International Journal of Advanced Technology in Engineering and Science
Vol. No. 11, Issue No. 06, June 2023
www.ijates.com
ijates
ISSN 2348 - 7550

Bagging Regressor with XGBoost Regressor as Base Estimator for Predictions in Indian Power-Energy Trading Market

Durgesh Tiwari[1], Dr. Sudhir Agarwal[2], Saurabh Singh[3],

[1] M.Tech Student, Department of Computer Science Engineering,
[1] Buddha Institute of Technology, Gorakhpur, Uttar Pradesh, India
dtiwari.it@gmail.com[1]
[2] Professor, Department of Computer Science Engineering,
[2] Buddha Institute of Technology, Gorakhpur, Uttar Pradesh, India
sagarwal22@bit.ac.in [2]
[3] Assistant Professor, Department of Computer Science Engineering,
[3] IET, Deen Dayal Upadhyay Gorakhpur University, Gorakhpur, Uttar Pradesh, India
saurabh9singh9@gmail.com[3]

Abstract:

For the power and energy sector, forecasting has been crucial. Thousands of studies on projecting electricity demand, costs, and renewable generation (such wind and solar power) have been written by academics and industry professionals. In this paper, examined how well various machine learning methodologies can forecast power costs. By taking into account previously underutilized predictive information like the pricing histories of neighboring nations, particularly extend ensemble techniques. We demonstrated that even during the current period; when unexpected changes are taking place, these factors significantly raise the quality of projections. In order to achieve high accuracy in power price forecasting, we proposed an ensemble method based solution for prediction of energy prices in Indian day ahead energy market.

***Keywords**: Electricity Price prediction, Forecasting, Machine learning, Ensemble approaches*

I. INTRODUCTION

The methods used to make business decisions include forecasting. The forecasting of load, generation, prices, and other factors is a necessity for the energy sector. All sectors of the energy industry use these forecasts for planning and running both power systems and commercial entities [1].

Electricity Price Forecasting (EPF) is a challenge that is getting harder and harder to address. For the energy transition to be successful, the applications made feasible by a price forecasting model are essential. They encourage smart applications like self-consumption or automobile battery optimization and enable owners of renewable energy generating means to benefit on the market by anticipating price changes [2].

Various aspects must be considered in order to comprehend electricity pricing at the same time. Energy transition strategies, for instance, raise the share of renewable energy in overall production [3] and enact fresh market laws such taxing carbon dioxide emissions. Additionally, cross-border connections are growing, and some markets, like the EPEX SPOT1, set prices for all of Europe, making it necessary to anticipate on a continental scale.

In addition, price spikes can be positive as well as negative as a result of the pricing algorithms [4] employed to balance generation and consumption. These spikes are challenging for conventional forecasting algorithms to handle and can cause significant losses for naïve business owners. Repeated lockdowns, which have had a significant impact on the European market lately, are notable for this time period in particular.

Machine Learning (ML) models, meanwhile, are getting better at handling complicated circumstances and addressing challenging tasks [5]. However, if the indicated methods and parameters are not fully reported, it can often be difficult to replicate them. Additionally, ML models are renowned for being opaque, challenging to understand, and frequently referred to as "black box" models. Data analysts typically base their decision to use or not utilize them on a single metric assessed simply on a single dataset.

In general, scholars and company owners are becoming more interested in EPF [6-7]. Every EPF publication suggests cutting-edge and effective solutions, but it is challenging to compare the literature since there are so many different markets, forecasting requirements, time spans, models, and methods [7].Additionally, it may be challenging to replicate the findings of a particular publication because specifics are sometimes left out, and stochastic processes may not be reproducible due to a simple lack of seeds. Another drawback is the absence of benchmarks for model comparison, which needs to be remedied in light of cutting-edge studies for other ML applications. Finally, in order to understand the phenomena on which the model is based and how it makes predictions, users of these models need explanations. This enables one to follow or disregard an unexpected prediction in a highly volatile market like the one for energy. We think it's crucial to describe any EPF model's output because doing so will make it clearer what each model actually captures and what it doesn't. This makes it easier to identify the characteristics that influence the prediction.

The organization of the paper is as follows; Section II presents a Machine Learning for EPF; Section III presents a literature survey; Section IV presents a proposed work; Section V shows the results and lastly section VI gives the conclusion.

II. MACHINE LEARNING FOR EPF

A field of computer science called machine learning (ML) proposes predicting models by using effective algorithms for learning from data. Because of the vast amount of data that is currently accessible and the increasing computational capability of machines, this topic has attracted a lot of attention in recent years. In the area of forecasting, ML models have been successful in resolving extremely challenging issues in multivariate Time Series regression as well as image processing [8]. We concentrate on these methods because we think that the potential of ML models in the area of EPF has not yet been completely explored. Also focus on the four models that are listed below. Also discuss the metrics, tests, and data preprocessing methods we used to compare them. Additionally, describe how we set values for hyper-parameters and lay out the recalibration method we employ to modify models in response to recent changes in the data.

SARIMA - Time Series Model

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is a powerful time series forecasting technique that captures both trend and seasonal patterns. It combines autoregressive (AR), integrated (I), and moving average (MA) components to model the data. The AR component captures the relationship between current and past values, while differencing in the component removes the trend. The MA component accounts for short-term fluctuations. SARIMA also incorporates a seasonal component to capture periodic patterns. By estimating parameters from autocorrelation functions, SARIMA can make accurate forecasts. However, it assumes linearity and stationarity, limiting its performance on non-linear or nonstationary data. Overall, SARIMA is widely used for diverse time series analyses.

XGBoost Regressor – Boosting Ensemble Methods

The XGBoost regressor is a powerful machine learning algorithm used for regression tasks. It combines gradient boosting principles with regression analysis to create a robust model for predicting continuous numerical values.

III. LITERATURE SURVEY

Yang et al [11] In order Current models for forecasting electricity prices ignore the importance of adaptive data preprocessing and efficient optimisation and selection strategies, instead relying only on straightforward hybridizations of data preprocessing and optimisation methods. With the aid of adaptive data preprocessing, complex optimisation strategies, kernel-based models, and the best model selection approach, this work develops a more precise model for predicting electricity prices. To be more specific, the best kernel-based ML models are developed using a leave-one-out optimization strategy based on the chaotic sine cosine algorithm. Effective data preprocessing outcomes are the aim of this method, which employs adaptive parameter-based variation mode decomposition. The first and last levels have nc and no neurons, which correspond to the second dimensions of X and Y, respectively. The other layers have (n1, • • •, n') neurons. With a grid search, these the hyper-parameter (', n1, • • •, n') are set. A gradient descent approach that reduces forecast errors to network weights is used to train the model.

Lago et al [12] LSTMDNN and GRU-DNN, two hybrid DNN, were proposed for predicting power spot prices for the following day. By including recurrent and regular layers, LSTMDNN models relationships between sequential time series and non-sequential data. A GRU layer was employed by GRUDNN, which trains more quickly than LSTM. EPEXBelgium dataset from January 1, 2010, to November 31, 2016, was used in this investigation. The sMAPE values for GRU-DNN and LSTM-DNN were 13.04 and 13.06, respectively.

Kuo et al [13] The suggested hybrid model was composed of CNN and LSTM layers. CNN first extracted the features, which were then provided to LSTM for forecasting. Historical 24-hour electricity prices were used as the model's input to generate the estimated price for the upcoming hour. The PJM Regulation Zone Preliminary Billing Data, which represents half-hourly capacity clearing rates for the regulation market in 2017, were used in the analysis. There had been 10 datasets used, with 3 months of data for each set and 1 month of data for testing. The average MAE for the suggested model was 8.85, which was lower than LSTM and CNN.

Qiao et al [14] A data preparation technique used in the prediction of electricity prices is the wavelet transform. This deterministic-based strategy doesn't provide reliable performance because the ordering and layers were selected based on experience. The results show that the suggested model outperforms existing AI models like BPNN in terms of forecasting accuracy. In the residential, commercial, and industrial electricity price scenarios, five order & four order five and seven layers, respectively, achieved the best WT-SAE-LSTM performance with MAPE values of 0.8606%, 0.4719%, and 0.4956%.

IV. PROPOSED WORK

In this research, various time series and machine learning models are used for prediction of 96 market clearing price of the next-day in day-ahead energy trading market. Concerning this issue, we started with two main research aims:

- How can energy price be predicted with highest accuracy, given the historical data is available?
- How can the underlying factors affect day-ahead energy prices?

Currently, machine learning on enormous datasets is frequently used for prediction; however, many ML algorithms require training data to be stored in memory on single machine, despite the fact that more scalable implementations of these algorithms have been proposed. Additionally, none of approaches are dynamic, domain-free, scalable, or accuracy-preserving. There are seven phases to the suggested strategy. The goals of first phase were to: a) collect real-time data from a general source; b) clean information and replace missing values; c) create a local ML model; d) apply dataset chunks to ML models for training and identify right models; and e) apply right model with new data. Various ML techniques are grouped as ensemble models to increase accuracy. In second stage of suggested methodology, we are using bagging regeressor with XGBoost as base estimator ensemble model to achieve high accuracy that is shown in below figure 1.

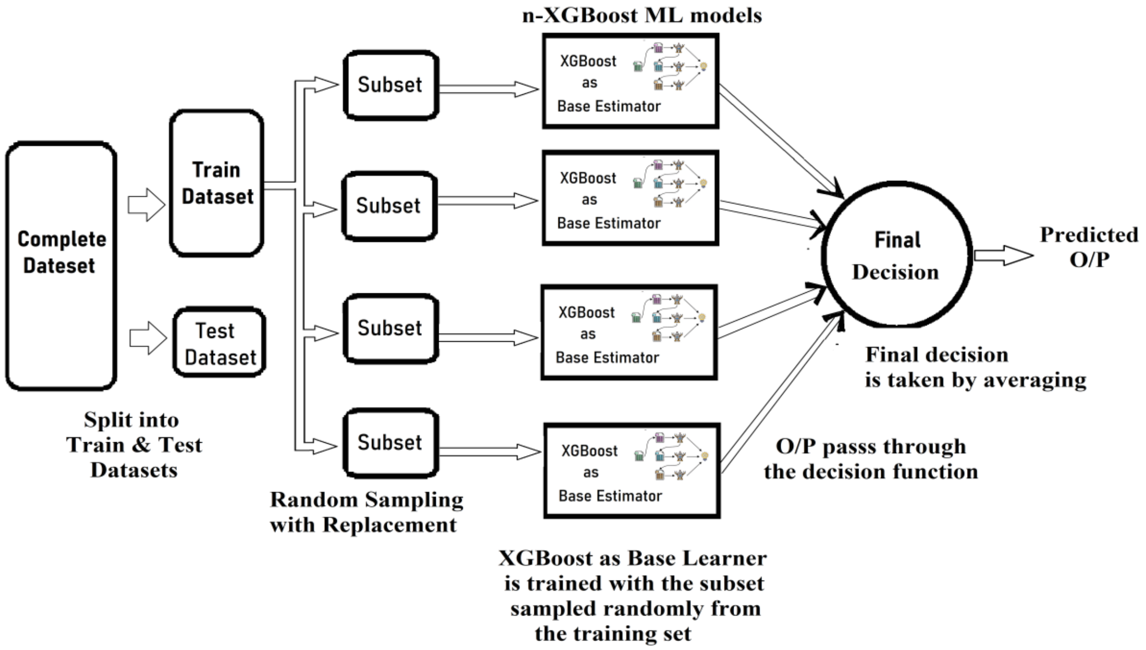

Figure 1: Proposed Bagging Regressor with XgBoost Regressor as base estimator model for EPF

We have contrasted final, combined, and trained model with original. Accuracy is preserved by this model. The suggested method outperforms all others in terms of efficiency.

V. RESULTS

Main goal of this research project is to quickly assess electricity prices in various markets using ML techniques. Features from entire data set are reduced, with some of features being eliminated as noise from entire data set, to increase computational speed and efficiency. Using feature selection methods, only crucial features are chosen. Classification problem is first used to evaluate prices, and it only provides information on prices that have an impact on market participants. Because marketers are unaware of prices beforehand, price volatility occurs. This could harm generating companies and have an impact on markets. Risk and bidding strategy to determine precise value of future prices are not reduced by identifying only prices. To lower risks, forecasting is done, and future price value is discovered. For average prices, electricity price forecasting is fairly accurate. They are unable to accurately respond to price spikes, though. However, for energy market participants to remain competitive in a competitive market, electricity price spikes are important. Results demonstrate the effectiveness in assessing performance by applying classification, forecasting, and spike prediction issues. The chart below (Fig.2) shows variations in electricity pricing during 2012 to 2021 as per dataset.

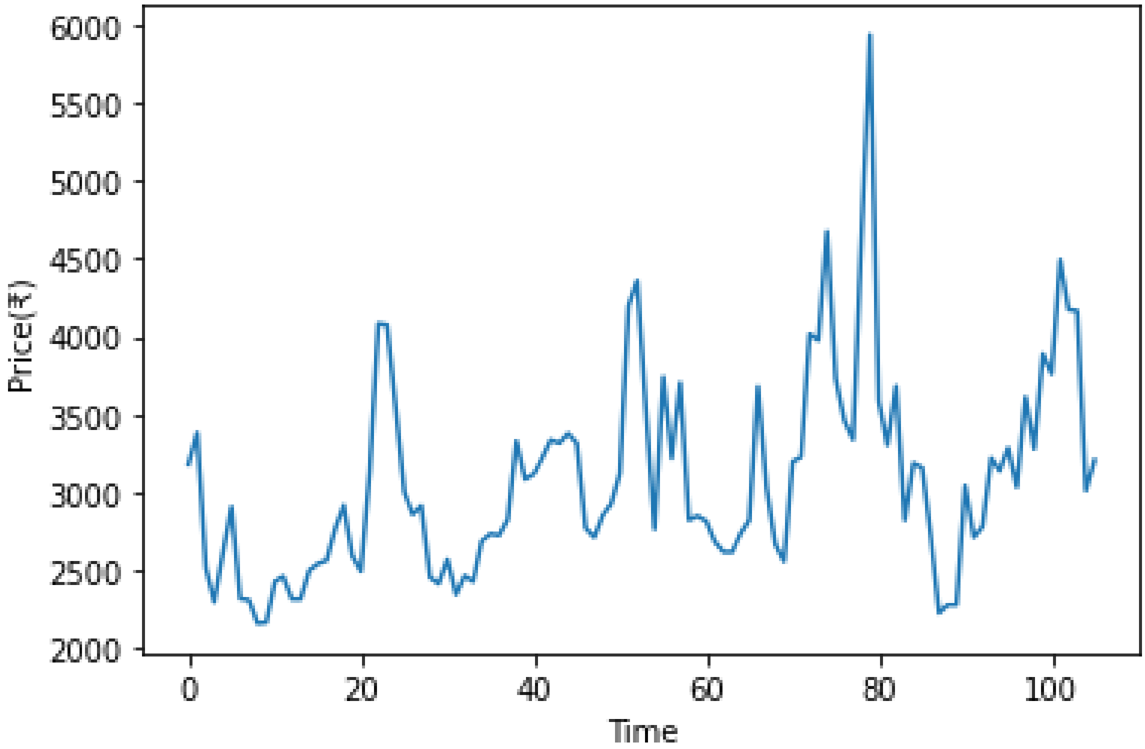

Figure 2: Electricity Price for 2013-2021

The below figure (Fig.3) shows decomposition of electricity price chart which helps in knowing Resid, Seasonality and Trend of price dataset. It can be clearly seen that in seasonal dataset, same pattern is repeating every year which means that our data is having seasonality. We can predict this by ARIMA and SariMax model.

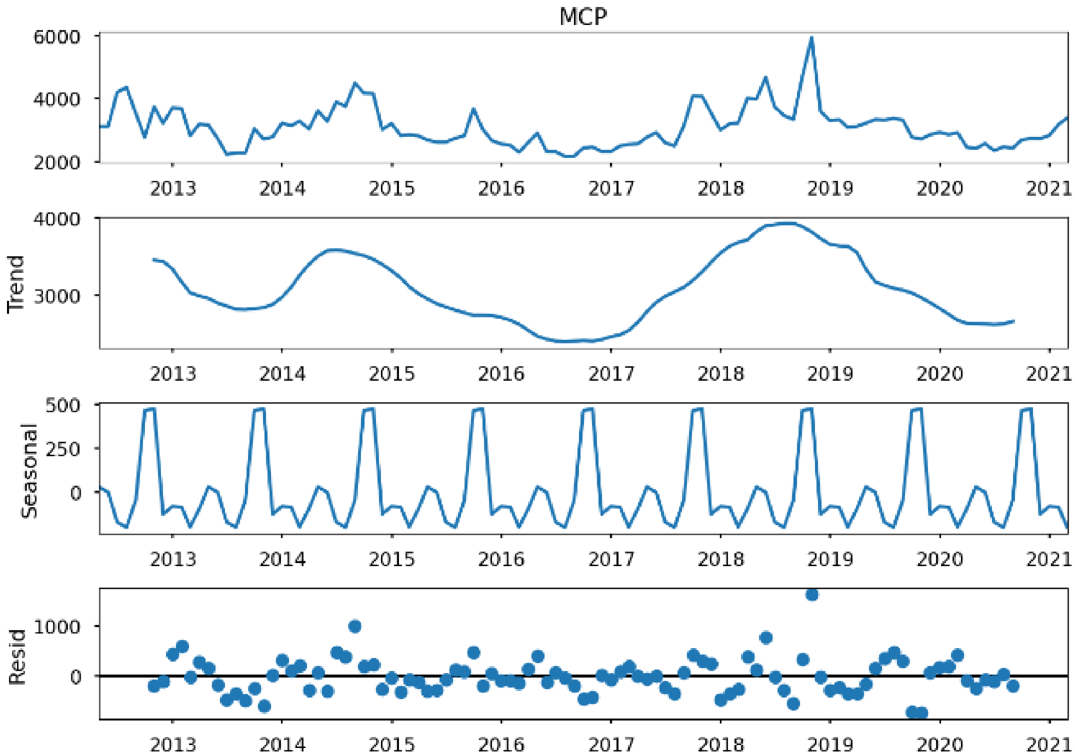

Figure 3: Trend Seasonality and Resid in MCP

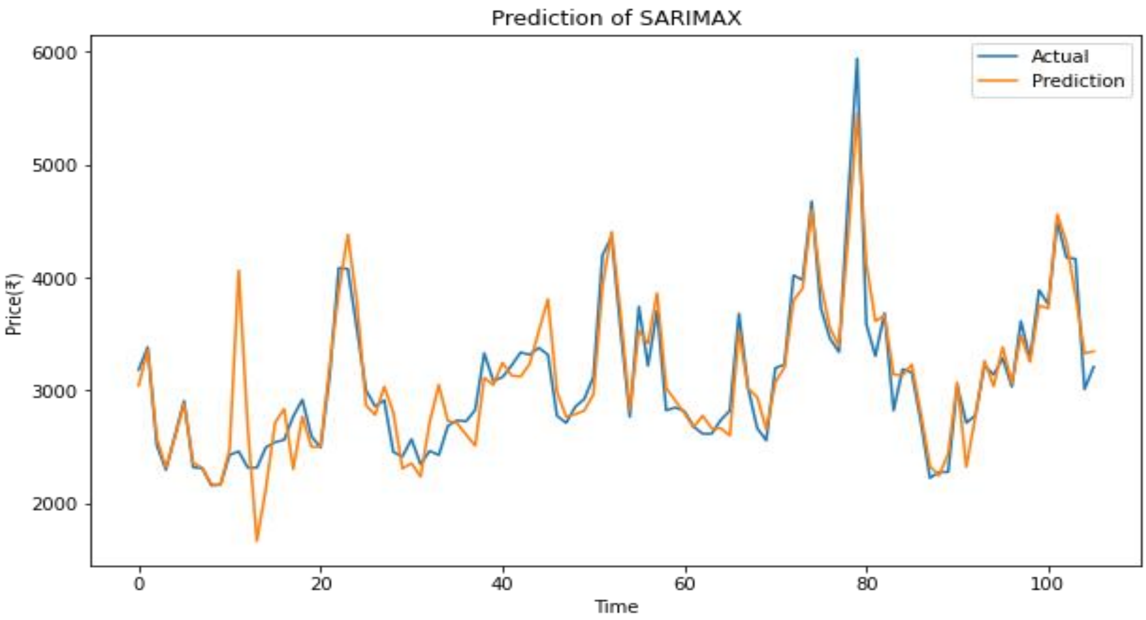

Figure 4: Price prediction of SARIMAX

The above figure (Fig.4) shows Prediction of SARIMAX model by plotting actual and predicted values. In the graph, Time is taken on x axis and price is taken on y axis to plot the values over the previous year data.

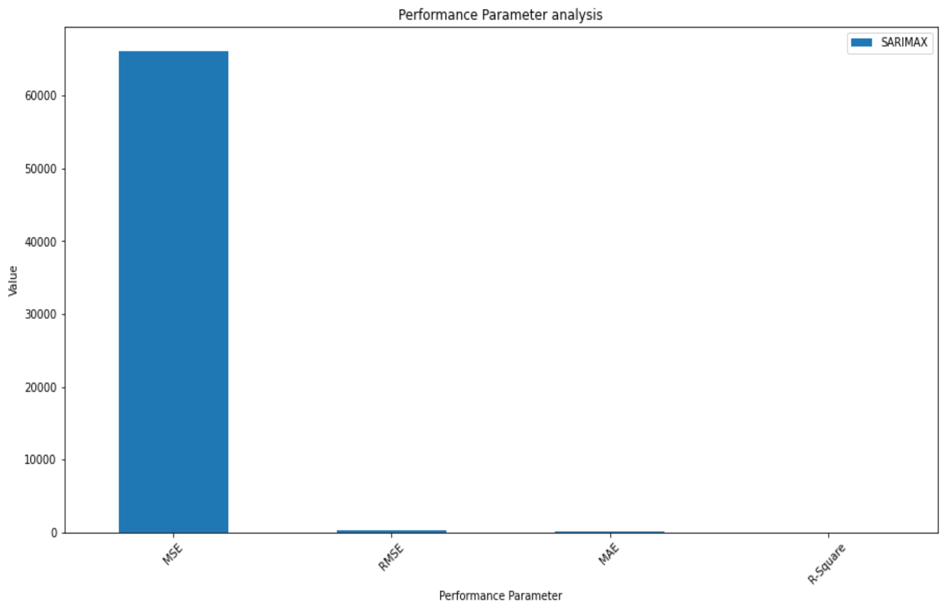

Figure 5: Performance parameters of SARIMAX

The above graph (Fig.5) shows performance parameters on x axis and value on y axis. The performance of SARIMAX model is evaluated with four evaluation metrics like Mean square error, Root mean square error, Mean absolute error and R-square. MSE has highest value (greater than 60,000), RMSE has 1000 value and MAE has lowest value (5000).

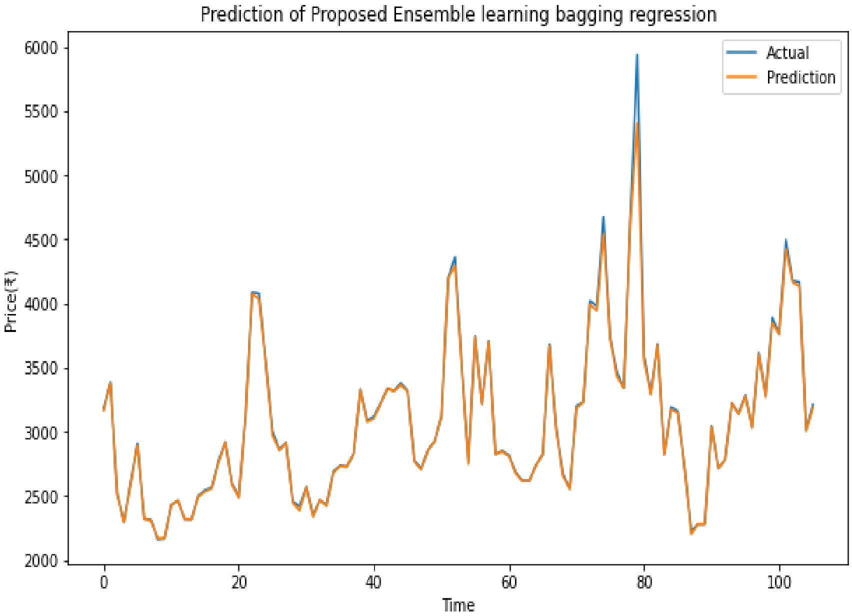

Figure 6: Prediction of proposed Bagging Regressor with XgBoost Regressor as base estimator

We have used Xgboost technique to predict electricity price. The above graph (Fig.6) shows actual values of dataset and predicted values as well. Price in rupees is taken on y axis and time is taken on x axis to plot the real and predicted values.

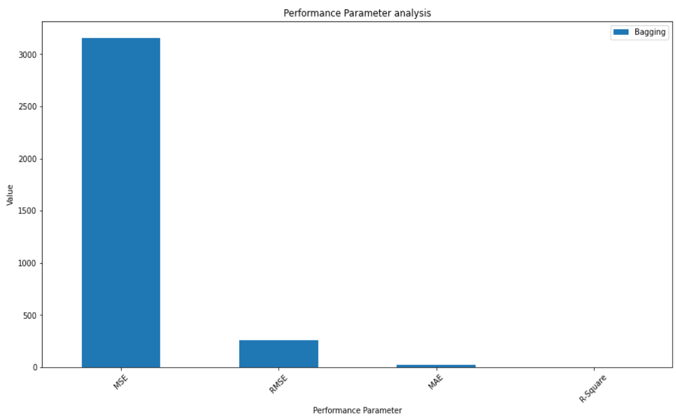

Figure 7: Performance Evaluation of proposed Bagging Regressor with XgBoost Regressor as base estimator

The above graph (Fig.7) shows values related to MSE, RMSE, MAE and R-square for XgBoost classifier. The error value for MAE, RMSE and MAE is 3000, 250 and 50 respectively as per performance parameter anlaysis.

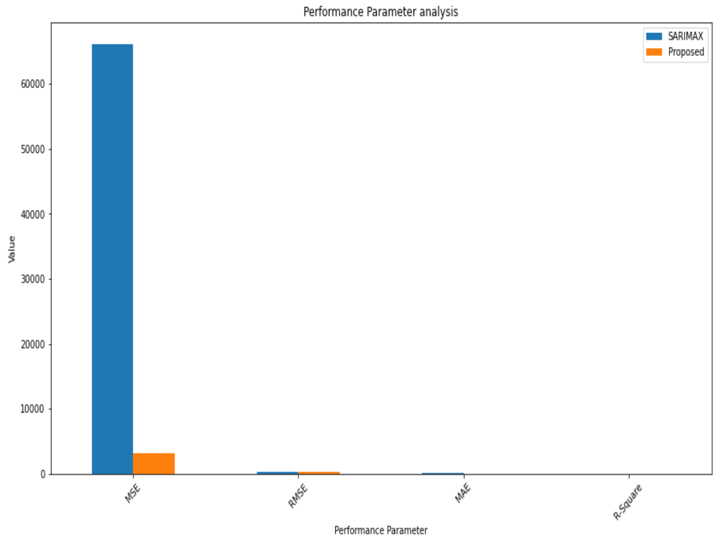

Figure 8: Comparison of Performance Parameters

The above figure (Fig.8) clearly shows variation of MSE and RMSE values in case of SARIMAX and proposed model. Both performance parameters have lesser value in case of proposed model when contrasted with existing model. Hence, we can say that our proposed model is better in terms of predicting electricity prices.

VI. CONCLUSION

Currently, machine learning on massive datasets is often utilised for prediction; however, despite the fact that more scalable implementations of these algorithms have been presented, many ML methods require training data to be stored in memory on a single machine. The uses for machine learning are numerous. Instead, it is also extending outside the energy sector. Because machine learning has such a wide application, some fields are where researchers are trying to revolutionize the world in the future. In this study, we only focus on forecasting energy prices. Additionally, a more accurate model for the same can be developed and utilized to forecast energy supply and demand. The findings of this research will help the next generation to understand the market and trade customs for the commodity of electricity. Additionally, this can help future researchers who use more complex technologies like deep learning to forecast energy costs, demand, and supply.

REFERENCES

[1] Bunn D, Andresen A, Chen D, Westgaard S. Analysis and forecasting of electricity price risks with quantile factor models. Energy J ,37:101–22,2016.

[2] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., and Arshad, H. State-of-the-art in artificial neural network applications: A survey. Heliyon 4, 11 ,2018.

[3] Amabile, L., Bresch-Pietri, D., El Hajje, G., Labbé, S., and Petit, N. Optimizing the self-consumption of residential photovoltaic energy and quantification of the impact of production forecast uncertainties. Advances in Applied Energy 2 ,2021.

[4] Anbazhagan, S., and Kumarappan, N. Day-ahead deregulated electricity market price forecasting using recurrent neural network. IEEE Systems Journal 7, 4 ,866–872,2013.

[5] Bergstra, J., and Bengio, Y. Random search for hyperparameter optimization. Journal of machine learning research 13, 2 2012.

[6] Breiman, L. Bagging predictors. Machine Learning 24, 2 , 123–140,1996.

[7] Burkart, N., and Huber, M. F. A survey on the explainability of supervised machine learning. J. Artif. Intell. Res. 70 ,245–317,2021.

[8] Krizhevsky, A., Sutskever, L., and Hinto, G. E. Imagenet classification with deep convolutional neural networks. Tech. rep., University of Toronto, 2012.

[9] Kampouraki, A., Manis, G., and Nikou, C. Heartbeat time series classification with support vector machines. IEEE Transactions on Information Technology in Biomedicine 13, 4 , 512–518,2009.

[10] Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., and Bagnall, A. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery, 35 401-449,2021.

[11] Yang, Wendong, Shaolong Sun, Yan Hao, and Shouyang Wang. "A novel machine learning-based electricity price forecasting model based on optimal model selection strategy." Energy 238 , 121989,2022.

[12] Lago, J., De Ridder, F., & De Schutter, B. , " Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms", Applied Energy, 221(February), 386–405,2018.

[13] Kuo, P.-H., & Huang, C.-J., "An electricity price forecasting model by hybrid structured deep neural networks",Sustainability, 10(4), 1280,2018.

[14] Qiao, W., & Yang, Z. (2020), "Forecast the electricity price of US using a wavelet transform-based hybrid model", Energy, 193, 116704