Enhancing Energy Market Operations and Planning by Leveraging Ensembled Models

Author:

GAUTHAMAN KUZHANTHAIVELU ARULKUMAR

Student Number: 22250076

Supervisor:

Dr. FABIANO PALLONETTO

Co-Supervisor:

Dr. SHAHID HUSSAIN

A thesis submitted in fulfilment of the requirements for the degree of MSc in Data Science and Analytics 2022-2023

In the

Department of Computer Science

Maynooth University

Abstract

The energy sector plays a crucial role in the global economy. Efficient operations and planning within energy markets are essential for maintaining reliability, affordability, and sustainability. Energy market operations are complicated by fluctuating energy prices, intermittent renewable energy sources, and limitations in infrastructure. Weather conditions, economic activity, and seasonal fluctuations affect energy demand and supply, making it challenging to maintain a stable grid. Advanced prediction techniques have emerged as valuable tools to address the challenges posed by fluctuating energy demand and electricity prices. This dissertation explores the application of individual and ensembled machine learning models for energy demand and electricity price prediction, to enhance energy market operations and planning. The study comprehensively analyses machine learning algorithms, including SARIMAX, Prophet, and ensemble methods like HistogramGradientBoostingRegressor, RandomForestRegressor, XGBoostRegressor and a Combined Regressor model. To overcome individual limitations like the overfitting of complex data and assumptions of additive seasonality, etc., an ensembled modelling approach is proposed, leveraging the complementary strengths of multiple algorithms. The ensemble models capture complex patterns and dependencies in historical energy data, improving the accuracy and reliability of future predictions. The accuracy and efficiency of these models are evaluated by extensive experiments utilizing real-world energy datasets and evaluation criteria like mean absolute error, root mean squared error, and mean absolute percentage error. It is found that ensembled learning models are more accurate, resilient, and flexible to market conditions than conventional approaches and individual algorithms in energy forecasting. These models can provide insights into forecasting load, supply optimization, risk management, and policy formulation for the market's operation and planning.

Declaration

As part of the MSc - Data Science and Analytics qualification, I hereby certify that the material I present for assessment is entirely my work and has not been taken from the work of others - save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:	Date:
GAUTHAMAN KUZHANTHAIVELU ARULKUMAR	07-08-2023

Acknowledgements

I am deeply grateful to Dr. Fabiano Pallonetto for his patience and feedback. Without his knowledge and expertise, I would not have been able to undertake this journey. I would also like to thank Dr. Shahid Hussain for his guidance and constant support in helping me complete this project.

Table of Contents

1. Introduction

1.1 Motivation

The energy sector is undergoing rapid changes due to various factors, such as increased demand, fluctuating energy sources, and evolving regulatory frameworks. Energy market operators and planners encounter difficulties in running efficient operations, allocating resources optimally, and making sustainable decisions. To tackle these issues, it is crucial to make accurate predictions of energy demand and electricity prices [1][2]. However, traditional energy market forecasting methods rely on statistical models [3], which have limitations in capturing complex non-linear relationships, adapting to changing market dynamics, and accurately predicting demand and price fluctuations. As a result, experts are exploring hybrid machine-learning models [4] that combine the strengths of multiple algorithms to overcome individual limitations and improve prediction accuracy.

Machine learning has become a powerful tool for extracting patterns and making predictions from complex datasets, particularly in energy demand and electricity price prediction. However, individual algorithms often lack the handling of specific data patterns or market dynamics. Combining multiple machine learning algorithms into one can help to address these limitations. This is known as ensemble learning, and it enables the algorithms to work together to better identify patterns and make more accurate predictions than any single algorithm would be able to do [5]. This dissertation aims to contribute to the advancement of energy market operations and planning. It does this by developing and applying combined/hybrid machine-learning models for energy demand and electricity price prediction. By addressing the shortcomings of traditional methods and individual algorithms, these models can produce more accurate and flexible predictions, supporting decision-making processes in the energy sector. The proposed models are tested with real-world data, and the results demonstrate their potential for energy market operations and planning. This enables market operators to optimize resource allocation, minimize costs, and improve system reliability. Bringing theory and practice together to forecast energy markets. Provides valuable insights for industry professionals, policymakers, and researchers seeking to enhance energy market operations and promote sustainable practices [6].

1.2 Time Series Forecasting

Time series forecasting uses statistics and modelling to produce forecasts and guide strategic decisions. This predictive analytics technique takes a historical dataset and uses it to create a model that can be used to predict future values [7]. This forecast can then be used to inform strategic decisions and help organizations plan for the future. For instance, by leveraging time series forecasting, a company may be able to project customer demand for a product over the next few months and adjust its production capacity accordingly. It provides insight into which outcomes are more likely, but it is not always accurate in predicting which outcomes are most probable. Forecasts are more accurate when the data is more complete [8]. It is common practice to use forecasting and prediction in conjunction with time series analysis to identify underlying causes for certain events and explain why these events occur. In addition, time series analysis allows for a more accurate prediction by isolating and identifying the underlying factors. Therefore, it is important to ensure that the data used for forecasting and prediction models is accurate and up to date, to produce reliable predictions. Furthermore, it is important to consider external factors that may influence predictions. To ensure accurate forecasts, these external factors must be considered and incorporated into the models. For instance, when predicting the prices of goods, the model must also consider external economic conditions, such as inflation or unemployment rates, to ensure that the predicted prices are accurate [9].

In many areas of business and research, time series forecasting is a crucial task. Statistical methods and machine learning methods are two co-existing approaches to time series forecasting, each with distinct advantages and disadvantages [10]. By fusing the most advantageous elements of statistics and machine learning, hybrid approaches promise to revolutionize time series forecasting. Although robust and adaptable statistical techniques like Holt-Winters and ARIMA frequently presume linear correlations in the data, this might impair predicting effectiveness. The benefits of cross-series information and universal approximation are provided by machine learning techniques like LSTMs and CNNs. They are, however, constrained by issues like the need for a lot of data and the length of processing. Although machine learning techniques can overcome the limitations of ARIMA, they provide inconsistent results for linear time series, rendering them unsuitable for time series modelling in the real world [11].

As a result, Machine Learning (ML) models enhance forecasting accuracy, flexibility, stability, tolerance to changing conditions, generalization, and managing complicated connections, as well as enhancing interpretability. These models may identify various data patterns or correlations, leading to more reliable and precise predictions. Additionally, they provide more versatility in managing various data kinds and predicting scenarios, which lessens instability or volatility. Hybrid ML models combine several modelling approaches, which can increase prediction accuracy and lower error rates [12]. Furthermore, they can handle big datasets that include complicated correlations and even spot minor patterns that other models would miss.

Furthermore, they are more flexible and can adapt to changing conditions more easily, making them more reliable in a variety of scenarios.

1.3 Project Aim

This thesis explores the effectiveness of ensemble learning based models in energy demand and electricity price forecasting. We aim to investigate whether these models can provide accurate and reliable predictions when applied to real-world datasets compared to conventional time-series models like Sarimax and Prophet. Subsequent sections will review the existing literature on energy demand and electricity price forecasting. We will then train and evaluate these ML models individually. The effectiveness of their forecasting abilities will be determined by comparing their performance metrics with the actual energy demand and electricity price data of Ireland. The results of the evaluations will be used to identify the most accurate ML model. Finally, the best model will be used to accurately predict Ireland's energy demand and electricity prices in the operations and planning by energy market operators.

1.4 Project Structure

The project will be divided into multiple chapters. The following information will be contained in each of them.

- **Chapter 2**: This chapter will provide a synopsis and literature evaluation of the pertinent works. An overview of earlier relevant work and articles will also be included.
- **Chapter 3**: In this chapter, a solution to energy demand and electricity price forecasting will be covered. This part will explore several models chosen for the experiment and choices made during the development and training of the models.
- **Chapter 4**: In this chapter, we examine results from model training on Ireland’s demand and price data and illustrate the performance of selected models through the chosen metrics.
- **Chapter 5**: We will analyze the work done in the preceding sections to conclude the thesis.

2. Background Literature Review

2.1 Literature Review

Energy forecasting is essential for managing power systems, trading, and integrating renewable energy. By identifying complicated patterns, hybrid ML models increase prediction accuracy. Their benefits and drawbacks in energy forecasting are examined in this literature review. The results of this review suggest that hybrid models have the potential to outperform traditional models. Further research is needed to identify the best hybrid model architectures for energy forecasting in different contexts.

Due to their capacity to examine intricate patterns and identify nonlinear relationships in energy data, machine learning (ML) models have attracted a lot of attention in the field of energy forecasting. Planning, decision-making, and resource management in the energy sector may all be enhanced with the help of ML models, which provide useful tools for forecasting energy consumption, prices, and other pertinent characteristics. The background on how ML models are used in energy forecasting is given in this part, along with a discussion of their benefits and potential drawbacks. By using machine learning models, we can identify patterns in data that may otherwise be difficult to detect. They can also be used to make predictions about future energy needs and prices, as well as to identify areas where energy efficiencies can be achieved. By leveraging the power of ML models, energy sector stakeholders will be able to make more informed decisions and better manage their resources.

In a variety of industries, including the residential, commercial, and industrial contexts, ML models have been frequently used to forecast energy demand. In energy demand forecasting applications, conventional ML techniques including linear regression, decision trees, and random forests have been successfully used. For instance, based on historical data, linear regression models have been used to forecast short-term electricity consumption [21]. Additionally successful in capturing complicated correlations between energy consumption and variables like weather, economic indicators, and time of day are decision trees and random forests [14]. Linear regression models are well-suited for short-term forecasting because they can capture linear trends in the data effectively. On the other hand, decision trees and random forests are better at capturing more complicated correlations, like those between energy consumption and weather, economic indicators, and time of day.

Forecasting energy prices is crucial for market participants, traders, and decision-makers. When considering variables like market dynamics, supply-demand imbalances, and price volatility, ML models have shown promise in accurately predicting energy prices. Based on historical pricing data and exogenous variables, support vector regression (SVR) has been used to forecast power prices [22]. To capture nonlinear correlations in pricing data, artificial neural networks (ANNs) have also been used, which has enhanced forecasting precision [23]. Both SVR and ANN have been successfully applied to power price forecasting. ANNs are better equipped to capture nonlinear correlations between different variables in pricing data and have been known to outperform more traditional forecasting methods like support vector regression (SVR).

However, it is still unclear which method works best for different market conditions and more research is needed to answer this question.

Energy forecasting can benefit from ML models' abilities to handle massive data volumes, find nonlinear correlations, and adjust to shifting patterns. This can include meteorological information, economic indicators, and historical trends, improving prediction accuracy. However, problems with data availability, quality, and interpretability still exist, making them challenging to comprehend. Complex patterns can be captured by both deep learning models and conventional statistical methods. However, resolving these problems is essential to improving ML models' precision and dependability in energy forecasting. To achieve successful energy forecasting, data needs to be collected, managed, and analyzed with the help of ML models. High-quality data and interpretable models are essential for reliable energy forecasting. Hybrid models, which combine multiple traditional ML methods, have been suggested as a possible solution to energy forecasting problems. Using these models, complex patterns can be captured with the interpretability of traditional machine learning methods, ensuring that future forecasts are reliable.

ML Models for Energy Demand Forecasting:

Energy demand forecasting activities frequently make use of hybrid machine learning models. To forecast short-term load, Wang et al. [13] suggested a hybrid model that combines wavelet decomposition with support vector regression (SVR), resulting in increased accuracy when compared to individual models. The study showed how non-linear regression methods and data decomposition approaches may effectively capture complicated load patterns.

For forecasting energy consumption, Suresh et al. [14] investigated hybrid models based on regression and machine learning methods. To increase prediction accuracy, they coupled a variety of regression models, such as linear regression and polynomial regression, with machine learning techniques, such as random forests and support vector machines. Their results showed that the hybrid models outperformed the individual models in terms of prediction accuracy, and thus offered a better approach for energy consumption forecasting. In the study conducted by Saravanan and [15], the objective was to predict the daily electricity consumption in an educational institution. They compared ARIMA models to other forecasting techniques for reliable short-term electricity consumption predictions in educational institutions. By fitting historical data, ARIMA models captured patterns and dynamics, enabling accurate predictions and informed decisions on energy usage and resource allocation.

Khalid et al. [16] conducted a study on energy demand forecasting in smart grids, focusing on a hybrid model that combined ARIMA (AutoRegressive Integrated Moving Average) with wavelet transform. By combining the benefits of wavelet transform with ARIMA, the researchers hoped to increase forecast accuracy. A mathematical method called the wavelet transform divides time series into several frequency components, enabling the examination of both high-frequency and low-frequency patterns. The hybrid ARIMA-wavelet model fared better in terms of forecasting energy demand accuracy than conventional ARIMA models. Wavelet transform was used in the hybrid model, which improved prediction accuracy by capturing both short-term variations and long-term trends. However, the wavelet transform is not without its limitations. One such limitation is that it is not able to deal with non-stationary data. This means that the wavelet transform is not suitable for forecasting energy demand in situations where there is a sudden change in the trend. Another limitation of the wavelet transform is that it is computationally intensive, which means that it is not always feasible to use it for large-scale energy demand forecasting.

To anticipate energy demand, the publication [17] introduces a hybrid ensemble model that combines wavelet transform with support vector regression (SVR). The model captures both high-frequency and low-frequency trends by decomposing energy demand time series into frequency components. Each dissected component is subjected to SVR, a machine learning technique, for regression analysis and forecasting future energy consumption. Using actual data, the authors validate the hybrid ensemble model and contrast it with more established SVR and wavelet transform-based models. The findings demonstrate that the hybrid ensemble model beats individual models and offers superior forecasting accuracy for energy demand. These methods work together to collect multi-resolution properties and extract significant information, producing improved results. The hybrid ensemble model can also capture nonlinear patterns in energy demand data more effectively and accurately, making it a reliable and valuable tool for energy demand forecasting. This model could be applied to other forecasting problems in the future.

ML Models for Electricity Price Forecasting:

Hybrid machine learning (ML) models have demonstrated promising outcomes in energy price predictions. For forecasting electricity prices, Aggarwal et al. [18] suggested a hybrid model that incorporates wavelet decomposition and random forests. Their research showed that combining the ensemble learning capabilities of random forests with wavelet decomposition to capture both high-frequency and low-frequency pricing components increases forecasting accuracy. The hybrid model was able to capture the non-linear features of the data that other models could not, resulting in a higher accuracy rate. This proves that hybrid ML models are a powerful tool for energy price predictions.

To estimate short-term demand, Shah et al [19] created a hybrid machine learning model that combines artificial neural networks (ANNs) and regression-based methods. To improve forecasting performance, the hybrid model combined several regression techniques, such as linear regression and autoregressive integrated moving averages (ARIMA), with ANNs' non-linear learning capabilities. The results of their study showed that the hybrid model was more accurate than either ANNs or regression-based methods alone. Furthermore, the hybrid model was able to account for seasonal and trend effects, helping to better capture short-term demand. However, the hybrid model is not perfect. The limitation of this method is that it requires a lot of data to produce accurate results. Another limitation is that it can be difficult to interpret the results of the hybrid model, making it difficult to explain to stakeholders why certain decisions are being made.

Zhang et al [20] created a sliding window method for forecasting day-ahead electricity prices using the SARIMAX model. To forecast electricity costs for the following day, the model takes exogenous factors like weather, demand projections, and economic indices into account. By dividing historical data into many windows using the sliding window technique, the model may be flexible and capture evolving trends over time. The study demonstrated the SARIMAX model's accuracy in predicting electricity costs by contrasting it with other benchmark models. A useful tool for day-ahead electricity price forecasting, the SARIMAX model with a sliding window method enables precise predictions and aids in decision-making in a variety of energy market applications. However, the sliding window technique can also have some drawbacks. One potential drawback is that it can be computationally intensive, which can make it difficult to use for large data sets. Additionally, the sliding window technique can be less accurate for data sets that are non-stationary, meaning that the data set has some trends or patterns that change over time.

For forecasting hourly short-term energy consumption, Zha et al. [39] suggest a hybrid model that combines SARIMA and Prophet. Prophet is a flexible forecasting model created by Facebook, whereas SARIMA is a commonly used model for capturing seasonal patterns and trends in time series data. The combined model, which Prophet manages non-linear growth or decline and incorporates domain-specific knowledge into, attempts to increase the accuracy of short-term energy demand forecasting by capturing seasonal and trend patterns. The SARIMA and Prophet models are trained by fitting them to past data on energy demand, and then the process forecasts energy demand for the coming hours. Since the combined model captures both seasonal patterns and non-linear trends, forecasting accuracy is improved above the performance of the individual SARIMA and Prophet models. However, the combined model is more difficult to interpret than either the SARIMA or Prophet model alone. In addition, the combined model is more computationally expensive, which may be a drawback in some applications.

An ensemble deep learning framework for estimating electric load is suggested in this [35]. To capture the intricate patterns in energy load data, the method integrates several deep learning models, including long short-term memory (LSTM) networks and deep belief networks (DBN). Comparing the ensemble method to individual models, the forecasting accuracy has been demonstrated to be greater. The proposed method is shown to be robust and efficient in estimating the electric load. It has the potential to be applied in electricity power management for better decision-making.

The study [36] focuses on the probabilistic ensemble model forecasting of day-ahead electricity prices. To help market players make better selections, the study examines and assesses the performance of several ensemble strategies, including bagging and random forest. The study results show that the ensemble model can improve the accuracy of electricity price forecasting by up to 20%. The study also suggests that bagging and random forest are the most effective strategies for improving the accuracy of price forecasting.

In the paper [37], a hybrid ensemble learning strategy is presented for forecasting short-term electricity prices. The least squares support vector regression (LSSVR) and random forest models are two that the ensemble model used to increase the reliability and accuracy of price forecasts in electricity markets. The LSSVR model is used to capture the nonlinear relationships and complex interactions between the input features and the output, while the random forest model is used to capture the nonlinear relationships between the input features and the output more accurately and efficiently. The two models are combined to create a hybrid model that is more accurate than either model on its own.

2.2 Model Architecture

2.2.1 SARIMAX Model

SARIMAX is a method of time series modelling that incorporates exogenous variables and includes autoregressive, moving average, and integrated components. It incorporates seasonal patterns in data and is an extension of the ARIMA model [24]. With AR capturing the link between observations and past values and MA considering residual errors, SARIMAX models analyze data for seasonality and trends. The time series may be impacted by exogenous factors like weather information, economic indicators, or policy changes. SARIMAX models can be used to forecast future values or examine the effects of exogenous variables. Mathematically we can represent the model like this,

$$\emptyset_p(L)\tilde{\emptyset}_p(L^s)\Delta^d\Delta_s^D y_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^s)\epsilon_t \tag{1}$$

Where,

- $\emptyset_p(L)$ is the non-seasonal autoregressive lag polynomial
- $\tilde{\emptyset}_p(L^s)$ is the seasonal autoregressive lag polynomial
- $\Delta^d\Delta_s^D y_t$ is the time series, differenced d times, and seasonally differenced D times
- $A(t)$ is the trend polynomial
- $\theta_q(L)$ is the non-seasonal moving average lag polynomial
- $\tilde{\theta}_Q(L^s)$ is the seasonal moving average lag polynomial
- ϵ_t is the error term

2.2.2 Prophet Model

The Facebook Core Data Science team created the time series forecasting model Prophet to solve issues with time series data forecastings, such as seasonality, trend shifts, and outliers [25]. It is a potent tool for predicting future values in time series data since it includes seasonal and trend components as well as numerous forecasting aspects. Multiple seasonalities and regressors are supported by Prophet's seasonality detection and modelling function, which employs a flexible additive model. It also uses a piecewise linear model to represent long-term data changes, allowing for non-linear growth or decline. Using solid statistical methods and imputing missing values, Prophet also manages outliers and missing data.

$$y(t) = g(t) + h(t) + s(t) + \epsilon_t \tag{2}$$

Where,

- y(t) is the additive regressive model
- g(t) is the trend factor
- h(t) is the holiday component
- s(t) is the seasonality component
- ϵ_t is the error term

Ensemble Models

An approach called ensemble modelling [26] mixes various separate models to increase the predictability and accuracy of results. It employs the idea of "wisdom of the crowd," whereby the choice made by a group of models generally performs better than that of a single model. Due to their capacity to identify a variety of patterns and minimize individual biases, ensemble models are common in domains like machine learning and time series forecasting. Bagging, boosting, and stacking are all tactics. The benefits of ensemble modelling include better model stability, less overfitting, and more accurate prediction. Thoughtful model selection and configuration may be necessary, as well as processing resources.

2.2.3 Histogram Gradient Boosting Regressor

With this estimator, any differentiable loss function can be optimized and an additive model is constructed in a forward stage-wise manner [27]. A regression tree is fitted on the negative gradient of the specified loss function at each level. Reducing the number of values for continuous input features can considerably speed up the development of decision trees. This can be accomplished by binning or discretizing values into a certain number of buckets. By doing this, tens of thousands of unique values for each characteristic can be reduced to only a few hundred. The binning of the input data can also be represented by effective data structures, such as histograms, and the algorithm for building trees can be further customized to make efficient use of histograms when building each tree.

2.2.4 Random Forest Regressor

The Random Forest Regressor is a regression ensemble learning technique that harnesses the potential of numerous decision trees to produce more precise predictions [28]. It is a member of the ensemble learning family, which combines many models to enhance resilience and predictive performance. The internal nodes of decision trees, which stand in for tests, the branches for results, and the leaf nodes for target values, serve as the building elements of the Random Forest Regressor. While random feature selection lessens the correlation between trees and enhances generalization, random aggregation (bagging) produces multiple subsets of training data.

2.2.5 XGBoost Regressor

An efficient and accurate machine learning approach for regression tasks is called XGBoost [29]. To increase precision and effectiveness, it makes use of gradient boosting, regularization, tree pruning, and weighted quantile drawing. Overfitting is penalized by XGBoost's regularization terms, whereas overfitting is avoided by tree pruning by bringing down tree complexity. The procedure is made faster and uses less memory thanks to the weighted quantile sketch data structure, which effectively manages data points and computes quantiles.

2.3 Performance Metrics Used

Root Mean Squared Error

Root Mean Squared Error, or RMSE, is a frequently employed metric for assessing the precision of a predictive model, particularly in the context of regression tasks. The average discrepancy between a dataset's actual (observed) values and the anticipated values is measured by RMSE. When measuring prediction errors, RMSE uses the same unit as the target variable. As a result, a lower RMSE number shows that the model's predictions are more accurate and are therefore closer to the actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}} \tag{3}$$

where N is the number of observations, $y(i)$ is the i-th measurement, and $\hat{y}(i)$ is its corresponding prediction.

Mean Squared Error

Mean Squared Error, known as MSE, is a frequently employed metric for assessing the effectiveness of predictive models, notably in regression assignments. An MSE is calculated as the average squared difference between an observed value and an anticipated value for a dataset. The squaring procedure used in MSE causes larger errors to be given more weight, which allows it to assess the average magnitude of the prediction errors. Similar to RMSE, a lower MSE value shows that the model is more accurate because its predictions are closer to the actual values.

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n} \tag{4}$$

where n is the number of observations, $y(i)$ is the i-th measurement, and $\hat{y}(i)$ is its corresponding prediction.

Mean Absolute Error

Mean Absolute Error, or MAE is a commonly used metric to assess how well predictive models perform, particularly when doing regression assignments. It calculates the difference between anticipated values and actual values (observed) in a dataset. Since the absolute differences are employed, MAE calculates the average size of the prediction errors without taking their direction into account. It gives a clear indication of how far the predictions made by the model, on average, differ from the observed values. A lower MAE score means that the model is more accurate because its predictions are more closely matched to the actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

where N is the number of observations, $y(i)$ is the i-th measurement, and $\hat{y}(i)$ is its corresponding prediction.

2.4 Software Used

Python was used for the entire project, along with several external libraries. Numpy and Pandas are standard data science libraries that we use to handle the data efficiently. In addition to Plotly, Seaborn and Matplotlib were used heavily to implement the graphs because of their added interactivity. Sklearn and Prophet were extensively utilized in training the selected models and performing predictions. Finally, model evaluation was done with the help of Scikit-Learn's metrics library, which allowed for quick and easy comparison between the different models.

3. Proposed Models for Energy Market Operations and Planning

3.1 Problem Statement

Energy market operations are complicated by fluctuations in supply and demand, intermittent renewable energy sources, infrastructure limitations, price volatility, and operational efficiency. Due to fluctuating energy demand and supply, which are affected by weather, economic activity, and seasonal fluctuations, maintaining a stable grid is challenging. The ageing of transmission and infrastructure systems may also result in inefficiencies and grid instability. A successful resource planning and management process requires control over fluctuating energy prices. Energy demand forecasting allows efficient resource planning, grid stability, and cost optimization by providing insight into future energy consumption. Accurate demand forecasting is essential for proper resource planning. With accurate data, utility companies can optimize their operations, reduce losses, and increase customer satisfaction. Accurate forecasts allow for effective power generation source dispatch and scheduling, ensuring the proper mix of energy supply is available to fulfil demand at various times. This will also help to meet the growing energy demand and reduce the risk of grid failure. The below figure (Fig.1) adapted from [41] shows a model energy market system and where the energy demand and electricity price forecasting will help in optimizing the operations.

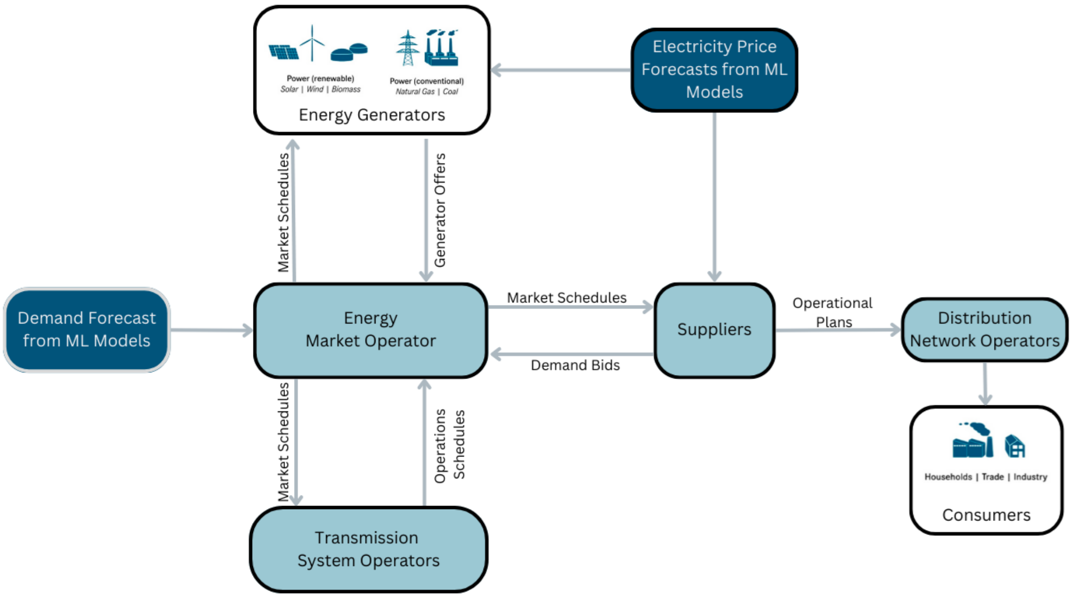

Figure 1 Energy Market Operation Overview

3.2 Approach

As the title indicates, the thesis aims to improve energy market operations and planning by predicting energy demand and day-ahead electricity prices using combined ML models. This chapter explores several modern ML models available for forecasting and critically evaluates their performance. The further sections are split into two phases, the first being modelling energy demand and forecasting future values. The second phase is to model and forecast day-ahead electricity prices in Ireland. The workflow of the thesis process is shown in Fig.2.

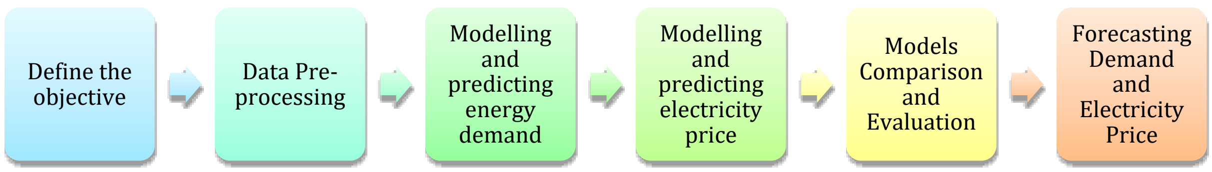

Figure 2 Process workflow design

3.3 Dataset

The open datasets of the Republic of Ireland used in this thesis, are collected in four parts: (1) the EIRGrid [30] provides information on the total energy generated, total demand, wind generation and wind availability for Ireland; (2) Additionally, the EnergyCharts [31] provides hourly data on the various sources from which Ireland generates energy; (3) The electricity wholesale prices dataset is obtained from the ENTSOE Transparency platform [32] which gives hourly market price of electricity in Irish markets; (4) Lastly, the weather information is gathered from the Met Eireann historical data [33]. The raw data is from 2019-01-01 00.00.00 - 2021-12-31 23.00.00 due to data availability. The input features from the collected datasets were critically examined and analyzed using a correlation matrix, box plots and density plots. Several features had to be some features were highly correlated with other variables resulting in duplicates, and the density plots showed zero distribution with constant values. After dropping these extra non-valued features, ten were selected for modelling in this thesis and their statistical overview is listed in Table.1.

Table 1 Overview of features used in the study

S.No.	Variable name	Count	Mean	Standard deviation	Variance	Min	25th Percentile	50th Percentile	75th Percentile	Max	missing %	Count of outliers
1	IE Generation	26304	3354.432	647.883	419752.487	1518.23	2890.750	3290.145	3773.003	5965.7	0.000	180
2	IE Wind Generation	26304	1143.887	832.905	693730.071	0	392.035	1005.695	1812.648	3571.3	0.000	0
3	SNSP	26304	0.362	0.168	0.028	0.0017	0.232	0.337	0.504	0.7428	0.000	0
4	Hydro Run-of-River	26074	90.651	64.084	4106.739	0	31.600	93.300	150.200	218.7	0.874	0
5	Fossil hard coal	26074	175.680	189.641	35963.617	0	0.000	102.700	252.900	761.5	0.874	887
6	Fossil oil	26074	291.051	189.246	35814.222	0	165.100	308.500	443.775	921.9	0.874	9
7	Fossil peat	26074	152.240	106.504	11343.105	0	65.800	112.100	243.400	347	0.874	0
8	Fossil gas	26074	866.854	393.419	154778.464	16.6	542.425	866.500	1167.575	1888.5	0.874	0
9	temp	26304	9.868	4.974	24.744	-5.6	6.400	9.700	13.500	26.3	0.000	38
10	rhum	26304	82.068	11.910	141.839	24	75.000	84.000	91.000	100	0.000	366
11	IE Demand	26304	3399.624	622.503	387509.569	2030.92	2879.953	3445.375	3868.453	5324.89	0.000	0
12	ElectricityPrice	26226	74.615	66.600	4435.601	-41.09	36.113	51.270	85.228	500	0.297	2918

The datasets were consolidated, and the complete data overview was analyzed for missing data and outliers. The missing data in the input features were handled by linear interpolation, thereby filling the gap in the time series. The range of non-outlier data is from Q1 - 1.5*(IQR) to Q3 + 1.5*(IQR) given that the criterion for outliers is 1.5 times IQR (Q3 - Q1). The flooring and ceiling method handled these outliers to make the data more standardized. Furthermore, four features were extracted from the original date time series: ' day', 'month', 'hour', and 'day of the week'. The final cleaned dataset had 14 features and 26304 records observed in hourly frequency. The last two rows ‘IE Demand’ and ‘ElectricityPrice’ in Table 1 are the target variables. The time series of the target variables at hourly intervals are plotted in Figure 3, which shows a slightly increasing trend and yearly seasonal pattern in the energy demand. While the electricity price is seen to be on a highly increasing trend with daily seasonality as well as some noise in the series.

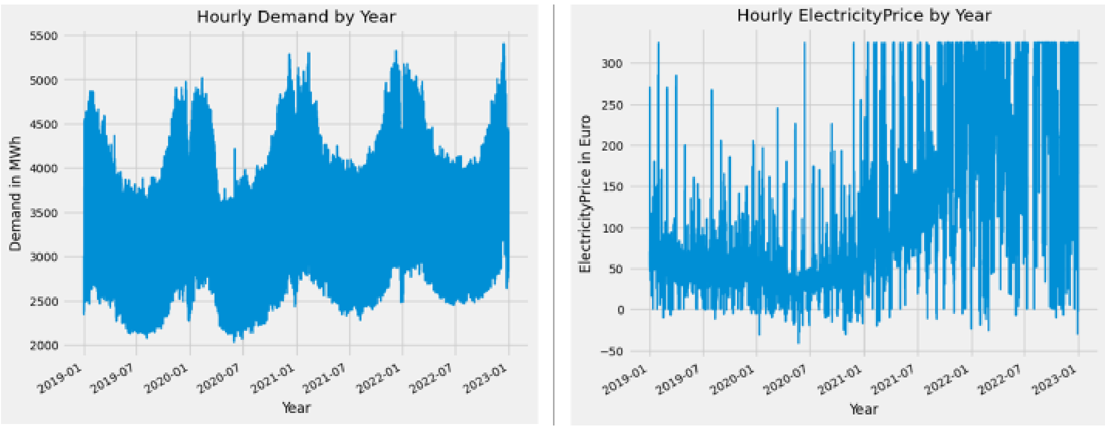

Figure 3 Time-series plot for Energy Demand and Electricity Price

The hourly demand and electricity price were analysed for yearly and daily trends and patterns. From Fig.4, it is evident that there are clear seasonal patterns in the total demand and electricity price. The demand is higher during the evening at 1600hrs while the lowest demand is recorded during the nighttime till early morning and gradually rises correlating to the working/business hours. The Electricity price is also in sync with the demand as the price is high during the daytime and less during the night.

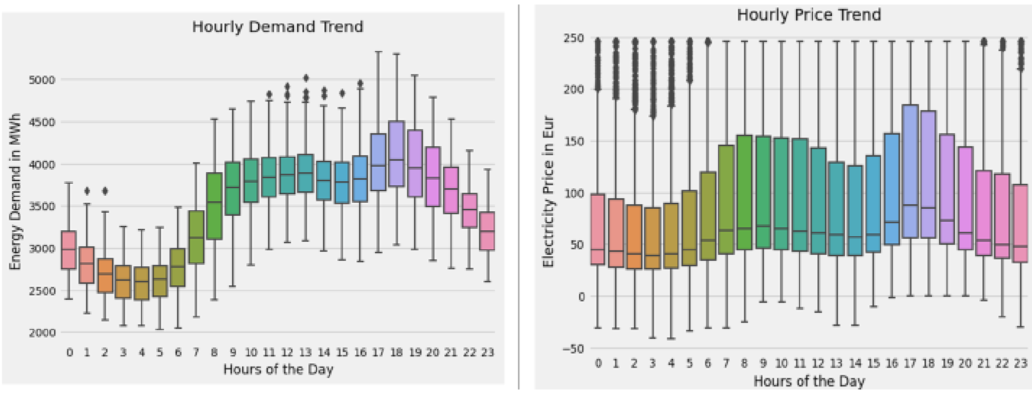

Figure 4 Hourly distribution of the target variables

The analysis of the monthly distribution of energy demand (Fig.5) shows a clear pattern of less energy demand during the summer due to climate controls like less need of using heaters [34]. It is also evident that the energy demands are higher during the winter months reaching the highest level over the year. Similarly, the electricity market prices are lesser during summer as there is less demand and reaches higher rates during winter in correlation with the demand trend.

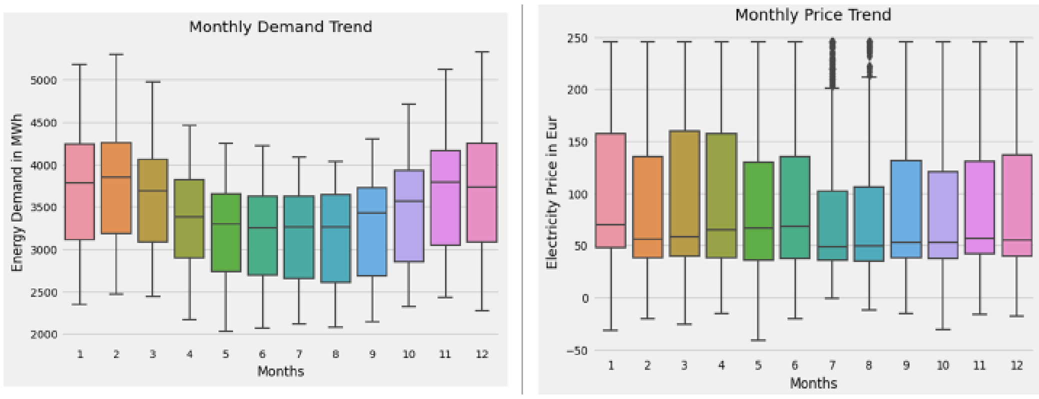

Figure 5 Monthly distribution of the target variables

3.4 Conceptual Design

3.4.1 Energy Demand Prediction

The first phase of our project is the prediction of energy demand in Ireland for a day ahead and a week ahead. The flowchart (Fig.6) describes this process on a higher level, describing the step-by-step approach. The selected models will be trained to predict energy demand and tuned to minimize the errors and the best resulting fit will be compared to other models evaluated by several performance metrics like RMSE, MAE, MSE, and MAPE. The best model will be used for demand forecasting.

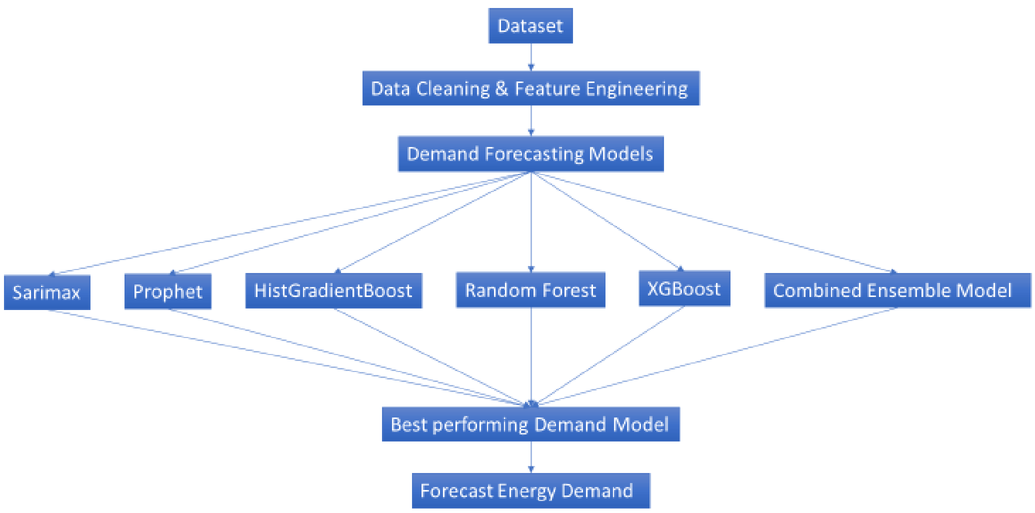

Figure 6 Modelling Approach for Energy Demand

3.4.2 Electricity Price Prediction

The second phase of our project is to predict electricity market prices for a day ahead and a week ahead. Here, the energy demand forecasted from the previous phase is added as a predictor along with the original predictors (Fig.7). The models are then trained and tuned separately and compared to evaluate the best-performing model with the selected metrics. The best-fitted model is then selected to predict the electricity market price.

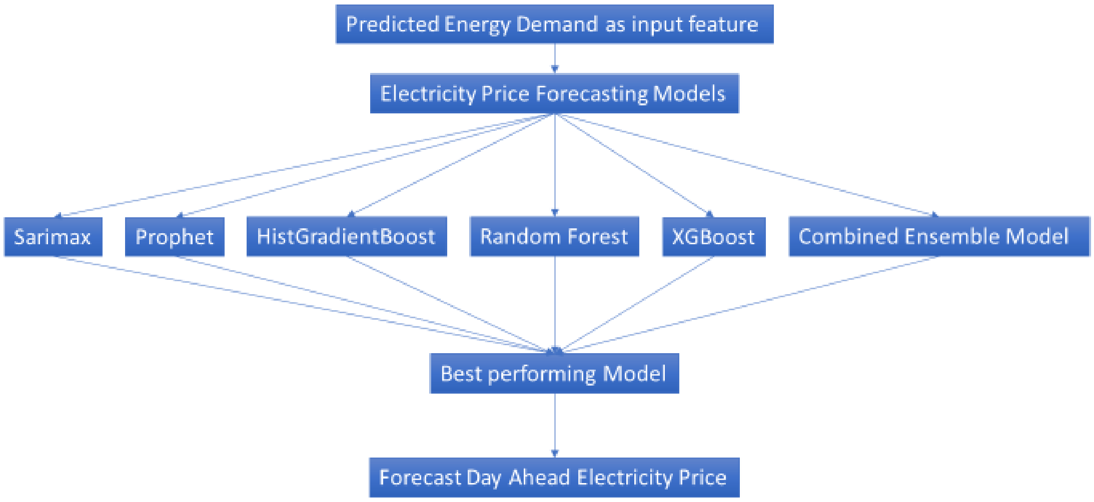

Figure 7 Modelling Approach for Electricity Price

3.5 Experimental Setup

The final cleaned dataset is then split into training and test set, where 80% of the data is split for the training set. Firstly, the selected models were trained on the original training set with all 14 predictors to forecast the energy demand. The models were then evaluated on the test set by predicting the target variables and the performance metrics were calculated for each model. Each of the individual models was tuned according to their specific hyperparameters to reduce the error and the models are refitted to reevaluate the performance. These are then compared and the best-fitted model is selected to predict the demand for the next day and a week ahead demand. Secondly, another set of models was trained along with the demand feature to forecast electricity prices as the process of model selection is repeated. The actual model outputs and results are explained in the next section.

3.5.1 Model 1: Sarimax

The Statsmodel API library for Python was used to train and tune the SARIMAX model. SARIMAX is represented by the notation SARIMA(p, d, q)(P, D, Q, s), where p is the order of the AR component, d denotes the differencing necessary to make the series stable, q denotes the moving average component, and P, D, and Q denotes the seasonal order component. While fitting this model to the training data, all these seven hyperparameters must be tuned to get the best possible fit with the least errors for the input data. This was done by performing a grid search for the hyperparameters, which fit multiple models in iteration and pick the best fit based on the performance metric chosen. In this case, the Mean Squared Error (MSE) was used to decide the best fit.

3.5.2 Model 2: Prophet

Prophet is a forecasting tool created by Facebook's Core Data Science team that aims to produce precise time series predictions with the least amount of configuration. Forecasting data with significant seasonal trends and numerous sources of uncertainty benefits, particularly from it. The time series data should be prepared in a specified format with two columns: "ds" (the timestamp) and "y" (the observed value) and additional features must be passed as exogenous features. Prophet is resistant to missing data and outliers since it estimates the model parameters using a Bayesian technique. The model's built-in imputation algorithm automatically fills in missing values and can handle time series with irregularly spaced data points. The model uses MCMC (Markov Chain Monte Carlo) sampling to estimate the model parameters and their uncertainties. The 'changepoint_prior_scale' and 'seasonality_prior_scale' are the two hyperparameters tuned using a grid search for reducing the model error. The final model is then used to forecast for several periods (days, weeks, or months) we want to anticipate in the future.

3.5.3 Model 3: HistGradientBoosting

HistGradientBoostingRegressor is a model available in sklearn.ensemble library for Python. This model accepts inputs for the target variable and other exogenous features. In our implementation, 'skforecast' a library which incorporates the sklearn's function for this model is used. This provides an easy and simpler way to model and configure the hyperparameters. The 'learning rate' and 'lags' were adjusted to reduce the model error by performing a grid search of all possible combinations.

3.5.4 Model 4: Random Forest Regressor

Similar to model 3, the skforecast's regressor was used with the random forest regressor from the sklearn library. The number of estimators and the lags to be used are the parameters that were tuned in this model. The effective parameters after grid search were found to be 100 estimators and lagged values of 50 resulted in the best fit.

3.5.5 Model 5: XGBoost Regressor

The XGBRegressor model from sklearn was trained using the skforecast regressor which is based on an ensembled learning technique. In this model, the learning rate was also tuned along with the number of estimators and lags. After training, the model was used to predict future values of the dataset. The results showed a high accuracy and a low root mean square error. This indicates that the model was able to capture the patterns in the data effectively.

3.5.6 Model 6: Combined Regressor

The combined regressor model is an ensemble estimator that fits several base regressors, each on the whole dataset [38]. An average of the individual predictions is calculated to form a final prediction. The Histogram Gradient Boosting, Random Forest, and XGBoost are the regressors fitted on the whole dataset and are then used to create this combined regressor model. To perform this task, sklearn provides a special algorithm called Voting Regressor, which gets the individually fitted models as input and calculates the average prediction of each model. This process ensures that the final prediction captures the best of all three ensemble models as each model has its pros and cons.

4. Results and Discussion

4.1 Energy Demand Models

The seasonal decomposing of the energy demand data in Fig. shows a clear sign of an increasing trend from the start of the year 2020. The decomposition plot (Fig.8) also clarifies the seasonality of the data which follows a yearly pattern.

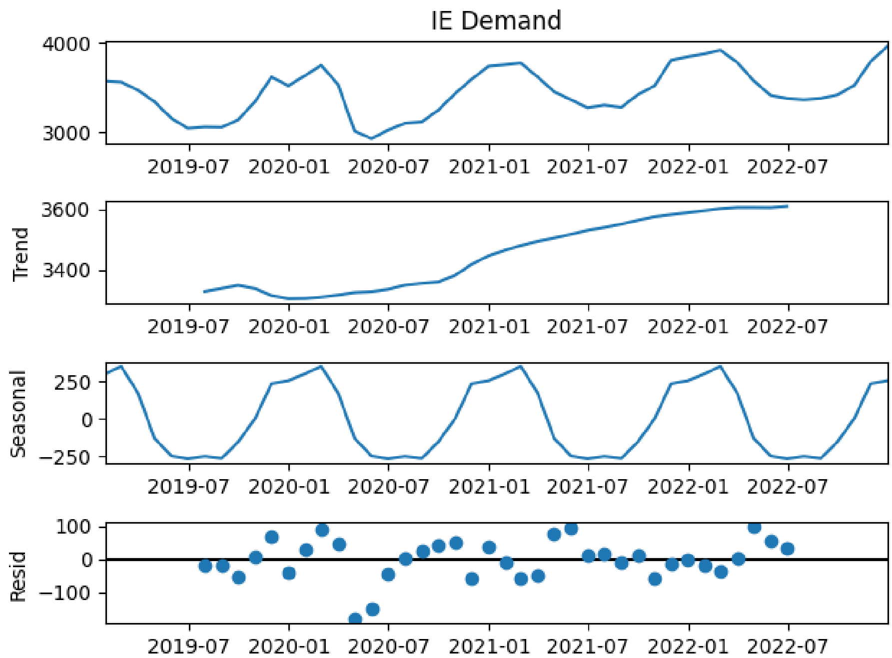

Figure 8 Seasonal Decomposition of Monthly Energy Demand

The performance metrics for all the models trained on the energy demand prediction are listed in Table 2. XGBoost and Combined Ensemble models are the better models according to the study results for our data.

Table 2 Evaluation Metrics of Energy Demand Models

Model	RMSE	MSE	MAE	R2
Sarimax	286.20	81915.86	221.71	0.787
Prophet	174.59	30483.47	139.55	0.920
Histogram Gradient Boosting	156.95	24634.07	125.27	0.936
RandomForest	220.99	48838.24	187.07	0.873
XGBoost	154.76	23950.80	123.46	0.937
Combined Regressor	159.46	25428.85	130.79	0.933

The Sarimax model was the first model fitted for the demand prediction data. The final tuned model (Fig.9) was not a great fit on the test data compared to the other models selected for our study, as it was not able to capture all the low points in the data, which resulted in an under-fitted model. The complex patterns in the data proved to be difficult to produce a better-fitted model. Further exploration and fine-tuning might be needed to improve this model for forecasting energy demand.

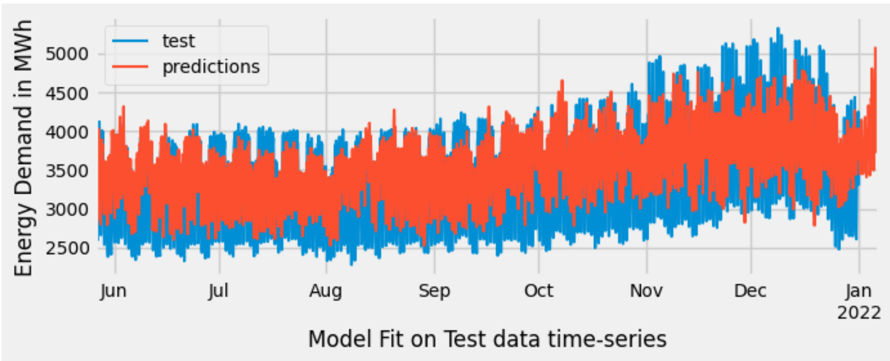

Figure 9 Energy Demand Test data Fit - Sarimax Model

The Prophet model resulted in a much better fit (Fig.10) than the conventional Sarimax model, as the variance in the data is captured well. The performance metrics also show that the model has much lower error.

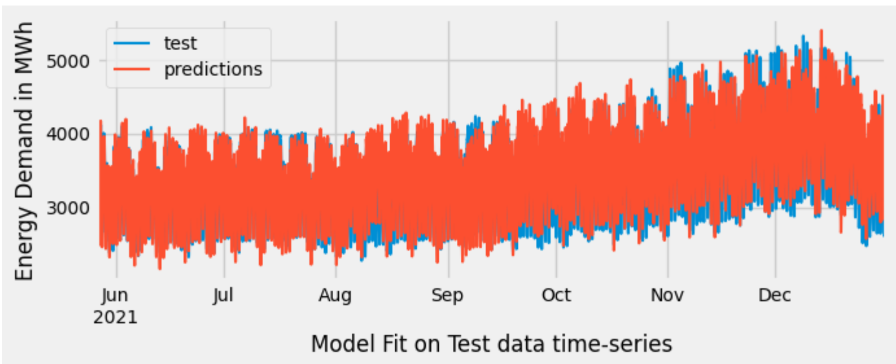

Figure 10 Energy Demand Test data Fit - Prophet Model

The Prophet model also provides additional decomposed time series (Fig.11), which shows clear seasonality and trends in the data on weekly, yearly and daily scales. The results confirm the original observations of energy demand being lesser on weekends and during summer months while reaching high demands on workdays and during winter.

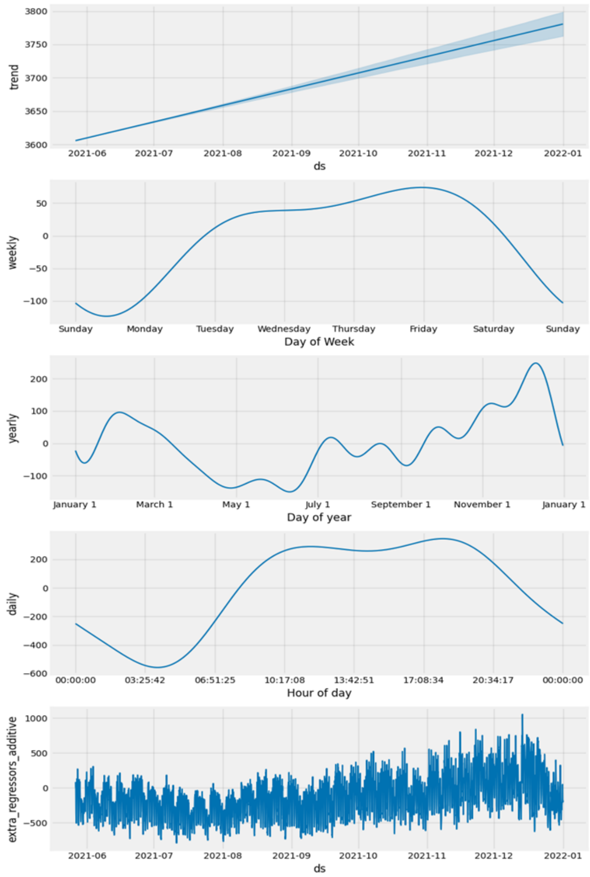

Figure 11 Decomposed trends for the Energy Demand variable

The Histogram Gradient Boosting regressor is an ensemble model, that is trained by building multiple weak learners. The residuals from the weak learners are then added to the previous prediction for the next iteration. This process is continued until the stopping criterion, which is usually based on the number of boosting rounds. The final model (Fig.12) has better performance than the Prophet and Sarimax models.

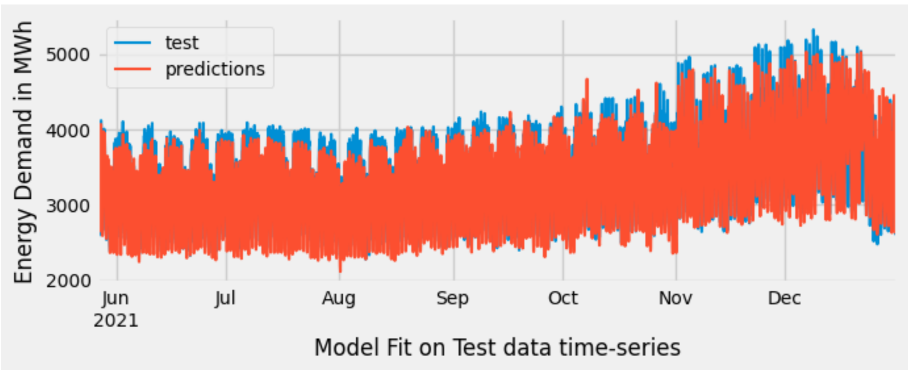

Figure 12 Energy Demand Test data Fit - Histogram Gradient Boosting Model

A Random Forest regressor creates multiple bootstrap samples at random from the original dataset to create an ensemble model. Each bootstrap sample is given a decision tree, which recursively divides the data into subsets with similar target values. Each decision tree, after being created, predicts the target value for a specific input, which is then added together to get the final prediction. The final prediction is based on the average of the projected values. From the results (Fig.13), we see that this model is the poor-performing model out of all the ensemble models. The fitted model was always falling short on the peaks in the testing dataset, leading to an under-fitted model.

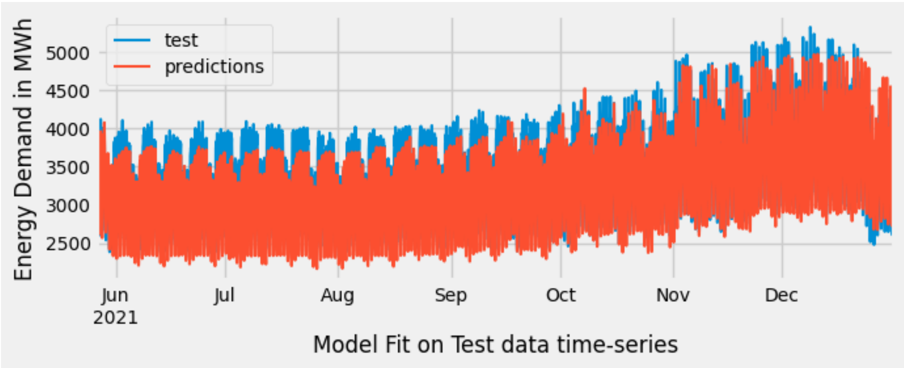

Figure 13 Energy Demand Test data Fit - Random Forest Regressor Model

The XGBoost Regressor is found to the be best-performing model for our dataset as observed from the results table. Since we have a lot of non-linear relationships between the predictor and target variable, this increases the complexity which is handled very well by the XGBoost regressor. This model also handles any missing data internally and has regularization techniques making the model more robust by avoiding overfitting and reducing the risk of reading noise in the data. This results in a better-fitted model (Fig.14) that is more accurate in our complex cases with many non-linear features. The results from Table.2 confirm that this model has the lowest error based on the performance metrics.

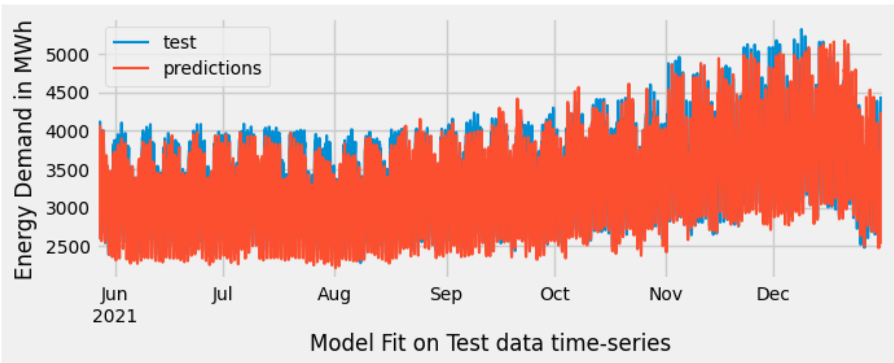

Figure 14 Energy Demand Test data Fit - XGBoosting Regressor Model

The final combined ensemble model brings the best of all the selected ensemble models. This results in creating a fine-tuned model which can produce better models by taking an average of the predictions from all other models. The fit (Fig.15) seems to have captured the seasonal pattern of energy demand perfectly compared to the individual models.

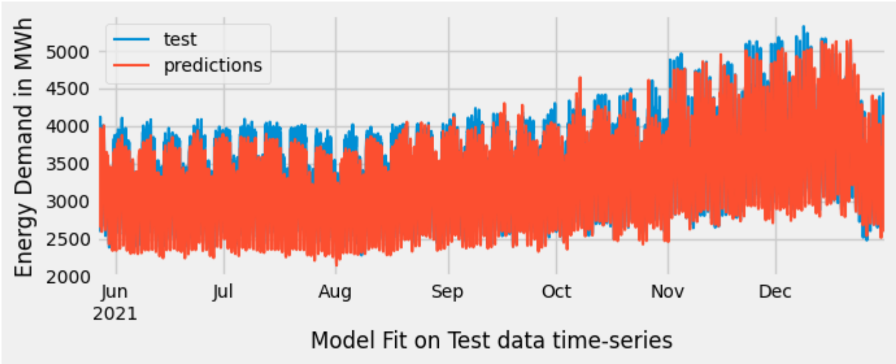

Figure 15 Energy Demand Test data Fit – Combined Regressor Model

Evaluation:

Comparing all the selected models for energy demand forecasting, the XGBoost regressor model which is an ensembled learning model has the best fit and performance. This model has the minimum MSE and an R2 score of more than 93.7% explaining more variation of the dependent variables than the other models. The Sarimax time-series model has the highest MSE values, which is significant from the other models and has a comparatively poor fit. From the R2 score, we see that only 78% of the variance is explained by dependent features, the lowest of all other models studied. The Random Forest regressor has the second highest MSE and was found to be a bit underfit, due to the complexity of the patterns in the data. Both the Ensembled Boosting models (HistGradientBoosting & XGBoosting) have performed well in terms of explaining the variability one percent more than the Prophet model and also have a lower RMSE. The combined regressor model is found to be the third best model which outperforms Random Forest and Prophet models with an RMSE score of 159 and R2 score of 93.3%. The day-ahead and week-ahead forecast of energy demand from all the models after the testing data is plotted in Fig.16. & Fig.17 respectively. The evaluation metrics for the demand models are plotted in Fig.18.

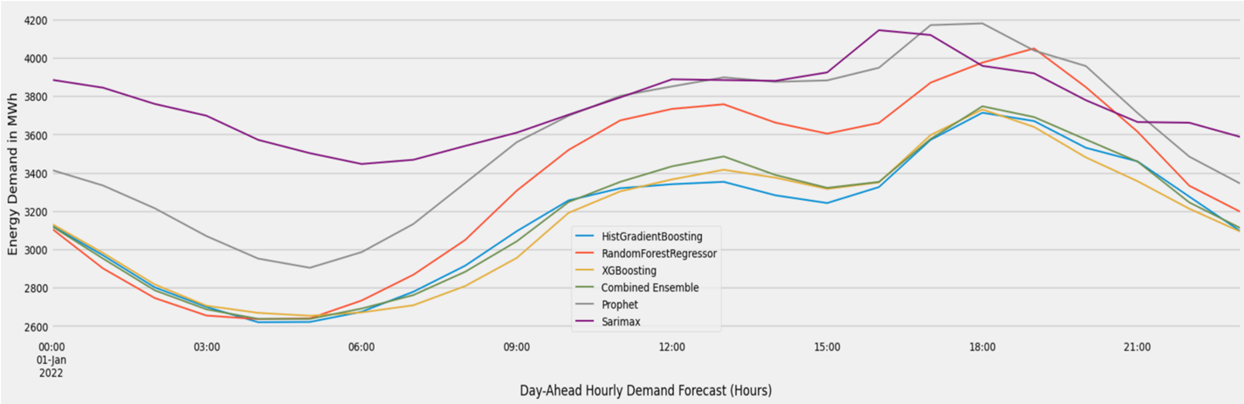

Figure 16 Day-ahead Prediction for Energy Demand

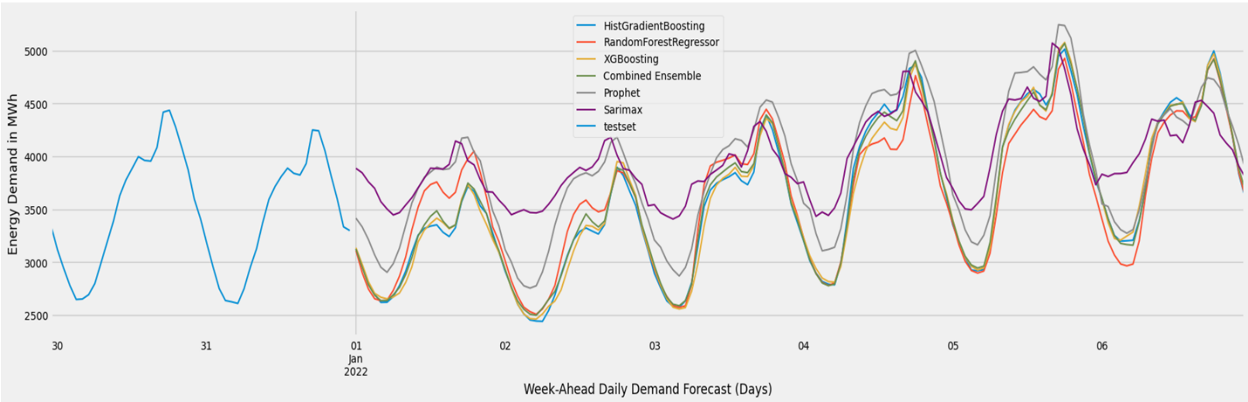

Figure 17 Week-Ahead Prediction for Energy Demand

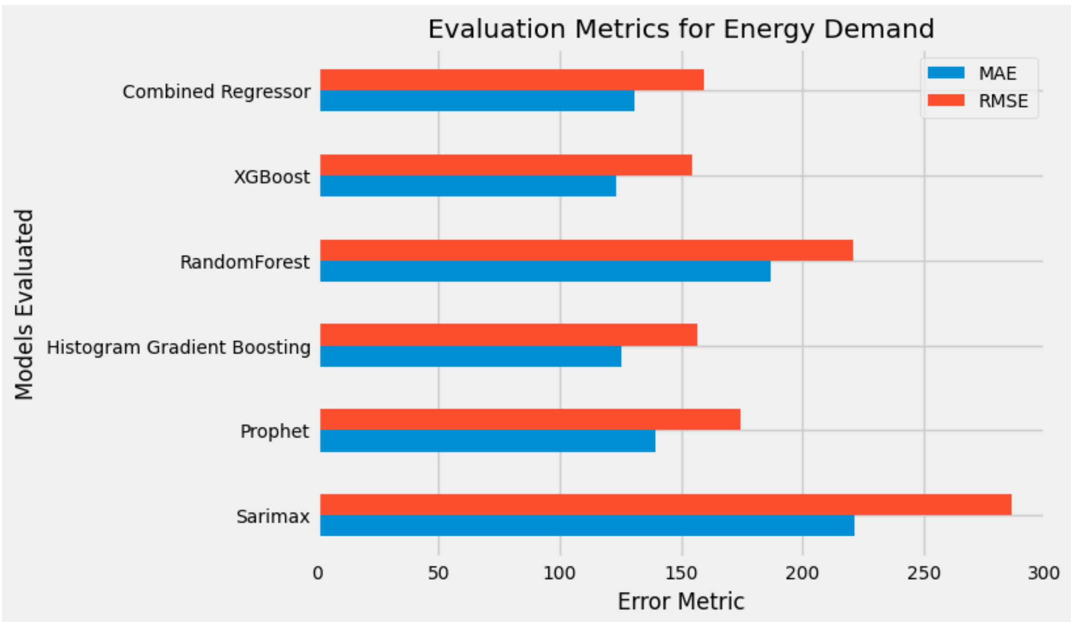

Figure 18 Comparison of Evaluation metrics for Demand models

4.2 Electricity Price Models

A similar approach of modelling and training is taken for the forecasting of electricity market prices, with the energy demand added as an additional predictor. The seasonal decomposition plot (Fig.19) of the electricity price data indicates a clear increase in the trend from the mid of 2020. There is also a seasonal component which follows a yearly cycle.

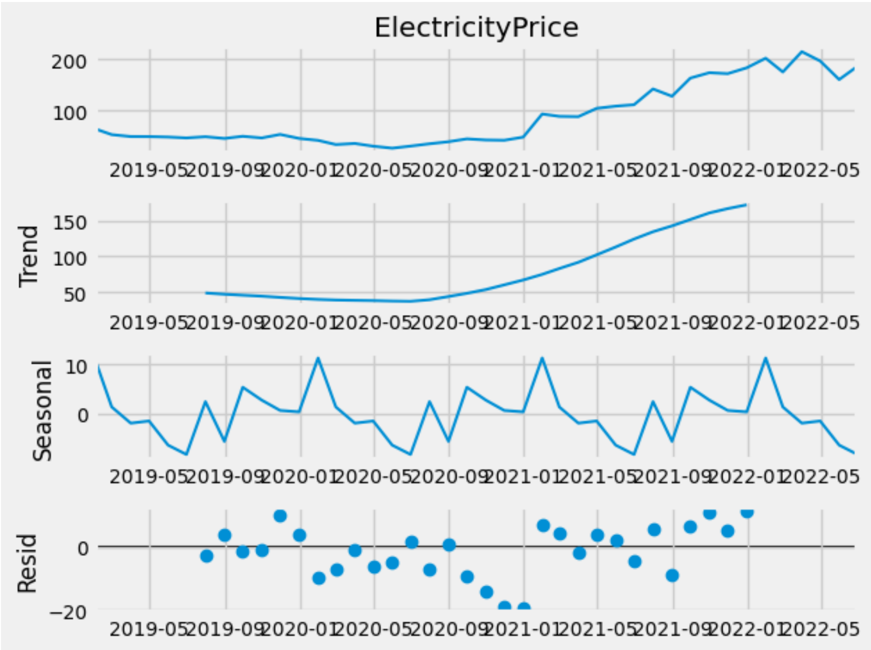

Figure 19 Seasonal Decomposition of Monthly Electricity Price Variable

The performance metrics for all the models trained on the electricity market price prediction are listed in Table 3. Prophet and Sarimax models are the better models according to the study results for our data.

Table 3 Evaluation Results of Electricity Price Models

Model	RMSE	MSE	MAE	R2
Sarimax	59.80	3576.19	48.85	0.136
Prophet	53.59	2871.96	41.55	0.306
Histogram Gradient Boosting	67.85	4604.60	57.72	-0.111
RandomForest	76.92	5917.58	64.35	-0.428
XGBoost	75.93	5765.92	63.33	-0.391
Combined Regressor	64.35	4141.77	55.31	0.002

The time-series model Sarimax has effectively captured the seasonal patterns and trends of electricity prices in the dataset (Fig.20). There was a lot of noise in the data with inconsistent patterns, the Sarimax model handled those effectively and produced accurate results. However, an R2 score of only 13% shows that very little variance of the dependent features is explained by this model.

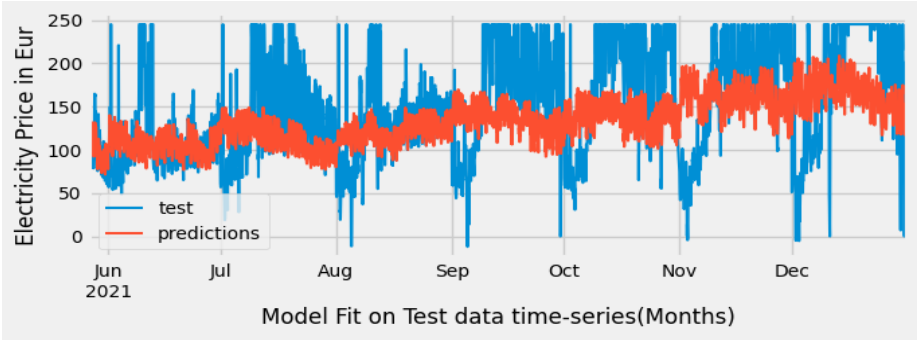

Figure 20 Electricity Price Test data Fit - Sarimax Model

The Prophet model (Fig.21) on the other hand has been even better than the Sarimax model. The Seasonal pattern is more accurate and due to its robustness, the output is more stationary. Noises in the data are handled much better and the seasonal components are better identified in this model. The performance metrics on our study show Prophet as the best model out of the study with the least MSE and RMSE. It also has a higher R2 score than all the models, with 30% which can be improved further with more detailed analysis and experimentation.

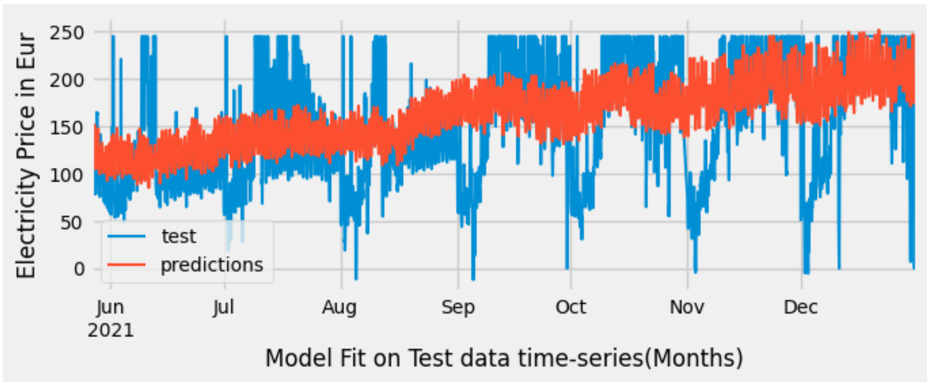

Figure 21 Electricity Price Test data Fit - Prophet Model

The decomposed series output (Fig.22) shows the weekly trends where the price is lower on Thursdays through Fridays and higher during the start of the week on Mondays. The daily seasonality reveals that the market electricity prices drop in the afternoon to the lowest point and peak in the early morning.

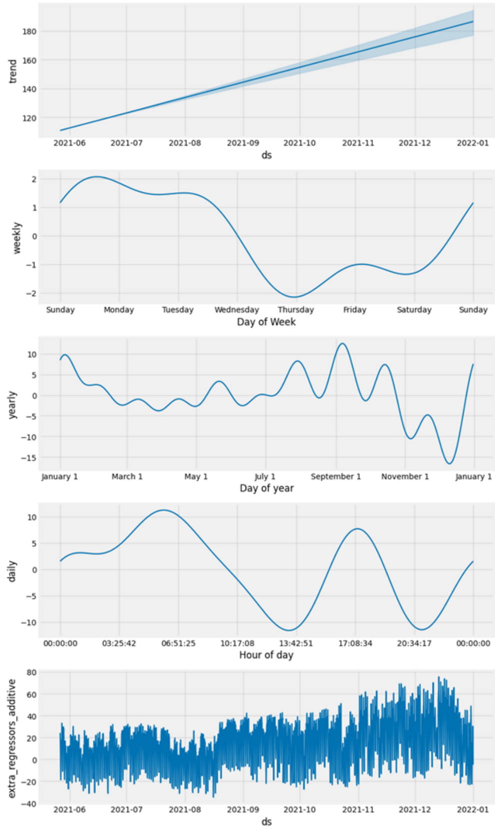

Figure 22 Decomposed trends for Electricity Price variable

The histogram gradient boosting regressor model for electricity price prediction was able to capture the seasonality pattern in the data quite well. However, it failed to capture the increasing trend in the long-term prediction. From Fig.23, we can see that the predictions were off from the test data towards the end of the forecast. The model is under-fitted and is mostly influenced by noise. Although, this model seemed to be the best-performing model compared to the other ensemble models studied.

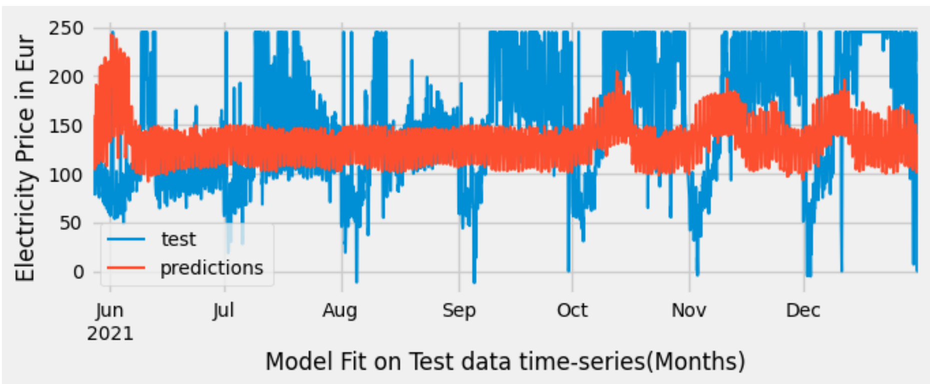

Figure 23 Electricity Price Test data Fit - Histogram Gradient Boosting Model

Another ensemble model used for our study is the Random Forest Regressor (Fig.24), which has a slightly higher MAE than the histogram gradient boosting model. The model was performing poorly on the test data, as the predictions were found to be off from actual values and no trend has been captured by the model. This was the least performing model with higher MSE and RMSE compared to all other individual models studied. Too much noise in the data has resulted in a poor fit in the decision trees causing the fit to perform poorer than the average predictions. This explains the negative R2 score value, showing a poor fit. This can be rectified by further analysis of the external features and fine-tuning the hyperparameters.

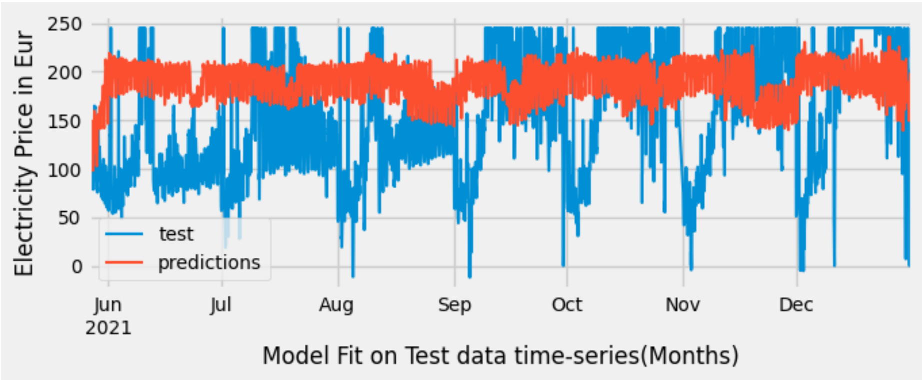

Figure 24 Electricity Price Test data Fit - Random Forest Regressor Model

The XGBoost regressor model fitted (Fig.25) on the electricity price data was slightly better in comparison to the Random Forest regressor. Though this model has captured the seasonality in the data it fails to handle the increasing trend, showing that this model might not be right for our data. The predictions on the test data had too much noise and didn't have standard variance.

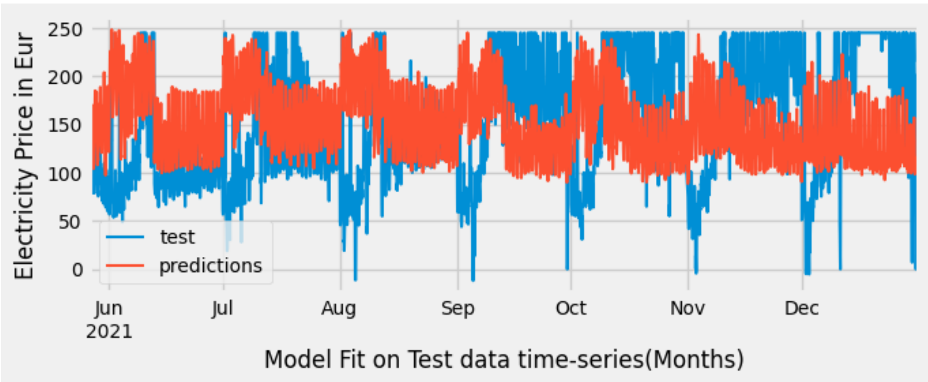

Figure 25 Electricity Price Test data Fit - XGBoost Regressor Model

The Combined ensemble model for electricity price was the average of other ensemble models. This model was found to be slightly better than all the other ensemble models, though the overall fit was not better than Sarimax and Prophet models. This seasonality was not properly captured in the model fit as it only shows a linear trend pattern (Fig.26).

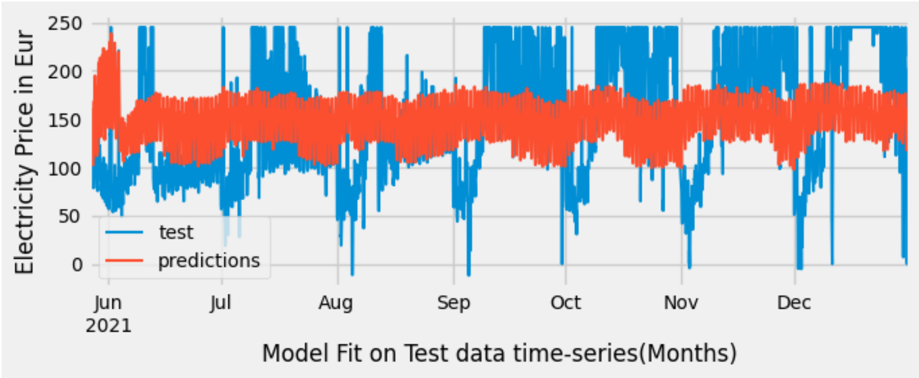

Figure 26 Electricity Price Test data Fit - Combined Regressor Model

Evaluation:

Comparing all the models for electricity market price prediction, the individual time-series models Sarimax and Prophet provided positive results for a weekly prediction (Table.3). The Prophet model has the lowest error metrics with an RMSE value of 53.5 and R2 score of 30.6% showing the best results compared with other models studied. While the ensembled regression models have higher RMSE and MAE values compared to the Prophet and Sarimax models. The negative R2 scores in the ensembled based learning models indicate that these model fits are poorer than the basic average model for the electricity price data. The day-ahead predictions and week-ahead predictions of electricity prices of all the models are displayed in Fig.27 & Fig.28 respectively. The evaluation metrics for the price models are plotted in Fig.29.

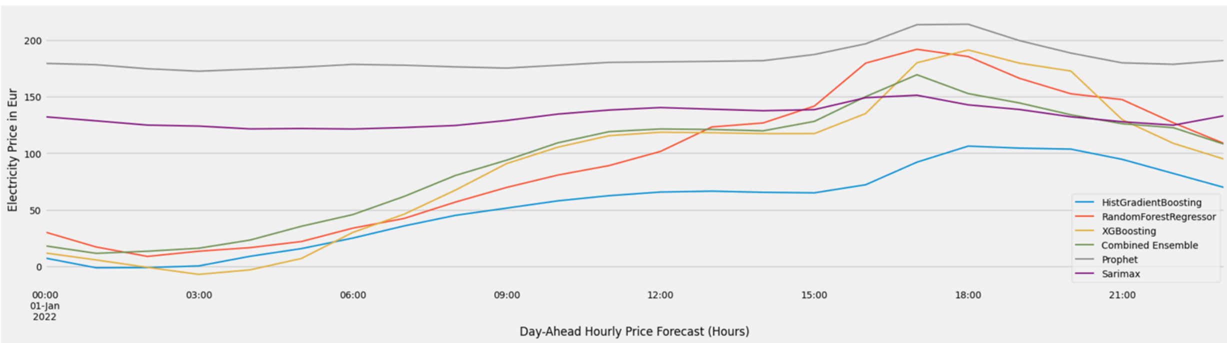

Figure 27 Day-ahead Prediction for Electricity Price

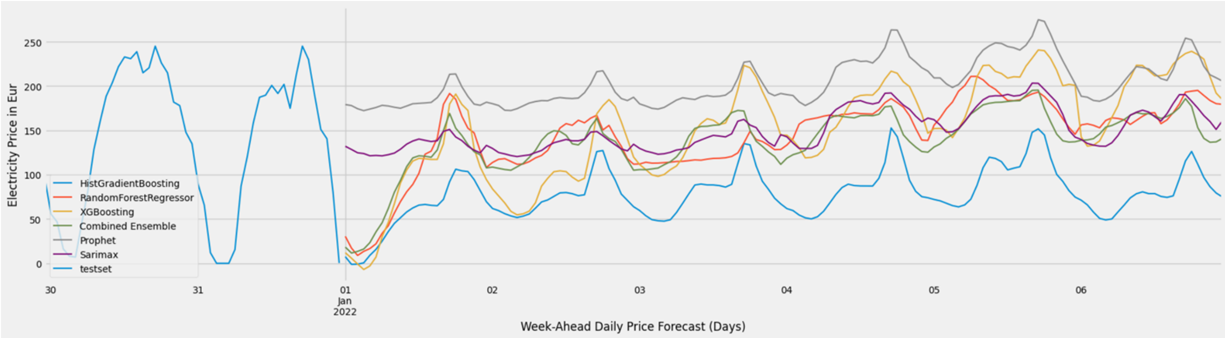

Figure 28 Week-ahead Prediction for Electricity Price

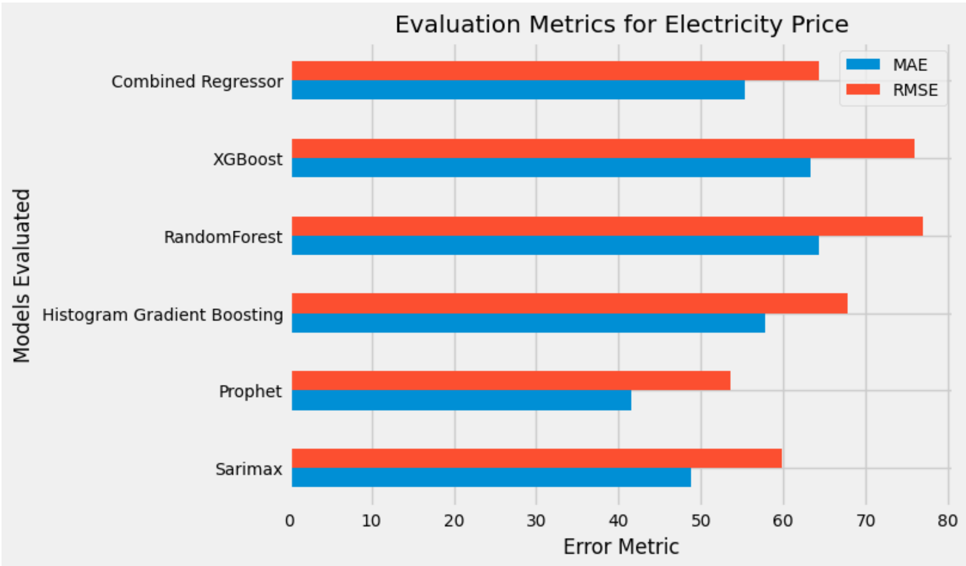

Figure 29 Comparison of Evaluation metrics for Price models

5. Conclusion

Modern energy planning operations need to be able to predict electricity demand and prices, which in turn allows better resource allocation and cost reduction. They also aid in load balancing and grid stability, which lessens the chance of blackouts brought on by supply and demand imbalances. To ensure the success of modern energy planning operations, careful consideration must be given to predicting electricity demand and prices, load balancing, and grid stability - all of which will lead to improved customer service, cost savings, and reliable, efficient, and cost-effective operations. In this study, we examined various time series and ensemble-based regression models for energy demand and wholesale electricity market price prediction. From the experimental results of the data collected for Ireland, the ensemble models were found to be providing more accurate predictions in the case of Energy demand for short-term load forecasting. The XGBoost ensemble regressor model was found to be best compared to other ensemble learning and time-series based models. In the case of the electricity price prediction, it is observed the Prophet model provides the best output. This model has been able to capture both linear and non-linear trends seamlessly resulting in better results in comparison with other models studied. The electricity price prediction might be further improved by considering external economic conditions, such as inflation or unemployment rates, to ensure that the predicted prices are accurate [9].

Energy forecasting is constrained and challenged by uncertainties, complexity, access to and quality of data, and rapid changes in the energy landscape. Despite these challenges, advances in data analytics, machine learning, and a deeper comprehension of energy dynamics have led to continued progress in energy forecasting. Further feature analysis and variable regularization might help to address these issues and in turn capture complex patterns of electricity price.

What's Next?

The electricity price prediction could still be improved in terms of accuracy. The Prophet model could be explored further by analyzing the input features and other regularization techniques. Additionally, the possibility of other factors that may be contributing to the electricity market price can be explored. Techniques like data augmentation as proposed by Bandara,K [40], using Global Forecasting Models(GFM) can be leveraged to improve accuracy of the models for even better energy demand and electricity price prediction.

Appendix-A

A.1 Code

GitHub Repository

https://github.com/Gauthaman-KA/ElectricityPrice-DemandForecasting

References

[1] M Shahidehpour, Hatim Yamin and Li, Z. (2002). Market operations in electric power systems: forecasting, scheduling, and risk management. New York: Institute Of Electrical And Electronics Engineers, Wiley-Interscience.

[2] Bunn, D.W. (2004). Modelling Prices in Competitive Electricity Markets. John Wiley & Sons.

[3] Debnath, K.B. and Mourshed, M., 2018. Forecasting methods in energy planning models. Renewable and Sustainable Energy Reviews, 88, pp.297-325.

[4] Syed, D., Abu-Rub, H., Ghrayeb, A. and Refaat, S.S., 2021. Household-level energy forecasting in smart buildings using a novel hybrid deep learning model. IEEE Access, 9, pp.33498-33511.

[5] Bühlmann, P., 2012. Bagging, boosting and ensemble methods. Handbook of computational statistics: Concepts and methods, pp.985-1022.

[6] Ascher, W. and Ascher, W., 1978. Forecasting: An appraisal for policy-makers and planners (Vol. 314). Baltimore: Johns Hopkins University Press.

[7] Webby, R. and O'Connor, M., 1996. Judgemental and statistical time series forecasting: a review of the literature. International Journal of Forecasting, 12(1), pp.91-118.

[8]Makridakis, S. and Wheelwright, S.C., 1977. Forecasting: issues & challenges for marketing management. Journal of Marketing, 41(4), pp.24-38.

[9] Habib, M.D. and Sarwar, S., 2013. Impact of foreign direct investment on employment level in Pakistan: A time series analysis. JL Pol'y & Globalization, 10, p.46.

[10] Wei, W., Ramalho, O., Malingre, L., Sivanantham, S., Little, J.C. and Mandin, C., 2019. Machine learning and statistical models for predicting indoor air quality. Indoor Air, 29(5), pp.704-726.

[11] Khandelwal, I., Adhikari, R., and Verma, G. (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. Procedia Computer Science, 48(1):173-179.

[12] Dong, B., Li, Z., Rahman, S.M. and Vega, R., 2016. A hybrid model approach for forecasting future residential electricity consumption. Energy and Buildings, 117, pp.341-351.

[13] Wang, Z., Zhang, J., & Guan, X. (2016). Hybrid model combining wavelet decomposition and support vector regression for short-term load forecasting. Energy, 109, 164-174.

[14] Suresh, K., Sudhakaran, G. B., & Babu, S. (2017). Energy demand forecasting using hybrid models based on regression and machine learning techniques. Energy, 135, 672-686.

[15] Saravanan, M., & Rajendran, P. (2015). Daily electricity consumption prediction using time series techniques. In 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-6). IEEE.

[16] Khalid, Z., Arshad, M. N., Arshad, A., Naeem, M. A., Khan, N., & Naqvi, R. A. (2021). Energy demand forecasting in smart grids using hybrid ARIMA-wavelet model. Energies, 14(7), 1863.

[17] Guo, H., Shi, H., Yang, L., & Dong, Z. (2018). Energy demand forecasting using a hybrid ensemble model based on wavelet transform and support vector regression. Energy, 144, 1047-1060.

[18] Aggarwal, A., Mittal, A., & Verma, A. (2018). A hybrid model based on wavelet decomposition and random forests for electricity load forecasting. Applied Energy, 228, 2070-2083.

[19] Shah, N., Dileep, M., & Bandyopadhyay, S. (2019). The hybrid machine learning model for short-term load forecasting. Electric Power Systems Research, 168, 101-110.

[20] Zhang, Y., Lin, H., Ji, B., & Xu, Y. (2019). Day-ahead electricity price forecasting based on SARIMAX model with a sliding window. In 2019 IEEE International Conference on Energy Internet (ICEI) (pp. 294-299). IEEE.

[21] Hong, T., Yan, D., Lee, S., & Shen, P. (2014). Short-term energy consumption forecasting with the ARIMA model using remote control data of commercial buildings. In 2014 IEEE PES General Meeting | Conference & Exposition (pp. 1-5). IEEE.

[22] Zhong, F., Hu, Z., Xia, Q., & Dong, Y. (2020). Energy price forecasting with SVR and LSTM networks. IEEE Transactions on Industrial Informatics, 16(5), 3394-3403.

[23] Li, C., Gooi, H. B., & Lu, Z. (2019). Ensemble deep learning for day-ahead electricity price forecasting. IEEE Transactions on Power Systems, 34(6), 4368-4377.

[24] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

[25] Taylor, S. J., & Letham, B. (2018). Forecasting at scale with Facebook Prophet. The Journal of Open Source Software, 3(22), 649.

[26] Brownlee, J. (2018). How to Improve Machine Learning Results with Ensemble Methods in Python. Machine Learning Mastery. Retrieved from https://machinelearningmastery.com/improve-machine-learning-results-with-ensemble-methods-in-python/

[27] Guryanov, Aleksei. (2019). Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees. 10.1007/978-3-030-37334-4_4.

[28] Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J.C.W., Van den Bossche, P., Van Mierlo, J. and Omar, N., 2018. Random forest regression for online capacity estimation of lithium-ion batteries. Applied energy, 232, pp.197-210.

[29] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[30] Eirgridgroup.com. (2018). Renewable Energy. [online] Available at: http://www.eirgridgroup.com/how-the-grid-works/renewables/ [Accessed 27 July 2023].

[31] "Energy-Charts." Energy-Charts.info, [online] Available at: http://www.energy-charts.info/index.html?l=en&c=IE. [Accessed 27 July 2023].

[32] transparency.entsoe.eu. (n.d.). ENTSO-E Transparency Platform. [online] Available at: http://transparency.entsoe.eu/dashboard/show [Accessed 27 Jul. 2023].

[33] www.met.ie. (n.d.). Historical Data - Met Éireann - The Irish Meteorological Service. [online] Available at: http://met.ie/climate/available-data/historical-data [Accessed 27 Jul. 2023].

[34] Semmler, T., McGrath, R., Steele-Dunne, S., Hanafin, J., Nolan, P. and Wang, S., 2010. Influence of climate change on heating and cooling energy demand in Ireland. International journal of climatology, 30(10), pp.1502-1511.

[35] Shen, Y., Gao, W., & Liu, W. (2017). "Electric Load Forecasting Using an Ensemble Deep Learning Framework." IEEE Transactions on Smart Grid, 8(4), 1778-1787.

[36] Montero-Manso, P., Martinez-Rojas, M., & Cidrás, J. (2020). "Probabilistic Forecasting of Day-Ahead Electricity Prices Using Ensemble Models." Applied Energy, 269, 115053.

[37] Xie, W., Gao, W., & Wei, W. (2018). "Short-Term Electricity Price Forecasting Based on a Hybrid Ensemble Learning Approach." IEEE Transactions on Power Systems, 33(2), 1943-1953.

[38] scikit-learn. (n.d.). sklearn.ensemble.VotingRegressor. [online] Available at: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html [Accessed 27 Jul. 2023].

[39] Zha, D., Fan, Y., Liu, X., & Lin, W. (2020). Hourly short-term energy demand forecasting using a combined model of SARIMA and Prophet. In 2020 IEEE International Conference on Power, Electrical, and Electronics (ICPEE) (pp. 421-426). IEEE.

[40] Bandara, K., Hewamalage, H., Liu, Y.H., Kang, Y. and Bergmeir, C., 2021. Improving the accuracy of global forecasting models using time series data augmentation. Pattern Recognition, 120, p.108148.

[41] Wattics. (2014). The Electricity Market in Ireland: Market Players & Suppliers Review. [online] Available at: https://www.wattics.com/electricity-market-ireland-wattics/ [Accessed 7 Aug. 2023].