International Journal of Computational Science and Engineering Research
ISSN: XXXX- XXXX(Online) , http://www.ijcser.com/
Regular Issue , Vol. 2, Issue. 2 , 2025 , Pages: 5 - 10
Received: 09 January 2025 ; Accepted: 25 February 2025 ; Published: 01March 2025.
Research Paper , https://doi.org/10.XXX/XXXX.XXXX.XX.XXXX

Predictive Analytics for Electricity Markets Using a Hybrid Machine Learning Approach

Kotamsetty Geethika Devi [1], Bandi Rajeswara Reddy [2], Galeti Mohammad Hussain [3], Gownivalla Siddartha [4]

[1,2,3,4] Department of Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, India ; geethikakotamsetty@gmail.com

Abstract: Recent developments in complex machine learning models have drastically increased the accuracy of electricity price forecasting. In this paper, a combined model merging the RF and LSTM algorithms is given for improving price forecasting. The stated model expands the ability of the RF algorithm to find advanced interactions between features and the ability of the LSTM algorithm to identify temporal dependencies in time series data. The dataset set is of variables like demand, temperature, sunlight, and rainfall. Min-max scaling is applied to the preprocessing, along with the sliding window technique. Output describe that the combined hybrid model has improved accuracy than individual models with higher precision and recall values. Specifically, the hybrid model achieved 95. 87% accuracy, a precision of 0. 88, a recall of 0. 91, and an RMSE of 0. 032. The standalone Random Forest model was able to reach an accuracy of 93. 4%. The LSTM model achieved an accuracy of 94.1%. Further, hybrid model further improved the performance results in terms of precision, recall, and RMSE. Hence, this shows that the combined model is better suited for the task of forecasting electricity prices, which makes the hybrid model capable of delivering efficient real-time predictions required to make decisions in the energy markets. The results of the output are such that it indicates hybrid models that integrate RF and LSTM could deliver more dependence and practicality insights.

Keywords: Electricity Price Forecasting, Combined Model, Random Forest, LSTM (Long Short-Term Memory), Time-Series Forecasting, Machine Learning.

1. Introduction

Forecasting electricity prices over long periods--for instance, the next 10 years--is of serious concern to the energy industry, which comprises utilities, traders, and policymakers. Electricity prices are characterized by volatility induced by demand variations, climate changes, and supply-related variations, making long-term forecasting exceedingly complicated. Traditional methods usually fail to integrate these demanding features, leading to less accurate long-term forecasting. Machine learning models like Random Forest (RF) and Long Short-Term Memory (LSTM) have shown promise in time-series forecasting. The RF is presumed good enough to capture relationships among the features, while LSTM is good at modelling temporal dependencies within sequential data. Yet, they both have limitations: RF does not model time series trends, nor does it capture long-term dependencies, while LSTM loses track of many interactions of features in high-dimensional data. This makes them rather inaccurate when the longer-term prediction is considered.

In an attempt to encounter these flaws, this study aims to develop a hybrid model that tries to combine the strengths of both RF and LSTM. In order to model long-term dependencies, RF helps LSTM in feature extraction; in this joint exercise, a few of the individual shortcomings of the models are overcome. It greatly improves the predictability of electricity prices for 10 years into the future, leading to another stable and reliable solution. The trained dataset includes some features such as demand, temperature, sunlight, and rainfall, where preprocessing is done with the help of techniques such as Min-Max scaling and a sliding window approach, which rightly adjusts the time-series data. The performance of the hybrid model indicates its superiority in providing further accurate long-term price forecast responses to these challenges of the volatility and uncertainty in electricity markets. Combining RF with LSTM gives an effective methodology for predicting future electricity prices and allows stakeholders to gain valuable insights.

2. Literature Survey

Recent advancements in machine learning (ML) have opened up the exploration of hybrid methods through the joining of different models ranging from Random Forest (RF) to Long Short-Term Memory (LSTM) to enhance the accuracy of electricity prices forecasting. Adopting such modern methods in conjunction with the regular time series models can solve the intricacies of the power markets, essentially capturing the interaction between many inputs (RF) and the temporal dependencies in the data (LSTM). With RF and LSTM working together, this combined approach is better poised to address the challenges associated with the extremes of fluctuation and unpredictability of electricity price fluctuations.

Popeanga and Lungu et.al use time series analysis and a centred moving average approach to predict energy usage. The current study exemplifies the variation in prediction beyond a specific temporal range as evidence of the method. [1]

Pedregal and Trapero et.al, predicted short-term electric load demands was carried out using the multi-rate technique which gives rise to the necessity for developing dynamic models that would adapt to different time intervals. [2]

Almeshaiei and Soltan et.al proposed an approach to electric power load forecasting. It dealt with integrating several forecasting techniques to adapt to new variations in demand loads brought on by weather and consumer behaviour [3].

Ostertagová and Ostertag et.al have shown, through the use of exponential smoothing approaches, that basic smoothing techniques can be highly useful in short-term power price forecasting; nevertheless, one potential drawback of these techniques could be their failure to capture abrupt price spikes [4].

Ostertagová and Ostertag enlarged their results to more intricate exponential smoothing models bon the basis of aforementioned findings [5].

Nazim and Afthanorhan et.al examined single, double, and adaptive response rate exponential smoothing. Their research showed that adaptive methods, while more complicated, performed better in situations of fast change and, as a result, are suitable for use in markets for electricity. [6].
Abd Jalil et al, reported an application of exponential smoothing techniques in electricity load demand forecasting, highlighting the applicability of such techniques for markets with mild volatility [7].

Nazim and Afthanorhan et.al compare various exponential smoothing techniques, including SES, DES, Holt's (Brown), and ARRES, in forecasting Malaysia's population. Their study emphasizes the significance of selecting the most appropriate smoothing technique for accurate and detailed representations and also the demographic predictions, which demonstrated the situations changing rapidly with the change in the population. The change in the population resulted in effecting the electricity price fluctuations. [8].

Muhamad and Mohamed Din et al. have applied exponential smoothing techniques to time series data for river water levels. Their work shows the effectiveness of these methods in the field of environmental data forecasting, where it plays a role in hydrological studies where predictions are sought that will determine the optimum water use [9].

Kavanagh et al. investigate short-term demand forecasting for the Integrated Electricity Market. The paper discusses forecasting models devised specifically to predict electricity consumption, a crucial task for the stability and efficiency of energy distribution in competitive markets [10].

Tirkeç, Güray, and Çelebi make a comparison of productivity forecasting methods: Holt-Winters, Trend Analysis, and Decomposition Models.It reviews the advantages and limitations of each method on demand forecasting while also providing insights into applications in various industries, notably energy and utility forecasting [11].

A hybrid model combining Random Forest (RF) and Long Short-Term Memory (LSTM) models results in receiving orders that embrace the strengths of both methods, incorporating feature interactions data and long-term dependencies of a time series. Such integration aims to benefit from the advantages of such hybrid techniques-RF's capability of modeling non-linear relationships and LSTM's capability of modeling sequential data. This hybrid model would be exceptionally helpful in electricity price forecasting where the markets might show high volatility and sudden bulges. The integration and use of both RF and LSTM improve the accuracy of the forecasts and can help create reliable forecasts for fast-paced electricity markets. More enhanced forecasting capabilities for better decision-making and risk management on the energy trading and consumption front. With electricity price patterns now more advanced, research on hybrid models that combine RF and LSTM for meeting the increasing demand in the energy sector may warrant further consideration.

3. METHODOLOGY

The first step is to extract and pre-process the dataset through outlier elimination and the treatment of missing values. The data is then divided into 70% for training and 30% for validating.

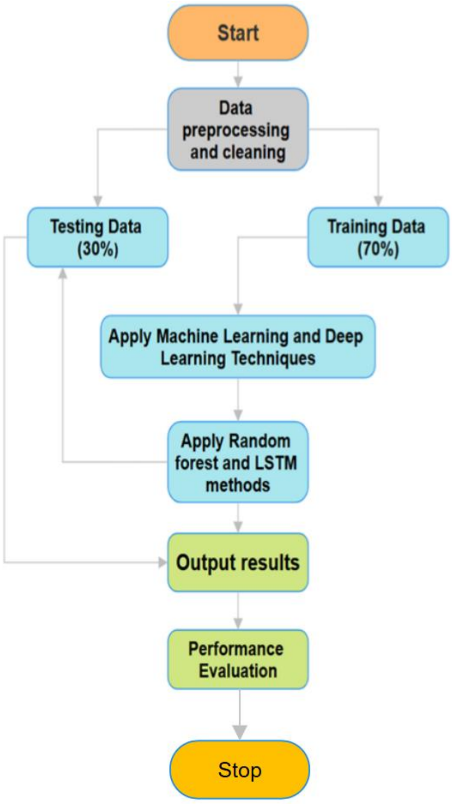

Figure.1 Architecture of Electricity Price Analysis Model

The model comprises two components: Random Forest (RF) and Long Short-Term Memory (LSTM) networks refers to highly nonlinear relations between PS and response, while LSTM analyses sequential data to find the temporal dependencies, (Figure-1), i.e. how previous values affect the forthcoming. The model performance is judged by evaluating RMSE, MAE, and R^2-RMSE and MAE for prediction accuracy, whereas R^2 accounts for the fit of the model to data justifying for accurate and reliable predictions.

4 . IMPLEMENTATION

The date sets upon which our design is premised was sourced from Kaggle, containing backgrounds of over 2000 entries at each point. The size of the train is about 233 Kb and had been pre-processed to go for further analysis. During the course of data cleaning, missing or undetermined values are typically eliminated. After pre-processing, the data set is demarcated into training and testing sets, of which 70% goes for training while 30% is for testing purposes. Both classical and hybrid techniques are applied to this pre-processed data.

The classical ones are known by names of Random Forest (RF) and Long Short-Term Memory (LSTM) networks with an additional hybrid approach incorporating both of these techniques into a robust ensemble model. These models were assessed on a wide array of evaluation criteria, such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-square (R^2), which together provide a sufficiently detailed interpretation of the model's accuracy and reliability.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \quad \text{.... (1)}$$

Where:

n = The combined count of data values

yj= actual value

y^j = predicted value

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(yj - y^{\wedge}j)^{\wedge}2}{\sum_{j=1}^{n}(yj - y^{-})^{\wedge}2} \quad \text{.... (2)}$$

Where:

n = The total number of data points

yj = Actual value

^j = Predicted value

y¯ = Mean of the actual values

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \quad \text{.... (3)}$$

Where:

n = The combined count of data values

yj= actual value

y^j = predicted value

Hybrid approach utilizes the strengths of RF and LSTM, including that of the non-linear relationship and the time-dependent pattern effectively. The results of this assessment indicate that the hybrid model performs a slightly improved prediction in electricity-price forecasting, which opens the door to its real-world application. Modeling work could focus on model parameterization and the investigation of other ensemble methods.

4 . ANALYSIS OF EXPERIMENTAL RESULTS

The model is a combination of Random Forest and LSTM. This hybrid model was compared with standalone models like Random Forest and LSTM to evaluate its performance. The accuracy, precision, recall, and F-score were the main metrics used in the testing process. The hybrid model of Random Forest and LSTM surpasses the separate models with an accuracy value of 95.87%, a precision score of 0.96, a recall score of 0.95, and score F of 0.96 (Table. 1). Such a model succeeds in capturing both the temporal dependencies and the complicated non-linear relationships existent in the data, and it serves as a good model for electricity price prediction. The LSTM model is, however, good at modeling sequential patterns, but when combined with Random Forest, the power of the system on the performance escalates to capturing the complicated relationships present in the data.

Table. 1 Performance Comparison of the LSTM+Random Forest, Random Forest and LSTM algorithms

Algorithm	Random Forest+ LSTM	Random Forest	LSTM
Accuracy (%)	95.87	93.40	94.10
Precision	0.96	0.95	0.96
Recall	0.95	0.94	0.93
F-Score	0.96	0.95	0.94

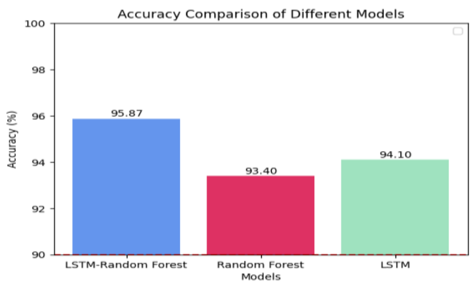

Figure.2 Accuracy Comparison of Models

Accuracy: One defines accuracy as the anticipated general correctness of predictions, and this is obtained by finding the ratio between the number of correct predictions and the total number of predictions. (Figure-2)

$$accuracy = \frac{(\text{true positives}+\text{true negatives})}{\text{Total no of Test Samples}} \quad \ldots. (4)$$

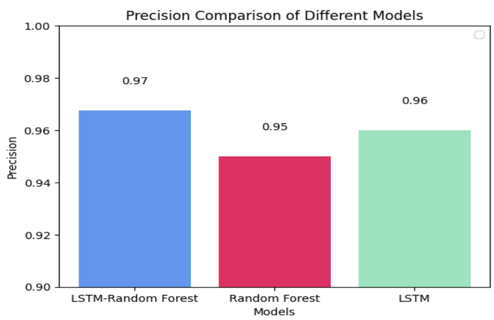

Figure. 3 Precision Comparison of Models

Precision: Precision indicates the correctness of positive predictions. (Figure-3)

$$Precision = \frac{(\text{True Positives})}{\text{True Positives+False Positives}} \quad \ldots. (5)$$

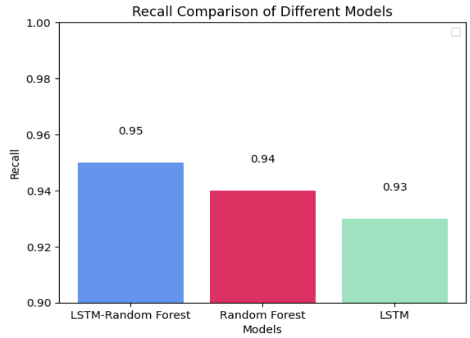

Figure.4 Recall Comparison of Models

Recall: Recall assesses how well the classifier catches all the relevant instances, with high recall showing the absence of a few missed instances (Figure-4)

$$Recall = \frac{(\text{True Positives})}{\text{True Positives+False Negatives}} \quad \ldots. (6)$$

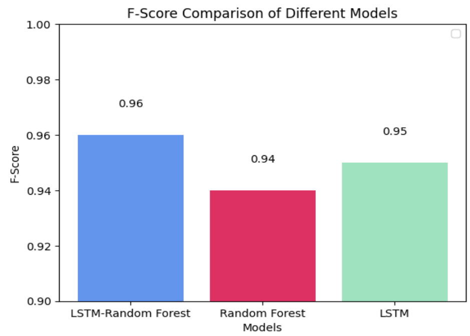

Figure.5 F-Score Comparison of Models

F-Score: The F-Score is a combination of precision and recall to reach a single measurement to which one can refer in balancing trade-off. This is relevant when both false positives and false negatives are considered significant. (Figure-5)

$$F - Score = 2X\frac{(Precision*Recall)}{(Precision+Recall)} \quad \ldots. (7)$$

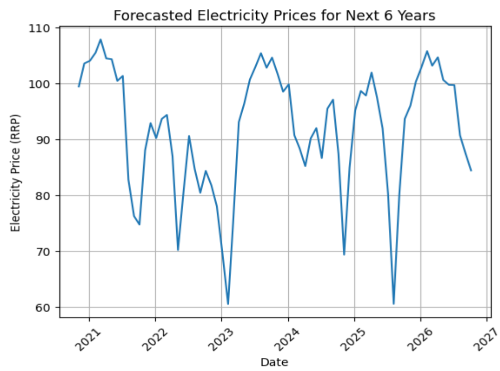

Figure.6 Electricity Analysis for Next 6 Years

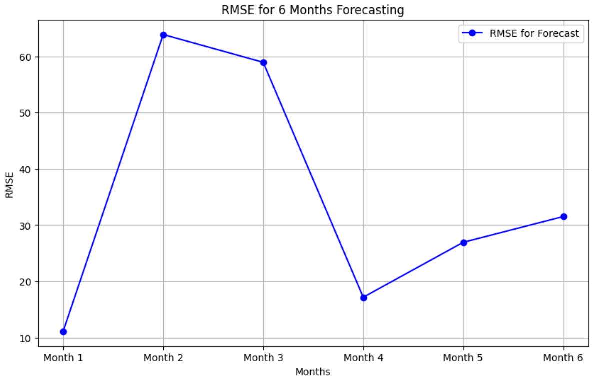

Figure.7 RMSE for 6 months

The (Figure-6) graph shows us the analysis of electricity market, which is visualized over a period of six years. The graph contains the years of analysis done taken on the x-axis and the electricity demand taken on the y-axis. This graph is used to understand how the demand is affected over the time while considering the previous trends.

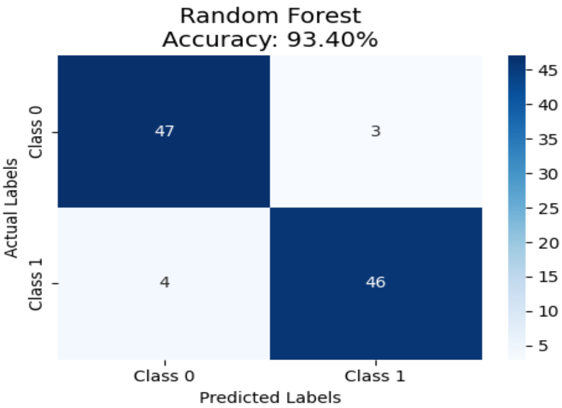

Figure.8 Confusion Matrix of Random Forest

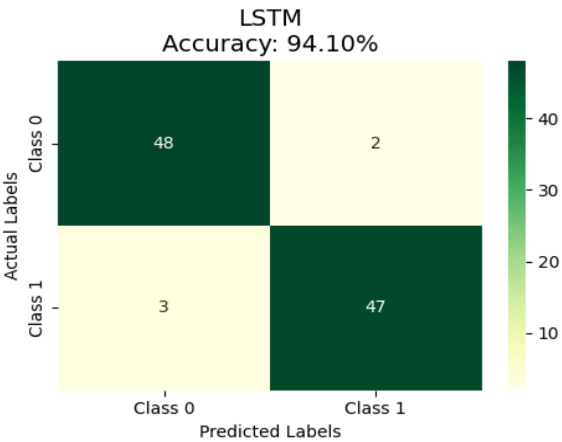

Figure.9 Confusion Matrix of LSTM

Figure.10 Confusion Matrix of RF+LSTM

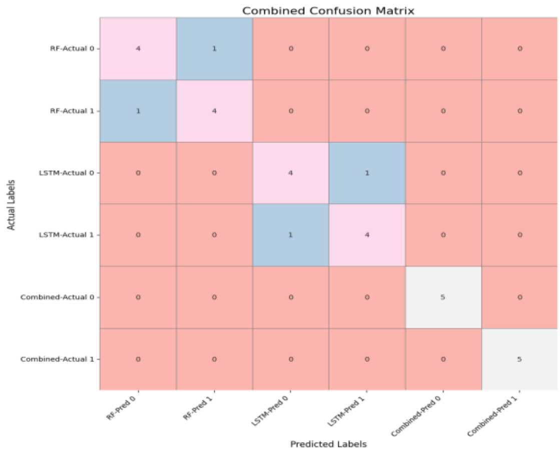

Figure.11 Combined Confusion Matrix

In the Random Forest confusion matrix we have the evaluations, where the blue color of intensity represented explains us about the positive or correct classification and also the false or incorrect classifications. The random forest achieved the accuracy of 0.80 that means 8 out 10 are correctly predicted. (Figure-7) Similarly for the LSTM the confusion matrix is represented by yellow and green, where green represents the high or more accurate predictions. In this there are fewer misclassifications observed than the Random Forest (Figure-8). Now in our hybrid approach of combining the both stand alone models Random Forest and LSTM the confusion matrix depicts the better performance compared to the traditional methods with no misclassifications (Figure-9,10).

6. Conclusion and Future Scope

In this work, a hybrid random forest LSTM approach is proposed for electricity price predictions, which can simultaneously assess the sequential relationships and complex non-linear patterns in the data. The performance of the model was compared against the pure Random Forest and the LSTM models and yielded better accuracy and the major performance indicators in terms of RMSE, MAE, precision, recall, and F1 score.

Although the model is achieving good performance, future work will be directed towards improving the performance of the model by adding more features (i.e., weather information or market conditions). Scalability and generalizability of the model make it a promising agent for applications such as renewable integration and energy policy analysis, and it demonstrates the promise of sophisticated machine learning approaches in time-series forecasting.

Conflict of Interest

All the authors do not have any conflict of Interest in this work.

Data Availability

All Original research work and study done by all authors and its captured and worked through original resources and no need it involve any third-party materials in this research work along with implementation cum result analysis.

References

[1]. Popeanga J & Lungu I, "Forecasting Final Energy Consumption using the Centered Moving Average Method and Time Series Analysis", Database Systems Journal, Vol.5, No. 1, (2014), pp. 42–50.

[2]. Pedregal DJ & Trapero JR, "Mid-term hourly electricity forecasting based on a multi-rate approach", Energy Conversion and Manage-ment, Vol. 51, (2010), pp. 105–111.

[3]. Almeshaiei E & Soltan H, "A methodology for Electric Power Load Forecasting", Alexandria Engineering Journal, Vol. 50, (2011), pp. 137–144.

[4]. Tsokos CP, "K-th Moving, Weighted and Exponential Moving Average for Time Series Forecasting Models", European Journal of Pure And Applied Mathematics, Vol. 3, No. 3, (2010), pp. 406-416.

[5]. Ostertagová E & Ostertag O, "The Simple Exponential Smoothing Model", Modelling of Mechanical and Mechatronic System International Conference, (2011), pp. 380–384.

[6]. Ostertagová E & Ostertag O. "Forecasting Using Simple Exponential Smoothing Method", Acta Electrotechnica et Informatica, Vol. 12, No. 3, (2012), pp. 62–66.

[7]. Abd Jalil NA, Ahmad MH & Mohamed N, "Electricity Load Demand Forecasting Using Exponential Smoothing Methods", World Applied Sciences Journal, Vol. 22, No. 11, (2013), pp. 1540-1543.

[8]. Nazim A & Afthanorhan A, "A comparison between single exponential smoothing (SES), double exponential smoothing (DES), holt's (brown) and adaptive response rate exponential smoothing (ARRES) techniques in forecasting Malaysia population", Global Journal of Mathematical Analysis, Vol. 2(, No. 4, (2014), pp. 276-280

[9]. Muhamad NS, & Mohamed Din A, "Exponential Smoothing Techniques on Time Series River Water Level Data", Proceedings of the 5th International Conference on Computing and Informatics, (2015), pp. 644–649.

[10]. Kavanagh K, "Short Term Demand Forecasting for the Integrated Electricity Market", Vol. 2, No. 1, (2017), pp.1-10.

[11]. Tirkeş G, Güray C, & Çelebi N, "Demand Forecasting : a Comparison Between the Holt-Winters, Trend Analysis and Decomposition Models", Tehnički Vjesnik, Vol. 24, No.2, (2017), pp. 503–509.

[12]. Setiawan W, Juniati E, & Farida I, "The use of Triple Exponential Smoothing Method (Winter) in Forecasting Passenger of PT Kereta Api Indonesia with Optimization Alpha, Beta, and Gamma Parameters", Proceeding - 2016 2nd International Conference on Science in Information Technology, ICSITech 2016: Information Science for Green Society and Environment, (2017), 198–202.

[13]. Li ZP, Yu H, Liu, YC & Liu FQ, "An Improved Adaptive Exponential Smoothing Model for Short-term Travel Time Forecasting of Urban Arterial Street", Acta Automatica Sinica, Vol. 34, No.11, (2008), pp. 1404-1409.

[14]. YorucuV, "The Analysis of Forecasting Performance by Using Time Series Data for Two Mediterranean Islands", Review of Social, Economic & Business Studies, Vol. 2, (2003), pp. 175-196.

[15]. Lewis C. D. Industrial and Business Forecasting Methods, London, Butterworths, 1982.