The International Journal of Engineering and Science (IJES)
|| Volume || 13 || Issue || 4 || Pages || PP 64-74 || 2024 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Forecasting and Its Applications in Electricity Markets

Chandanam Soumithri *Student*, Dr. Kanna Bhaskar *Professor & Head of the Department*
Department of Electrical and Electronics Engineering,
JNTUH College of Engineering, Hyderabad

--***ABSTRACT***--
Forecasting is determining what is going to happen in the future by analysing what happened in the past and what is going on now. It relies on past and current data and analysis of trends. The most popular subfields of energy forecasting include - Load Forecasting, Electricity Price Forecasting, Wind power Forecasting, Solar Power Forecasting. Electricity Price Forecasting (EPF) is a branch of energy forecasting which focuses on predicting the spot and forward prices in wholesale electricity markets. Since early 1990s, the procedure of deregulation and the creation of competitive power markets have been reshaping. But power is a completely unique commodity, it is far economically non-storable and power system requires a constant balance among manufacturing and power intake. Due to the deregulation of retail electricity market, consumers can choose retail electric suppliers freely, and market entities are facing fierce competition because of the increasing number of new entrants. Under these circumstances, forecasting the changes in all market entities, when market share stabilized, is important for suppliers making marketing decisions. Over the past three decades, accurate modelling and forecasting of electricity prices has grown to be a key problem in competitive energy markets. Electricity demand forecasting is a significant and essential procedure for making plans periodical operations and facility growth inside the strength zone. This project presents an extensive review of the established approaches to electricity price forecasting. In these situations, one can use successfully a simple and fast method of analysis and forecasting, such as the method of Extreme Gradient Boosting (XGBoost). It summarizes the influencing factors of price behaviour and proposes an extended taxonomy of price forecasting methods.
Keywords: *Forecasting, Electricity Load Forecasting, Electricity Price Forecasting, Electricity Markets*

Date of Submission: 02-04-2024 Date of acceptance: 13-04-2024

I. Introduction:

Under restructuring of electrical power industry, different contributors namely generation companies and consumers of energy want to meet in a market to determine on the electricity price. In the contemporary deregulated state of affairs, the forecasting of electricity and price has emerged as one of the most important studies fields in electrical engineering. Load forecasting of researches and academicians are engaged within the hobby of growing tools and algorithms for load and price forecasting. In actual electricity markets, price curve well known shows notably richer shape than load curve and has the following characteristics: high frequency, non-constant mean, and variance, more than one seasonality, calendar effect, excessive level of and high percentage of uncommon charge moves. These kinds of characteristics can be attributed to the subsequent motives, which distinguish electricity from other commodities: (i) non storable nature of electrical power, (ii) the requirement of maintaining constant stability between demand and supply, (iii) inelastic nature of demand for over quick term, and (iv) oligopolistic generation side uncertainties. Similarly, to those marketplace equilibrium is also influenced by way of both load and generation aspect uncertainties. Therefore, price forecasting equipment are important for all market members for their survival under new deregulated surroundings. Even correct load forecasts cannot guarantee profits and the market place threat because of trading is full size because intense volatility of energy prices. Frequently there is a time lag between awareness of an impending event or need and occurrence of that event. This lead time is the main reason for planning and forecasting. If the lead time is zero or very small, there is no need for planning. If the lead time is long, and the outcome of the final event is conditional on identifiable factors, planning can perform an important role. In such situations, forecasting is needed to determine when an event will occur or a need arise, so that appropriate actions can be taken.

In management and administrative situations, the need for planning is great because the lead time for decision making ranges from several years (for the case of capital investments) to a few days or hours (for transportation or production schedules) to a few seconds (for telecommunication routing or electrical utility loading). Forecasting is an important aid in effective and efficient planning. The trend to be able to more accurately predict a wider variety of events, particularly those in the economic/business environment, will continue to provide a better base from which to plan. Formal forecasting methods are the means by which this improvement is occurring.

Regardless of these improvements, two important comments must be kept in view. The first is that successful forecasting is not always directly useful to managers and others. A second important point is the distinction between uncontrollable external events (originating with the national economy, governments, customers, and competitors) and controllable internal events (such as marketing or manufacturing decisions within the firm). The success of a company depends on both types of events, but forecasting applies directly to the former, while decision making applies directly to the decision-making latter. Planning is the link that integrates both.

A wide variety of forecasting methods are available to management (see, for example, Makridakis and Wheelwright, 1989). These range from the most naive methods, such as use of the most recent observation as a forecast, to highly complex approaches such as neural nets and econometric systems of simultaneous equations. In addition, the widespread introduction of computers has led to readily available software for applying forecasting techniques. Complementing such software and hardware has been the availability of data describing the state of economic events (GNP, consumption, etc.) and natural phenomena (temperature, rainfall, etc.). These data in conjunction with organizational statistics (sales, prices, advertising, etc.) and technological know-how provide the base of past information needed for the various forecasting methods.

Some of the areas in which forecasting currently plays an important role are:

1.Scheduling: Efficient use of resources requires the scheduling of production, transportation, cash, personnel, and so on. Forecasts of the level of demand for product, material, labour, financing, or service are an essential input to such scheduling.

2. Acquiring resources: The lead time for acquiring raw materials, hiring personnel, or buying machinery and equipment can vary from a few days to several years. Forecasting is required to determine future resource requirements.

3.Determining resource requirements: all organizations must determine what resources they want to have in the long-term. Such decisions depend on market opportunities, environmental factors, and the internal development of financial, human, product, and technological resources. These determinations all require good forecasts and managers who can interpret the predictions and make appropriate decisions.

Although there are many different areas requiring forecasts, the preceding three categories are typical of the short-, medium-, and long-term forecasting requirements of today's organizations.

An overview of Forecasting Techniques:

Forecasting situations vary widely in their time horizons, factors determining actual outcomes, types of data patterns, and many other aspects. To deal with such diverse applications, several techniques have been developed. These fall into two major categories: quantitative and qualitative methods.

QUANTITATIVE: Sufficient quantitative information is available.

- Time series: Predicting the continuation of historical patterns such as the growth in sales or gross national product.
- Explanatory: Understanding how explanatory variables such as prices and advertising affect sales.

Quantitative forecasting can be applied when three conditions quantitative exist: forecasting.

1. Information about the past is available.
2. This information can be quantified in the form of numerical data.
3. It can be assumed that some aspects of the past pattern will continue in future.

QUALITATIVE: Little or no quantitative information is available, but sufficient qualitative knowledge exists.

- Predicting the speed of telecommunications around the year 2020.
- Forecasting how a large increase in oil prices will affect the consumption of oil.

Qualitative forecasting methods, on the other hand, do not require forecasting data in the same manner as quantitative forecasting methods. The inputs required depend on the specific method and are mainly the product of judgment and accumulated knowledge. Qualitative approaches often require inputs from a number of specially trained people.

UNPREDICTABLE: Little or no information is available.

- Predicting the effects of interplanetary travel.
- Predicting the discovery of a new, very cheap form of energy that produces no pollution.

The basic steps in a forecasting task:

1. Problem Definition
2. Gathering Information
3. Preliminary (exploratory) analysis
4. Choosing and fitting models

5. Using and evaluating a forecasting model

Review of modelling process:

Almost all the review offers their own classifications of the numerous methods which have been advanced for reading and predicting energy expenses. Some of them are better, some are worse, but all have many stuffs in commonplace. Without lack of generality, we take the category of Weron (2006) as a place to begin, with six business of models. We then modify it through combining the first two organizations into one large elegance (due to the decreasing popularity of manufacturing- price models and the increasing use of simulation models. **Error! Reference source not found.** shows the different modelling approaches.

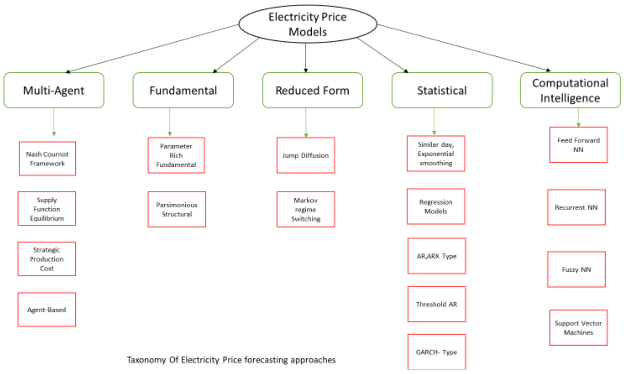


Figure 1: Taxonomy of electricity Price Forecasting approaches

- *Multi-agent* (*multi-agent simulation*, *equilibrium*, *game theoretic*) models, which simulate the operation of a system of heterogeneous agents (generating units, companies) interacting with each other, and build the price process by matching the demand and supply in the market.
- *Fundamental* (*structural*) methods, which describe the price dynamics by modelling the impacts of important physical and economic factors on the price of electricity.
- *Reduced-form* (*quantitative*, *stochastic*) models, which characterize the statistical properties of electricity prices over time, with the ultimate objective of derivatives evaluation and risk management.
- *Statistical* (*econometric*, *technical analysis*) approaches, which are either direct applications of the statistical techniques of load forecasting or power market implementations of econometric models.
- *Computational intelligence* (*artificial intelligence-based*, *non-parametric*, *non-linear statistical*) techniques, which combine elements of learning, evolution, and fuzziness to create approaches that are capable of adapting to complex dynamic systems, and may be regarded as 'intelligent' in this sense.

Load Forecasting Procedure:

Load forecasting is divided into three types depending on the forecasting horizon: short-term load forecasting (STLF), medium-term load forecasting (MTLF) and long-term load forecasting (LTLF)

Short Term Forecast - Short time forecasting is done for day-to-day operation, to ensure generation capacity for a week, and for maintaining the required spinning reserve. Hence it is usually done 24 hours ahead of when the weather forecast for the following day because available from the meteorological office. Application of short time load forecasting STF are

- To meet the short-term demand with the most economic commitment of generation sources.
- To access the power system security based on the information available to the dispatches to prepare the necessary corrective action.

Medium-term Forecast - This forecasting is done for 5 to 6 years and play a major role while planning the size of a power plant and construction and installation of the equipment in power plants, Applications of medium-term load forecasting MTLF

- For the estimation of fuel (Coal, diesel, water etc)
- For the estimation of peak power & energy requirement for each month of the coming year.

Long-term forecast - Long-term forecast may extend over a period of 20 years or even more and in advance, in order to facilitate various plan like, the preparation of maintenance schedule of the generating units, plan for future expansion of the generating capacity enter into agreements for energy interchange with neighbouring utilities.

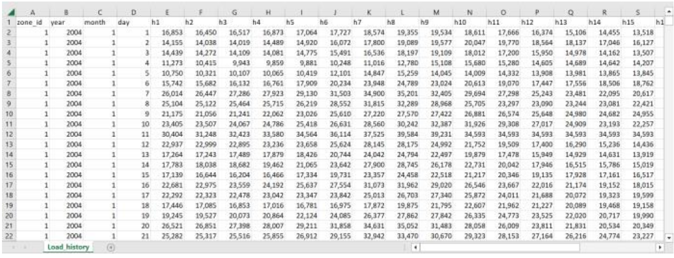
II. METHODOLOGY:

Data Gathering:

The Global Energy Forecasting Competition (GEFCom2012) attracted hundreds of participants worldwide, who contributed many novel ideas to the energy forecasting field. This paper introduces both tracks of GEFCom2012, hierarchical load forecasting, with details on the aspects of the problem and the data. The historical data collected for 3 years for all the twelve months. The parameters found in this dataset are as follows: zone_id, Year, Month (mentioned as 1 to 12), Day (mentioned as 1 to 7), Load in megawatts (MW) for hourly daily. Figure 2 shows a sample of a day datasheet.

XGBoost:

XGBoost, or Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It is designed to handle large datasets and is widely used for regression, classification, and ranking problems. XGBoost is built upon several machine learning concepts and algorithms, including



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	zone_id	year	month	day	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	h13	h14	h15	h1
2	1	2004	1	1	16,853	16,450	16,517	16,873	17,064	17,727	18,574	19,355	19,534	18,611	17,666	16,374	15,106	14,455	13,518	
3	1	2004	1	2	14,155	14,038	14,019	14,489	14,920	16,072	17,800	19,089	19,577	20,047	19,770	18,564	18,137	17,046	16,127	
4	1	2004	1	3	14,439	14,272	14,109	14,081	14,775	15,491	16,536	18,197	19,109	18,012	17,200	15,950	14,978	14,162	13,507	
5	1	2004	1	4	11,273	10,415	9,943	9,859	9,881	10,248	11,016	12,780	15,108	15,680	15,280	14,605	14,689	14,642	14,207	
6	1	2004	1	5	10,750	10,321	10,107	10,065	10,419	12,101	14,847	15,259	14,045	14,009	14,332	13,908	13,981	13,865	13,845	
7	1	2004	1	6	15,742	15,682	16,132	16,761	17,909	20,234	23,948	24,789	23,024	20,613	19,070	17,447	17,556	18,506	18,762	
8	1	2004	1	7	26,014	26,447	27,286	27,923	29,130	31,503	34,900	35,201	32,405	29,694	27,298	25,243	23,481	22,095	20,617	
9	1	2004	1	8	25,104	25,122	25,464	25,715	26,219	28,552	31,815	32,289	28,968	25,705	23,297	23,090	23,244	23,081	22,421	
10	1	2004	1	9	21,175	21,056	21,241	22,062	23,026	25,610	27,220	27,570	27,422	26,881	26,574	25,648	24,980	24,682	24,955	
11	1	2004	1	10	23,405	23,507	24,067	24,786	25,418	26,631	28,560	30,242	32,387	31,926	29,308	27,017	24,909	23,193	22,257	
12	1	2004	1	11	30,404	31,248	32,423	33,580	34,564	36,114	37,525	39,584	39,231	34,593	34,593	34,593	34,593	34,593	34,593	
13	1	2004	1	12	22,937	22,999	22,895	23,236	23,658	25,624	28,145	28,175	24,992	21,752	19,509	17,400	16,290	15,236	14,436	
14	1	2004	1	13	17,264	17,243	17,489	17,879	18,426	20,744	24,042	24,794	22,497	19,879	17,478	15,949	14,929	14,631	13,919	
15	1	2004	1	14	17,783	18,038	18,682	19,462	21,065	23,642	27,900	28,745	26,178	22,731	20,042	17,946	16,515	15,786	15,019	
16	1	2004	1	15	17,139	16,644	16,204	16,466	17,334	19,731	23,357	24,458	22,518	21,217	20,346	19,135	17,928	17,161	16,517	
17	1	2004	1	16	22,681	22,975	23,559	24,192	25,637	27,554	31,073	31,962	29,020	26,546	23,667	22,016	21,174	19,152	18,015	
18	1	2004	1	17	22,292	22,323	22,478	23,042	23,347	23,842	25,013	26,703	27,340	25,872	24,011	21,688	20,072	19,323	19,599	
19	1	2004	1	18	17,446	17,085	16,853	17,016	16,781	16,975	17,872	19,875	21,795	22,607	21,962	21,227	20,089	19,468	19,158	
20	1	2004	1	19	19,245	19,527	20,073	20,864	22,124	24,085	26,377	27,862	27,842	26,335	24,773	23,525	22,020	20,717	19,990	
21	1	2004	1	20	26,521	26,851	27,398	28,007	29,211	31,858	34,631	35,052	31,483	28,058	26,009	23,811	21,831	20,534	20,349	
22	1	2004	1	21	25,282	25,317	25,516	25,855	26,912	29,155	32,942	33,470	30,670	29,323	28,153	27,164	26,216	24,774	23,227	

Load_history

Figure 1: Spreadsheet of the Load data taken from GEFC

supervised machine learning, decision trees, ensemble learning, and gradient boosting.

XGBoost is a classification algorithm widely used in machine learning for its speed, efficiency, and performance on large datasets. It is an open-source library that excels in training and testing models on substantial data volumes, making it popular across various domains. XGBoost implements gradient boosting decision trees, offering features like regularization to prevent overfitting, handling of missing values, and parallel learning capabilities for scalability.

This algorithm has gained significant attention for its computational efficiency, feature importance analysis, and ability to handle real-world data effectively without extensive preprocessing. XGBoost's popularity stems from its speed, accuracy, and versatility in tasks such as regression, classification, and ranking.

Supervised machine learning involves training a model to find patterns in a dataset with labels and features, which can then be used to predict labels on new data. Decision trees are a type of machine learning algorithm that create a tree-like model to make predictions. Ensemble learning combines multiple models to improve accuracy, while gradient boosting is a technique that creates a series of models and combines them to create an overall model that is more accurate than any individual model. XGBoost includes features like parallelization, regularization, non-linearity, and cross-validation to enhance its performance and generalization capabilities. It can be parallelized to train with multiple CPU cores, includes different regularization penalties to avoid overfitting, can detect and learn from non-linear data patterns, and comes with built-in cross-validation for model evaluation. Additionally, XGBoost is scalable and can run distributed thanks to distributed servers and clusters like Hadoop and Spark, allowing it to process enormous amounts of data.

XGBoost Benefits and Attributes:

1. High accuracy: XGBoost is known for its accuracy and has been shown to outperform other machine learning algorithms in many predictive modeling tasks.
2. Scalability: It is highly scalable and can handle large datasets with millions of rows and columns.
3. Efficiency: It is designed to be computationally efficient and can quickly train models on large datasets.
4. Flexibility: It supports a variety of data types and objectives, including regression, classification, and ranking problems.
5. Regularization: It incorporates regularization techniques to avoid overfitting and improve generalization performance.
6. Interpretability: It provides feature importance scores that can help users understand which features are most important for making predictions.
7. Open-source: XGBoost is an open-source library that is widely used and supported by the data science community.

Methodological and technical aspects related to the XGBoost method:
XGBoost (Extreme Gradient Boosting) is an ensemble learning method that combines the predictions from multiple individual models (typically decision trees) to make a final prediction. In simple mathematical terms, it can be represented as follows:
Let's consider a dataset with 'n' samples and 'm' features, where the input data is represented as 'X' and the corresponding target values are represented as 'y'.

- The base model can be represented as:

$y^i=f(xi)$
where y^i is the predicted output for sample i, and $f(xi)$ is the prediction function.

- The objective function for XGBoost is a sum of a loss function L and a regularization term Ω:

$Obj=\sum i=1nL(yi,y^i)+\sum k=1K\Omega(fk)$
where K is the number of base learners (trees), and fk represents the prediction of the kth tree.

- To optimize the objective function, XGBoost performs gradient boosting by iteratively adding new base learners (trees) to minimize the objective function. At each iteration, a new tree hk is trained to correct the errors made by the existing ensemble:

$y^i(t)=y^i(t-1)+hk(xi)$
where $y^i(t)$ represents the ensemble prediction after t iterations, and $k(xi)$ is the prediction of the new tree hk.

- XGBoost uses a weighted sum of the predictions from all trees to make the final prediction:

$y^i=\sum k=1Khk(xi)$
where y^i is the final prediction for sample i.
In summary, XGBoost combines the predictions from multiple trees by iteratively adding new trees to minimize a predefined objective function, thereby improving the overall predictive performance

Forecasting through this method implies running through the following stages:

1. **Data Preparation:**

Clean the dataset by handling missing values, outliers, and encoding categorical variables if necessary.
Split the dataset into training and testing sets.

2. **Model Training:**

Initialize an XGBoost model with appropriate hyperparameters such as learning rate, maximum depth of trees, and number of estimators (trees).
Train the model on the training data using gradient boosting.
The model iteratively adds new trees to minimize the objective function, correcting errors made by the previous ensemble.

3. **Model Evaluation:**

Evaluate the trained model's performance on the testing set using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or root mean squared error (RMSE).
Assess the model's accuracy and generalization ability.

4. **Hyperparameter Tuning (Optional):**

Perform hyperparameter tuning using techniques such as grid search or random search to find the optimal combination of hyperparameters.
This step aims to improve the model's performance further.

5. **Forecasting:**

Once the model is trained and evaluated, use it to make forecasts on new or unseen data.
Input the relevant features of the new data into the trained model to predict the target variable.
Generate forecasts for future time periods or events.

6. **Model Monitoring and Maintenance:**

Monitor the model's performance over time and retrain it periodically with updated data if necessary.
Keep track of any changes in data patterns or underlying relationships that may affect the model's accuracy.
By following these stages, one can effectively use XGBoost for forecasting tasks, whether it's predicting sales, stock prices, demand, or any other time-series data
The fundamental hierarchical layout of the extreme gradient boosting tree model is depicted in Figure 2.

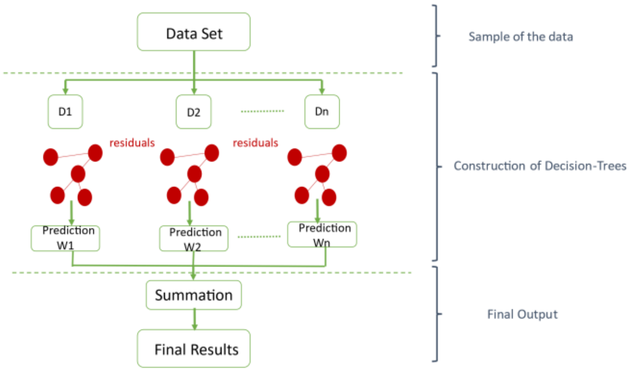


Figure 2: The structure of the XGBoost Algorithm

The Dataset formatted and its structure is displayed from the Excel in Figure 3.

	A	B	C	D	E	F	G
1	Month	Holiday	Day	Hour	Temp	Previous Loa	Present Loa
2	10	0	1	1	58	10,463	11,077
3	10	0	1	2	58	9,695	10,544
4	10	0	1	3	57	9,480	10,398
5	10	0	1	4	57	9,421	10,467
6	10	0	1	5	57	9,364	10,917
7	10	0	1	6	56	9,659	12,308
8	10	0	1	7	56	10,811	15,484
9	10	0	1	8	58	11,943	15,667
10	10	0	1	9	67	14,057	14,614
11	10	0	1	10	72	14,795	14,131
12	10	0	1	11	73	15,190	14,302
13	10	0	1	12	74	14,908	14,454
14	10	0	1	13	75	14,732	14,782
15	10	0	1	14	73	14,967	15,434
16	10	0	1	15	75	15,354	15,713
17	10	0	1	16	73	15,876	16,432
18	10	0	1	17	72	16,264	17,819
19	10	0	1	18	72	16,585	18,790
20	10	0	1	19	69	16,688	18,545
21	10	0	1	20	66	17,528	19,223
22	10	0	1	21	66	17,604	18,970
23	10	0	1	22	65	15,831	17,072
24	10	0	1	23	63	14,182	14,572
25	10	0	1	24	61	12,359	12,597
26	10	0	2	1	60	11,156	11,380
27	10	0	2	2	60	10,561	10,924
28	10	0	2	3	61	10,337	10,660

Figure 3: Spread sheet interface of the Training Data set

The basic, level wise structure of the code is shown in the Figure 4.

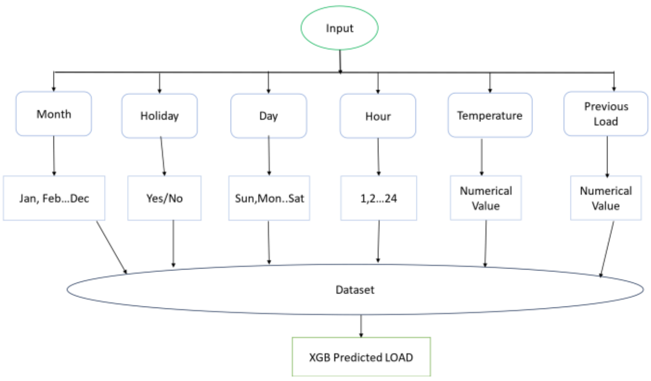


Figure 4: Structure of the code

The dataset's input format depicted in the coding is illustrated in the Figure 5.

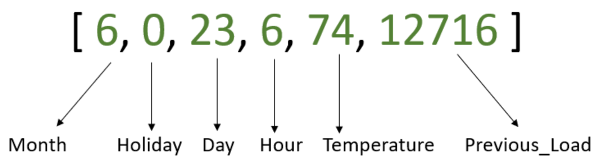


Figure 5: Input Form

Assumptions when creating a tree:
Below are some of the assumptions we make when using a tree:
XGBoost tree creation relies on several key assumptions to construct effective ensemble models. It assumes an additive model structure, where predictions are the sum of individual trees. Each tree is considered a weak learner, meaning it performs slightly better than random guessing. Through gradient boosting, XGBoost iteratively improves predictions by minimizing residuals left by preceding trees. Additionally, trees are assumed to interact, allowing subsequent ones to correct errors and enhance overall performance. Regularization is also employed to control complexity, ensuring models generalize well to unseen data and avoid overfitting. These assumptions guide the construction and optimization of XGBoost trees, enabling them to effectively learn from data and make accurate predictions.

How do XGBoost trees work?

1. Initial Prediction:
XGBoost starts with a simple prediction for all data points, usually the average of the target variable.

2. **Building the First Tree**:
The first decision tree is created to predict the discrepancies (residuals) between the initial prediction and actual target values.

3. Updating Predictions:
The predictions are updated by adding the first tree's predictions to the initial prediction.

4. Building Subsequent Trees:
Additional trees are built to predict the remaining residuals after previous trees' predictions are added.

5. Adding Trees to the Ensemble:
All trees' predictions are combined to make the final prediction, with each tree's contribution weighted based on performance.

6. Regularization and Hyperparameters:
Regularization techniques and hyperparameters are used to control tree complexity and prevent overfitting.

7. Gradient Boosting Optimization:
Gradient boosting optimizes the ensemble by iteratively minimizing the objective function's gradient, updating model parameters accordingly.

Building a XGBoost Tree:

- Load the dataset. It consists of 6 features, Month, Holiday, Day, Hour, Temperature, Previous Load and Actual Load.
- We will take Month, Holiday, Day, Hour, Temperature and Previous Load as our independent variables X. Actual Load is our dependent variable y.
- The next step is to split the dataset into training and test.
- Perform feature scaling
- Fit the model in the XGBoost Algorithm.
- Make predictions

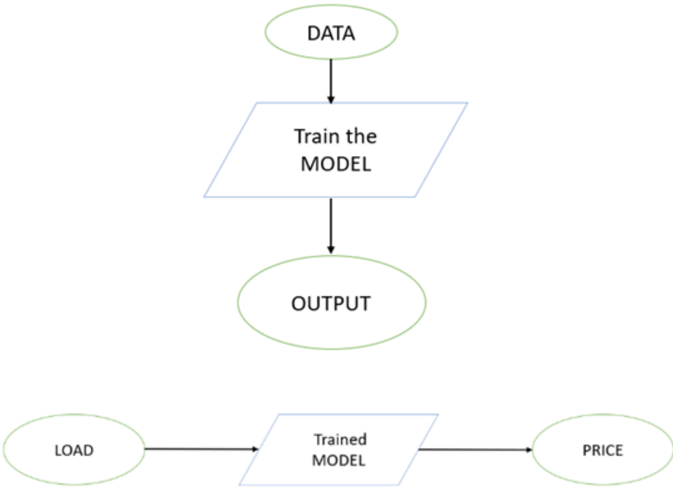


Figure 6:Structure of XGBoost Regression

Input=

```
([[6,0,23,6,74,12716],[6,0,23,7,75,13437],[6,0,23,8,78,14985],[6,0,23,9,80,17256],[6,0,23,10,80,19408],
  [6,0,23,17,82,28850],[6,0,23,18,81,29948],[6,0,23,19,81,29478],[6,0,23,20,79,27933],[6,0,23,21,77,25866],
  [6,0,23,22,75,24607],[6,0,24,1,74,16423],[6,0,24,2,71,14689],[6,0,24,3,72,13652],[6,0,24,4,70,13058],
  [6,0,24,5,68,12851],[6,0,24,13,82,24442],[6,0,24,14,82,25887],[6,0,24,15,82,27603],[6,0,24,16,83,28688],
  [6,0,24,22,71,24941],[6,0,24,23,69,22224],[6,0,24,24,67,19163],[6,0,25,12,87,21402],[6,0,25,13,88,21225],
  [6,0,25,14,87,22450],[6,0,25,15,87,24756],[6,0,25,16,86,27680],[6,0,25,17,85,30270],[6,0,25,18,85,32036],
  [6,0,25,19,83,33031],[6,0,26,1,74,15010],[6,0,26,2,74,13528],[6,0,26,3,74,13229],[6,0,26,4,74,12954],
  [6,0,27,24,76,20526],[6,0,28,1,76,17850],[6,0,28,2,76,16047],[6,0,28,3,78,15042],[6,0,29,20,88,27483],
  [6,0,29,21,85,25699],[6,0,29,22,84,24851],[6,0,29,23,76,22659],[6,0,29,24,74,19708]])
```

Output =

```
[12824.414 14312.713 16108.973 16899.613 17906.842 28114.371 30592.287
 29636.992 27526.285 25875.895 24935.895 16551.758 13960.587 12635.768
 11721.464 11578.948 21812.506 24117.844 25254.566 26688.336 24472.49
 20027.568 16837.229 20168.494 24186.127 26564.473 27603.488 28322.242
 29142.969 28781.521 31616.37  14567.833 13990.053 12986.571 13230.449
 18304.83  13408.457 13785.667 12477.281 27529.781 30347.111 28079.898
 21374.035 18101.482]
```

After utilizing XGBoost to forecast the load, the Figure 7 displays the comparison between the Actual load and the Predicted load.

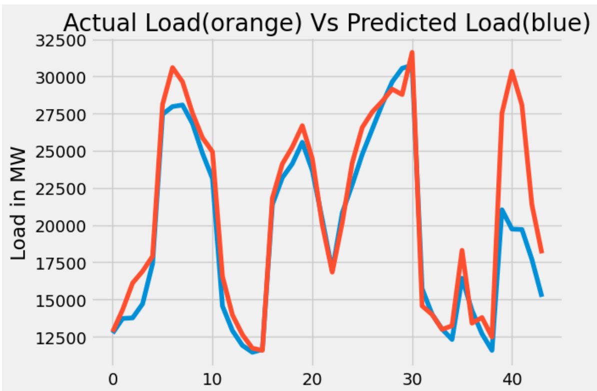


Figure 7: Comparison of Actual load and Forecasted load

Table 1: *MSE and RMSE values*

	MSE	RMSE
XGBoost Algorithm	0.01361	0.11668

For the load input range of 12000 to 22000 we got Mean Squared Error (MSE) as 0.013613360241088219 and the Root Mean Squared Error (RMSE) as 0.11667630539697517 as shown in Table 1Table 1.

XGBoost Regression:

XG regression, also known as Extreme Gradient Boosting Regression, is a powerful machine learning algorithm that belongs to the ensemble learning methods. It works by building multiple decision trees sequentially, where each tree corrects the errors of the previous one. This iterative process allows XG regression to make accurate predictions by combining the outputs of these individual trees. The key features of XG regression include its ability to handle complex relationships between variables, manage missing data effectively, and prevent overfitting through regularization techniques. It is widely used in various fields such as finance, healthcare, and marketing due to its high predictive accuracy and efficiency. In summary, XG regression is a sophisticated regression technique that leverages ensemble learning to create a robust predictive model by aggregating the strength of multiple decision trees in a sequential manner.

The historical data collected for 4 years for all the twelve months. The parameters found in this dataset are as follows:

Load in megawatts (MW) for hourly daily and Price in $/MWh. Figure 8 shows a sample of a day datasheet.

	A	B	C	D
1	Load MWh	Price $/MWh		
2	24,000	53.00		
3	24,800	40.50		
4	7,200	40.75		
5	6,400	42.00		
6	11,200	55.00		
7	2,400	50.00		
8	12,800	58.00		
9	39,200	50.50		
10	10,400	57.40		
11	14,400	41.50		
12	16,000	39.25		
13	16,000	40.25		
14	15,200	40.05		
15	23,200	53.00		
16	9,600	45.00		
17	8,000	43.00		
18	18,400	44.00		
19	12,800	50.00		
20	19,200	43.50		
21	8,800	36.75		
22	8,800	36.75		
23	10,400	35.25		
24	13,600	37.00		

Figure 8: Spread Sheet interface of the training Data Set

Problem Analysis:

In this data, we have one independent variable *Load* and one dependent variable *Price* which we have to predict. In this problem, we must build a XGBoost Regression Model which will study the correlation between the Load and Price of the Electricity markets and predict the price for the load based on the load in a particular hour.

- Importing libraries
- Importing dataset
- Splitting the dataset into the Training set and Test set
- Training the XGBoost Regression model on the training set
- Predicting the Results

Input =

```
([[16971.814],[23737.307],[20331.914],[16459.236],
       [19371.162],[21165.318],[16443.037],[18560.498],[1000]])
```

Output =

```
[41.477543 35.13089  40.276505 45.68983  40.47629  37.264107 45.68983
 39.117268 35.31538 ]
```

The predicted prices can be utilized in electric markets for bidding strategies, highlighting the practical implications of the forecasting results on market operations and decision-making processes.

III. Conclusion:

Forecasting electricity prices is complex, in the increasingly competitive electricity markets. It is essential for all market participants to forecast prices one day ahead. Accurately predicting such prices makes it easier for energy suppliers to adjust them bidding tactics and meanwhile allows consumers to come up with a hedging plan themselves against high prices. The non-storability characteristics of electricity prevent using inventories to smooth out supply and demand shocks, which causes increased uncertainty in electricity prices. Therefore, the price forecast can be accurate enable electricity producers to optimally allocate their resources for dynamic management demand from different regions.

The empirical findings are significant for policy makers, electricity and electricity suppliers' generators and regular consumers. In particular, it can enable accurate forecasting power generators to balance the uncertain demand and supply from different regions, which may eventually lead to a reduction in electricity price variability. In addition, these the findings are important with respect to network stability and economic profitability market participants. Specifically, with a better understanding of the variations in electricity prices, network operators may be better able to avoid such large differences in the network example of fluctuating electricity prices. A better understanding of the exact forecasting significantly contributes to the economic advantages of market entities. In moreover, given the increase, these findings are of great interest to policy makers redirecting resources to produce clean energy. Accurate prediction This framework can enable policy makers to design a plan for better network integration systems and electricity prices in different regions. This can lead to a reduction disproportion of supply and demand and finally reducing the uncertainty of electricity prices.

Because understanding and capturing price dynamics is important for electricity numerical results that robust prognostic methods generate showing the importance of properly designing the forecasting process, policy implications for market efficiency and predictability in electricity markets. However, this study only focused on data statistics of the electricity market and did not include important operational one's factors such as the profile of electricity generation sources and seasonality. And the combined method will be interesting for investigating the main influencing factors electricity prices. This too represents a direction for future research.

References

[1]. Grzegorz Dudek "Probabilistic Forecasting of Electricity Prices Using Kernel Regression" 2018 15th International Conference on the European Energy Market (EEM).

[2]. José Portela González; Antonio Muñoz San Muñoz San Roque; Estrella Alonso Pérez "Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting" IEEE Transactions on Power Systems (Volume: 33, Issue: 1, Jan. 2018)

[3]. Catherine McHugh; Sonya Coleman; Dermot Kerr; Daniel McGlynn "Forecasting Day-ahead Electricity Prices with A SARIMAX Model" 2019 IEEE Symposium Series on Computational Intelligence (SSCI)

[4]. Muhammad Furqan Azam; Muhammad Shahzad Younis "Multi-Horizon Electricity Load and Price Forecasting Using an Interpretable Multi-Head Self-Attention and EEMD-Based Framework"

[5]. Mughees Ali; Zahoor Ali Khan; Sana Mujeeb; Shahid Abbas; Nadeem Javaid "Short-Term Electricity Price and Load Forecasting using Enhanced Support Vector Machine and K-Nearest Neighbor" 2019 Sixth HCT Information Technology Trends (ITT)

[6]. José Portela González; Antonio Muñoz San Muñoz San Roque; Estrella Alonso Pérez "Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting-2017" IEEE Transactions on Power Systems (Volume: 33, Issue: 1, Jan. 2018)

[7]. Kanna Bhaskar and S. N. Singh "AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network" IEEE Transactions on Sustainable Energy (Volume: 3, Issue: 2, April 2012)

[8]. R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future", Intl. J. of Forecasting, vol. 30, no. 4, pp. 1030-1081, 2014.

[9]. Takuji Matsumoto, Misao Endo "Electricity price forecast based on weekly weather forecast and its application to arbitrage in the forward market" 2021 11th International Conference on Power, Energy and Electrical Engineering (CPEEE)

[10]. Li Shejiao , Wen Chuanbo , Wang Songwei , Chen Zhiguo "A Hybrid Forecasting Algorithm and Its Application to Economic Analysis" The Proceedings of the Multiconference on "Computational Engineering in Systems Applications"

[11]. C. Albanese, H. Lo, S. Tompaidis **"A numerical algorithm for pricing electricity derivatives for jump-diffusion processes based on continuous time lattices"** European Journal of Operational Research, 222 (2) (2012), pp. 361-368

[12]. S.K. Aggarwal, L.M. Saini, A. Kumar **"Electricity price forecasting in Ontario electricity market using wavelet transform in artificial neural network-based model"** International Journal of Control, Automation and Systems, 6 (5) (2008), pp. 639-650

[13]. S.K. Aggarwal, L.M. Saini, A. Kumar **"Electricity price forecasting in deregulated markets: A review and evaluation"** International Journal of Electrical Power and Energy Systems, 31 (2009), pp. 13-22

[14]. S.K. Aggarwal, L.M. Saini, A. Kumar "Short term price forecasting in deregulated electricity markets. A review of statistical models and key issues" International Journal of Energy Sector Management, 3 (4) (2009), pp. 333-358

[15]. R. Aïd, L. Campi, N. Langrené "A structural risk-neutral model for pricing and hedging power derivatives" Mathematical Finance, 23 (3) (2013), pp. 387-438

[16]. W. Charytoniuk, E.D. Box, W.-J. Lee; M.-S. Chen, P.Kotas, P. Van Olinda"Neural-network-based demand forecasting in a deregulated environment" IEEE Transactions on Industry Applications (Volume: 36, Issue: 3, May-June 2000)

[17]. G. E. P. Box and G. M. Jenkins, "Time Series Analysis-Forecasting and Control", CA, San Francisco: Holden-Day, 1976.

[18]. A. D. Papalexopoulos, S. Hao and T.-M. Peng, "An implementation of a neural network-based load forecasting model for the EMS", IEEE Trans. Power Syst., vol. 9, pp. 1956-1962, Nov. 1994.

[19]. Cai Chang-chun; Wu Min, "Support vector machines with similar day's training sample application in short-term load forecasting" 2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies

[20]. Ju Ping, Wang Junkai, Zhao Xiayang. Ninety-six points short-term load forecasting-theory & applications [J]. Automation of Electric Power Systems, 2001, 25(22): 32-36.

[21]. T. M. Peng, N. F. Hubele and G. G. Karady, "An adaptive neural network approach to one-week ahead forecasting", IEEE Trans. Power Syst., vol. 8, pp. 1195-1203, Aug. 1993.

[22]. C. N. Lu, H. T. Wu and B. Vemuri, "Neural network based short term load forecasting", IEEE Trans. Power Syst., vol. 8, pp. 336-342, Feb. 1993.

[23]. Feng Han; Qing Zhang; Xu Zhang; Tingjiao Li, "Application of fuzzy neural network to power system short-term load forecast",2010 International Conference On Computer Design and Applications

[24]. Venkata Ramana K, M. Kathiravan "A Novel XG-Boost Approach for Electricity Load Balancing and Demand Prediction" 2023 ICSCNA, IEEE Xplore 01 January 2024.