting result

Conclusion

- Load Forecasting is an important application of machine learning in electrical engineering which is significantly gaining importance due to increase complexity in transmission and load networks.
- Implemented and conducted literature survey of various state of art techniques of load forecasting
- Statistical techniques like SMA, WMA, ARIMA provide good performance in short term load forecasting but fail to map complex temporal patterns in load time series analysis.
- Deep Learning Technique like FFNN, LSTM provide a ray of hope in understanding these complex patterns and also account other factors like weather, humidity into forecasting of load.
- Various state of art techniques of time series forecasting like CNN, RNN or auto-regressive transformers have not been in implemented in load forecasting which can provide great results in big volume of data.
- We aim to prototype these algorithms on real-time dataset of Delhi Load Dispatch Centre to benchmark and understand practical importance in load forecasting by incorporating other factors like weather.

Acknowledgement

- Under guidance:

  Ms. Sobhita Mehar, Assistant Professor, Dept of Electrical Engineering, IIT BHU, Varanasi

- References:

  - Dataset - https://www.delhisldc.org/
  - https://arxiv.org/pdf/1906.04818.pdf
  - https://paperswithcode.com/task/load-forecasting
  - https://arxiv.org/pdf/2007.04517.pdf
  - https://arxiv.org/pdf/2208.00728v1.pdf
  - https://arxiv.org/abs/1906.04818
  - https://caciitg.com/resources/tsa/
  - https://paperswithcode.com/paper/es-drnn-a-hybrid-exponential-smoothing-and
  - https://paperswithcode.com/paper/optimal-adaptive-prediction-intervals-for

Thank You!!