Comparative Analysis of ML Models for Electricity Price Forecasting

Malti Bansal*, Aditya Raj, Aman Raj
Department of Electronics & Communication Engineering,
Delhi Technological University (DTU), Delhi- 110042, India
*maltibansal@gmail.com

Abstract— **Precise forecasting of electricity prices is crucial for optimizing energy consumption, ensuring market stability, and facilitating well-informed decision-making within the electricity sector. However, the intricate dynamics and inherent volatility of electricity markets present significant challenges to conventional statistical methods. Consequently, machine learning (ML) algorithms have emerged as powerful tools for addressing this complex forecasting task. This study undertakes a comprehensive comparative analysis of prominent ML algorithms for electricity price forecasting, assessing their performance across various forecasting horizons and market conditions. The research investigates the comparative performance of 18 ML algorithms, focusing primarily on regressors such as the Random Forest Regressor, Extra Trees Regressor, LGBM Regressor, and Decision Tree Regressor, among others. To thoroughly evaluate the forecasting accuracy and robustness of each model, a variety of evaluation metrics are employed, including mean absolute error (MAE) and root mean squared error (RMSE). An important aspect of this research involves examining exogenous factors, such as weather patterns and electricity generation sources, and their influence on the forecasting performance of ML algorithms.**

Keywords— **Electricity price forecasting, Machine learning algorithms, Regressors, Random Forest Regressor, Extra Trees Regressor, Light Gradient Boosting Machine Regressor, Decision Tree Regressor.**

I. INTRODUCTION

Imagine a world without electricity—not just the flip of a switch, but the absence of modern conveniences [1]. Now, envision this force as unpredictable, its cost fluctuating like a stormy sea, leaving you in perpetual scramble. This reality lacks accurate electricity price forecasting, essential for guiding budgets to national policies [2]. For Energy Consumers, budgeting becomes a tightrope walk. Without reliable forecasts, planning energy usage feels like navigating a minefield. Each appliance switch or flick of the light could hide a budget-busting surprise [3]. Precise forecasts empower you to predict bills, optimize usage, and save hundreds annually [4]. Grid operators face a balancing act, where reliability is paramount. Keeping lights on isn't just flipping switches; it's a dance between supply and demand, with price as a key player. Accurate forecasts enable operators to anticipate fluctuations, manage reserves, and ensure a stable, reliable power supply, reducing blackout risks and bolstering trust [5, 6]. For investors, the market is a maze. Investing in energy resembles sailing uncharted waters. Clear price forecasts are crucial for predicting project profitability, assessing trends, and making sound decisions [7]. Accurate forecasts act as a lighthouse, guiding investors to promising ventures like solar or wind farms, mitigating risks, and fostering growth [2]. Policymakers face a guessing game without reliable predictions. Shaping the energy landscape demands foresight. Accurate forecasts lay the foundation for crafting effective policies, incentivizing renewables, and ensuring a secure future [5, 6]. Electricity price forecasting isn't merely digits on a screen; it's the unseen force shaping our energy reality. It differentiates between chaos and prosperity, between a teetering grid and dependable power. It's the compass guiding investors and policymakers toward a sustainable future [1, 7]. In essence, accurate electricity price forecasting is essential. It empowers individuals, stabilizes markets, guides investments, and shapes our energy future.

II. ELECTRICITY MARKET SCENARIO

The global electricity market is evolving rapidly due to environmental concerns, technological advancements, and changing consumer preferences. This dynamic presents significant challenges and opportunities for the energy sector. Key trends highlight the need for dynamic pricing to adapt to market changes. Decarbonization, driven by a shift to cleaner energy sources like solar and wind, disrupts traditional market structures. Consumers are increasingly participating in energy production through decentralized resources, necessitating innovative grid solutions. Fluctuating fuel prices and extreme weather contribute to electricity price fluctuations, requiring flexible grid management. Evolving regulatory landscapes pose challenges for electricity suppliers but also offer opportunities. Empowering consumers through Demand-Side Management with smart grids and dynamic pricing schemes can enhance grid efficiency and reduce peak demand, ensuring a sustainable market. However, efficient energy consumption scheduling relies on consumer awareness of future electricity prices. This scheduling of energy consumption can be efficiently achieved only if the consumers are aware of the future electricity prices.

III. DATASETS

Delving into electricity price forecasting requires a solid dataset foundation. We utilize two datasets, each offering insights into energy price dynamics in Spain: Weather Features Dataset and Energy Dataset.

Fig. 1: Map Highlighting 5 major cities of Spain

IV. PREPROCESSING OF DATA

The preprocessing of data included the handling of the two datasets first on an individual level and

then by merging the two.

Energy Dataset: As a first step, the unnecessary columns were dropped followed by the handling of duplicates and Null values. The Null values were predicted by the use of Linear Interpolation and the duplicates were handled by only keeping the first occurrence of the set of values.

Weather Dataset: The same procedure was followed for the weather dataset, however involving an extra step which was concerned with keeping the weather features separated for the 5 cities in order to filter the data such that it has the same number of observations for each city. Handling of outliers was accomplished by first setting up their values to Null which were later replaced by using Linear Interpolation.

After having dealt with the datasets individually, the datasets were merged into a single data frame.

V. FEATURE ENGINEERING AND SELECTION

Feature Engineering: In the realm of feature engineering, several enhancements were introduced to the dataset. This involved generating features denoting the hour, day, and month, alongside incorporating business hours in Spain, diverging from the conventional 9 AM-5 PM norm.

Feature Selection: The process of feature selection entailed leveraging the sklearn library in Python to identify the top k features. Through iterative testing to minimize error, a value of 25 was determined for k, signifying the most salient features for predicting the values of 'price actual’.

VI. MACHINE LEARNING MODELS

Linear Regression: Linear regression, a cornerstone of machine learning, uncovers linear relationships between a dependent variable (y) and one or more independent variables (X). This algorithm employs a linear equation to model relationships, facilitating trend understanding and predictions.

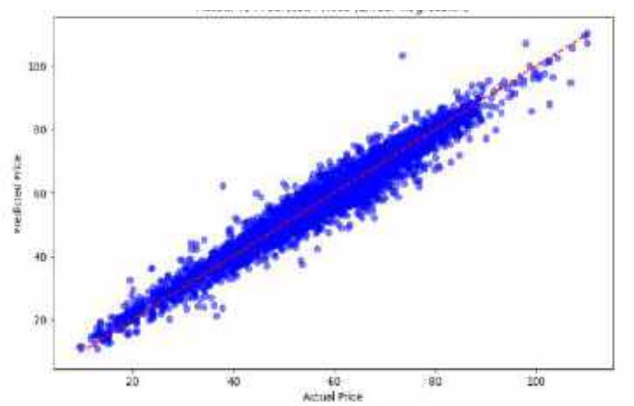


Fig 2: Scatter plot - LR

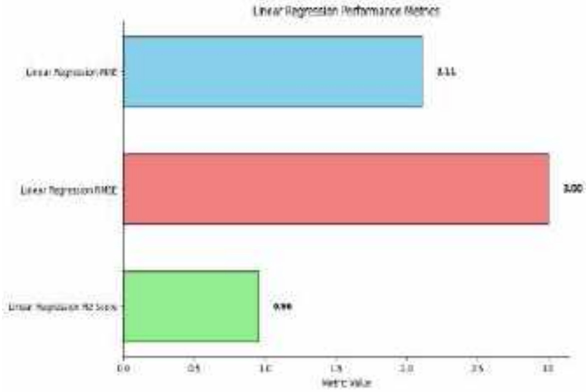


Fig 3: Performance metrics of LR

The MAE suggests a decent level of accuracy at 2.11, while the RMSE of 3.00 indicates greater variability in prediction errors compared to the MAE. This variability suggests some predictions were significantly off, possibly due to the model's sensitivity to outliers or complex dynamics. The R-squared score of 0.955 indicates that the LR model captured a significant portion of the variance in electricity prices, explaining a large portion of price fluctuations based on the included features.

Random Forest Regressor: The RFR demonstrated high accuracy in electricity price forecasting, capturing a substantial portion of price variance compared to baseline models [4]. Its ability to handle complex nonlinear relationships significantly impacted results [10]. By combining multiple decision trees, the RFR effectively learned intricate patterns and interactions among features, leading to more realistic forecasts [10].

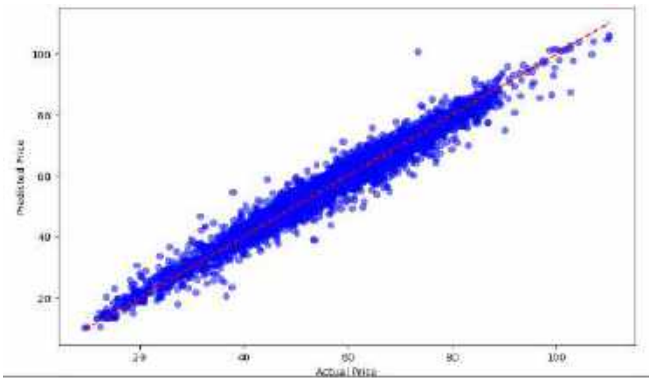


Fig 4: Scatter Plot - RFR

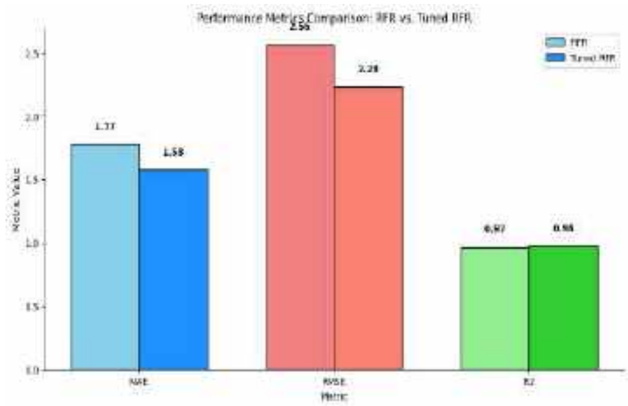


Fig 5: Performance Metrics - RFR and Tuned RFR

The tuned RFR has the best performance among all models considered under this research with the best metrics achieved for all three parameters of MAE, RMSE and R^2.

Tuning Parameters: N estimators [100, 200, 300], Max Depth [None, 5, 10, 20], Minimum Samples Split [2, 5, 10], Minimum Samples Leaf [1, 2, 4]. Tuning was achieved by using the GridSearchCV method.

Decision Tree Regressor: Metrics show that the DTR achieved moderate accuracy compared to the LR model. While MAE and RMSE are slightly higher [14], the R2 score remains relatively high, indicating that the DTR captured a significant portion of the variance in electricity prices [17]. The DTR offers a good balance between accuracy and interpretability for electricity price forecasting, effectively capturing non-linear relationships and providing valuable insights into price-influencing factors. However, its susceptibility to overfitting necessitates careful model selection and regularization techniques. When the model is tuned, the Mean Average Error exhibits a significant 16.2% reduction from 2.59048 to 2.17285, indicating a lower average error in predictions and improved model performance. Likewise, for RMSE, there is a notable 17.4% decrease from 3.73104 to 3.08097, suggesting reduced variability in prediction errors and better handling of outliers and complex dynamics. Although the increase is marginal for the R2 value, from 0.93120 to 0.95309, it still signifies an improvement in capturing the variance in electricity prices.

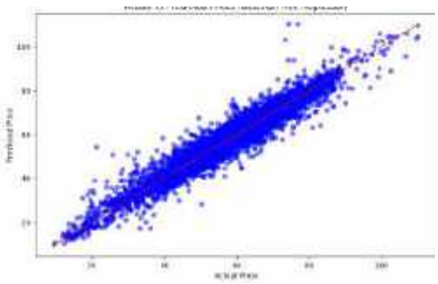


Fig 6: Scatter Plot - DTR

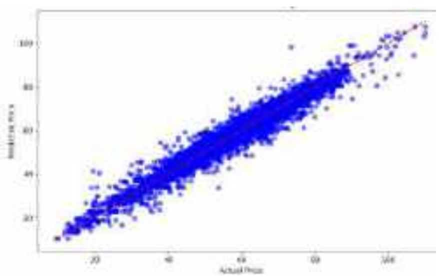


Fig 7: Scatter Plot - DTR (T)

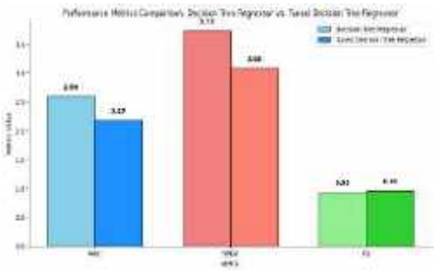


Fig 8: Performance metrics - DTR and DTR (T)

Tuning Parameters: Max Depth [5, 10, 15], Maximum Samples Split [2, 5, 10], Minimum Leaf Samples [1, 2, 4]. Tuning achieved by using the GridSearchCV method.

Gradient Boosting Regressor: By combining multiple weak learners, GBR achieves superior accuracy compared to individual models, with lower MAE (1.79688) and RMSE (2.56107). Its iterative learning process enables it to focus on areas where previous trees performed poorly, leading to improved error correction and increased robustness to outliers and complex dynamics. GBR assigns weights to features based on their importance in the final model, offering valuable insights into factors influencing electricity prices. Building on the impressive performance of the GBR, the tuned version demonstrates further improvements in accuracy, consolidating its potential for reliable electricity price forecasting. A significant 7.5% reduction in MAE compared to the original GBR (from 1.79688 to 1.66151) indicates even more precise average predictions. RMSE sees a 7.9% decrease from the original GBR (from 2.56107 to 2.36961), showcasing significant improvement in handling outliers and complex dynamics. Additionally, the R2 metric reaches an impressive 0.97283, surpassing the original GBR's 0.96758, indicating an even greater ability to capture variance in electricity prices.

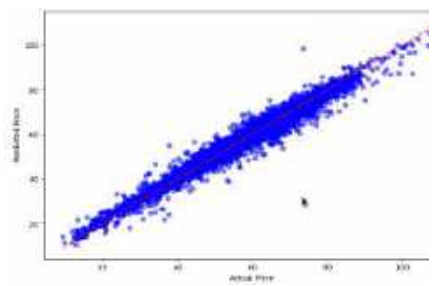


Fig 9: Scatter Plot GBR

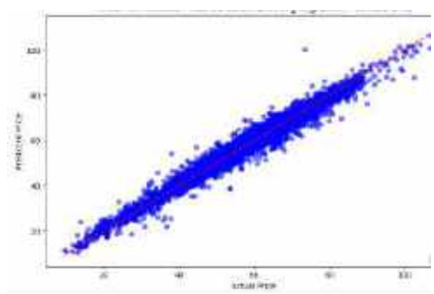


Fig 10: Scatter Plot - GBR Tuned

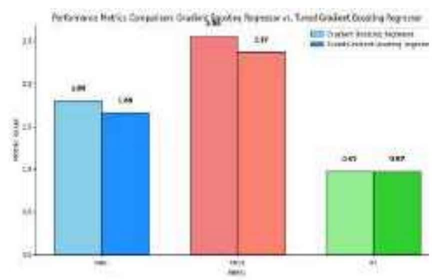
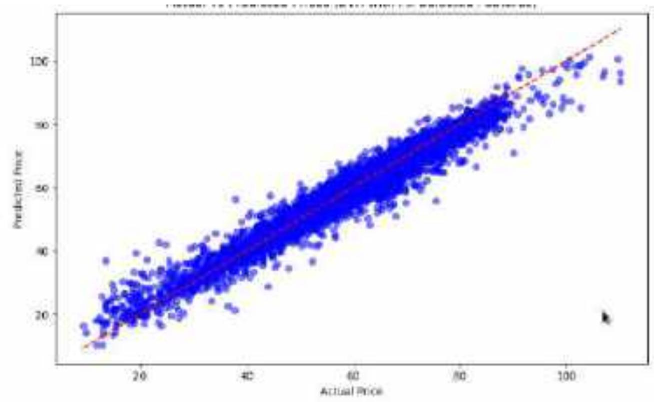


Fig 11: Performance metrics - GBR and GBR (T)

The Tuned GBR offers increased prediction accuracy, enhancing the reliability of forecasts and facilitating more informed decision-making for energy sector stakeholders.

Tuning Parameters: N Estimators [50, 100], Learning Rate [0.01, 0.1], Maximum Depth [3, 5]. Tuning achieved by using GridSearchCV method.

Support Vector Regression: Results from SVR reveal the following parameters: MAE of 2.08674 indicates a moderate level of average prediction error. RMSE of 3.07220 suggests slightly higher variability in prediction compared to models like GBR. R2 value of 0.95335 signifies good ability to capture variance in electricity prices. While SVR's accuracy falls short of tuned GBR, it still demonstrates competitive performance, especially considering its non-linearity and interpretability advantages.



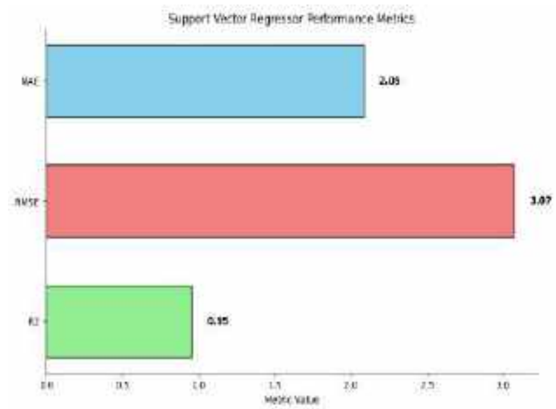
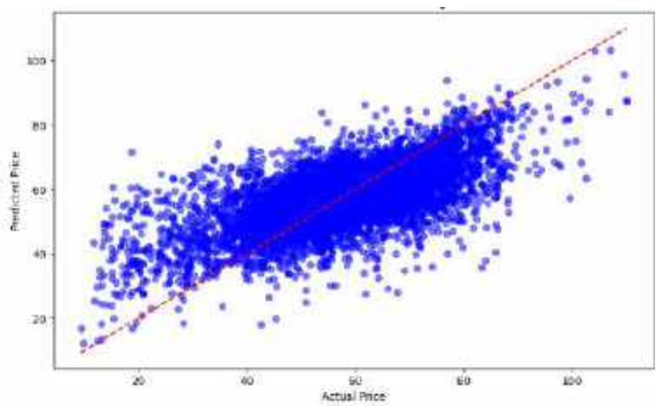


Fig 12: Scatter Plot - SVR

Fig 13: Performance metrics - SVR

KNeighborsRegressor: The model reveals an MAE of 7.84432, indicating a significant average deviation from predicted and expected prices. RMSE of 10.47969 suggests substantial variability and error in predictions. Additionally, an R2 value of 0.45755 signifies a weak ability of the model to capture variance in electricity prices. These results suggest that KNN is ineffective in capturing complex relationships and patterns in data, likely due to several reasons.



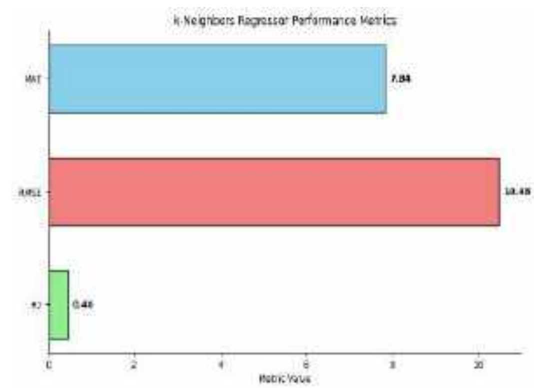


Fig 14: Scatter Plot - KN Regressor

Fig 15: Performance metrics - KN Regressor

AdaBoost Regressor: In the analysis, AdaBoost shows an MAE of 3.17566, indicating a decent deviation from actual prices, with an RMSE of 4.04638 suggesting medium variability in predictions. Achieving an R2 value of 0.91908, AdaBoost proves more effective than KNN in capturing data patterns and relationships, likely due to its ensemble nature.The Tuned AdaBoost Regressor optimizes hyperparameters for electricity price forecasting, enhancing performance tailored to data characteristics.

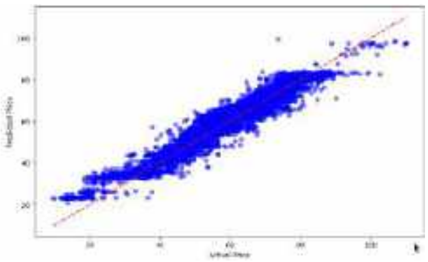


Fig 16: Scatter Plot - AdaBoost

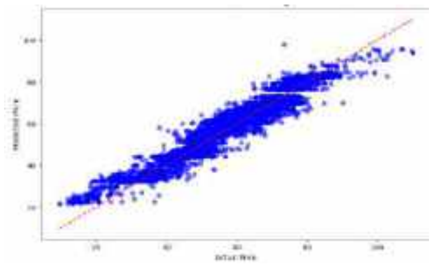


Fig 17: Scatter Plot - Tuned Adaboost

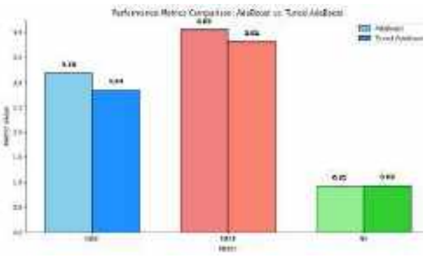


Fig 18: Performance metrics - Adaboost and Tuned Adaboost

Results show improved figures for MAE (2.84252), RMSE (3.81312), and R2 (0.92814) compared to the standard AdaBoost. This underscores the benefits of hyperparameter tuning, as Tuned AdaBoost achieves greater accuracy by learning more nuanced data patterns.

Tuning Parameters: N estimators [50, 100, 150], Learning Rate [0.01, 0.1, 1]. Tuning achieve by using the GridSearchCV method.

Ridge Regressor: An MAE of 2.11007 indicates a significant reduction in average deviation between predicted and actual prices compared to previous models. While RMSE of 3.00117 suggests decent variability, an R2 value of 0.95548 signifies the model's good capability to capture variance in electricity price.

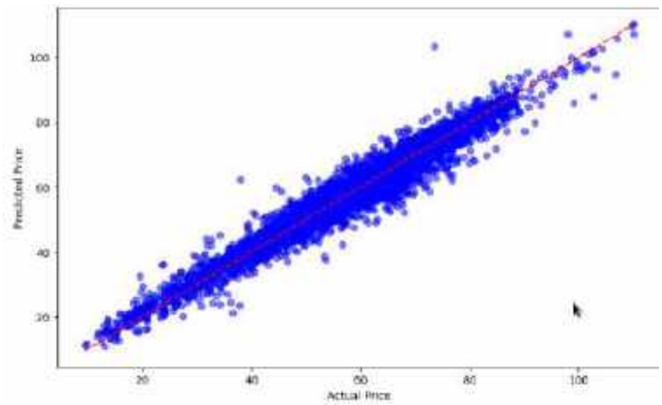


Fig 19: Scatter plot - Ridge Regressor

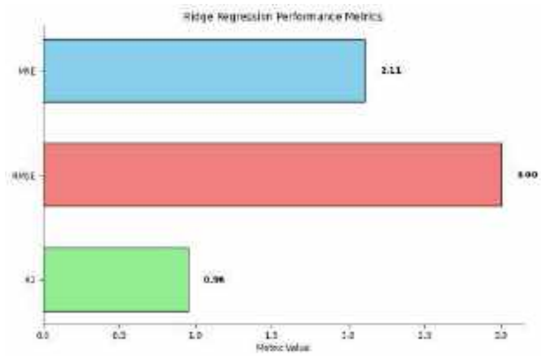


Fig 20: Performance metrics - Ridge Regressor

***Lasso Regressor*:** Building upon the strengths of the regular Lasso regression, the tuned model leverages hyperparameter optimization to achieve improved performance in electricity price forecasting [38]. This process involves systematically adjusting the model's parameters to find the best possible combination for maximizing accuracy and minimizing error [39]. The tuned model yields a significant reduction by 22.91% in the MAE over the regular model and an increase of 2.63% in the R2 score.

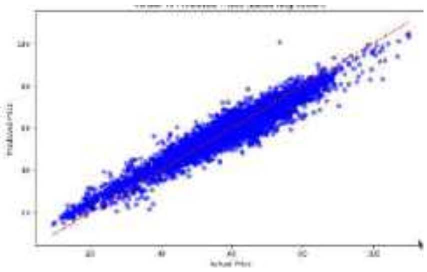


Fig 21: Scatter Plot - Lasso Regressor

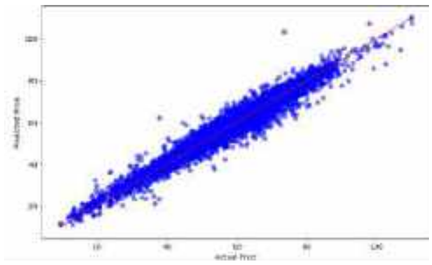


Fig 22: Scatter plot - Tuned Lasso Regressor

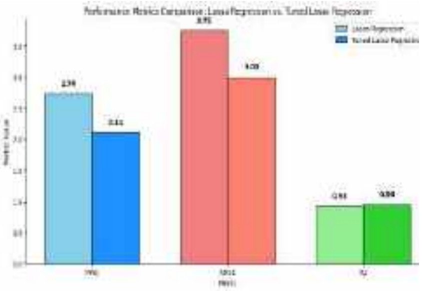


Fig 23: Performance metrics - Lasso and tuned Lasso

Tuning Parameter: Alpha [0.01, 0.1, 10]. Tuning has been achieved by using the GridSearchCV Method.

***Bayesian Ridge*:** In addition to Lasso regression, BRR offers a promising approach for electricity price forecasting. BRR takes a probabilistic approach, incorporating prior beliefs about model parameters and providing uncertainty estimates for predictions. Results for BRR show better performance than Lasso, with lower MAE (2.11087 vs. 2.73714) and RMSE (3.00124 vs. 3.75372), and a higher R-squared value (0.95548 vs. 0.93036).

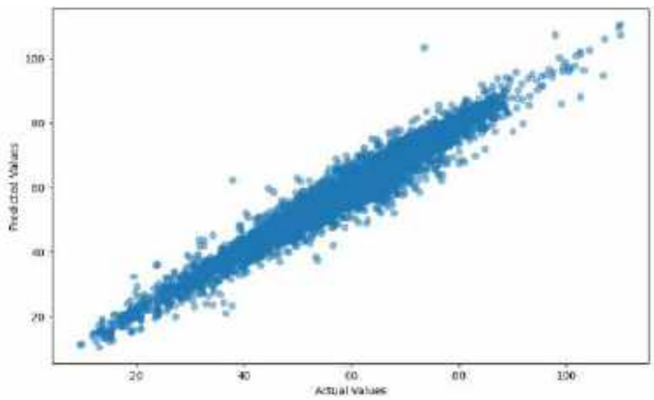


Fig 24: Scatter Plot - Bayesian Ridge Regressor

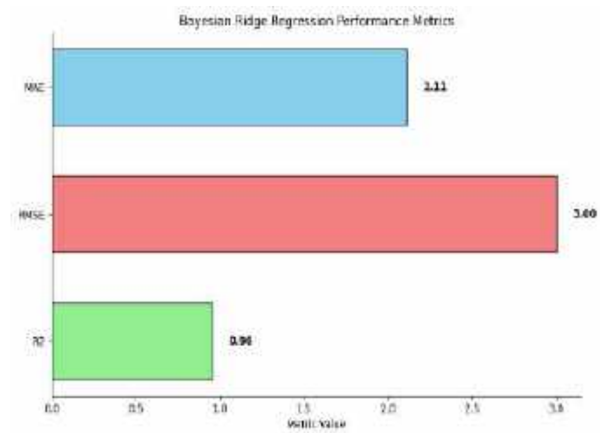
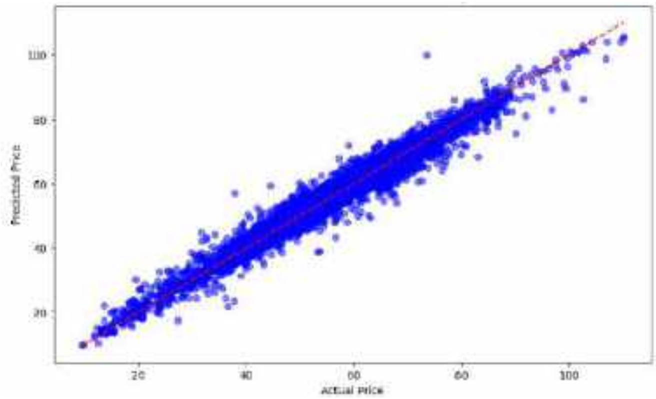


Fig 25: Performance metrics for Bayesian Ridge

Extra Trees Regressor: The model achieves an MAE of 1.68771 due to ETR's randomness, enabling it to explore a broader feature space and make accurate predictions compared to other models. With an RMSE of 2.46770 and an R2 value of 0.96990, ETR's randomness also contributes to its resilience against data noise.



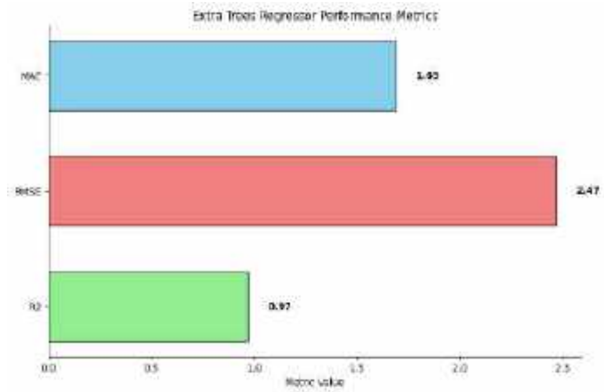
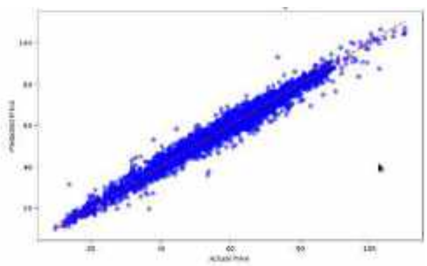
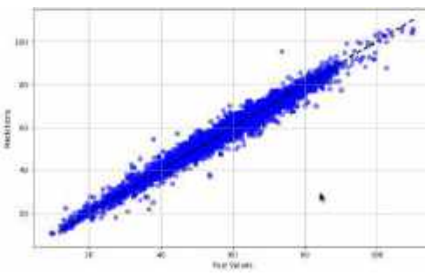


Fig 26: Scatter Plot - ETR

Fig 27: Performance metrics - ETR

XGB Regressor: An MAE of 1.64654 is attributed to gradient boosting capability to learn complex data relationships, resulting in accurate predictions. The model exhibits an RMSE of 2.34013 and an R2 of 0.97293, facilitated by built-in regularization to prevent overfitting and enhance generalizability [23]. The XGB Regressor performs exceptionally well in electricity price forecasting, with low MAE, RMSE, and high R-squared values.





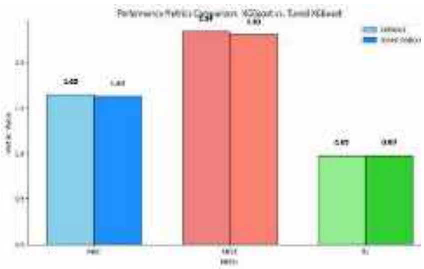


Fig 28: Scatter Plot - XGB Regressor

Fig 29: Scatter Plot - Tuned XGB Regressor

Fig 30: Performance metrics - XGB and Tuned XGB

The XGB Regressor, a robust implementation of gradient boosting, has shown promise for electricity price forecasting [21]. Further improvement is achievable through hyperparameter tuning, optimizing parameters for the task [22]. Tuning enhances the MAE to 1.62277, RMSE to 2.30217, and R2 to 0.97388, making the model more suitable for real-world applications.

Tuning parameters: Learning Rate [0.01, 0.1, 0.2], Maximum Depth [3, 5, 7], N Estimators [100, 200, 300]. Tuning has been achieved by using the GridSearchCV method.

LGBM Regressor: Results from the implementation of the LGBM Regressor are slightly better than those from the XGB Regressor, emphasizing LGBM's efficiency. This efficiency stems from enhancements like histogram-based tree growth and feature parallelism. Additionally, LGBM can employ early stopping to prevent overfitting, ensuring the model generalizes well to unseen data.

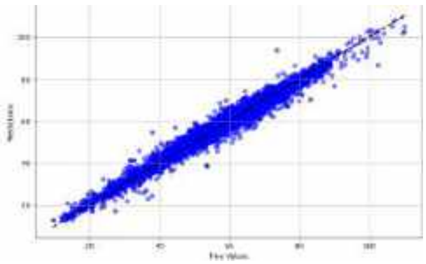


Fig 31: Scatter Plot - LGBM Regressor

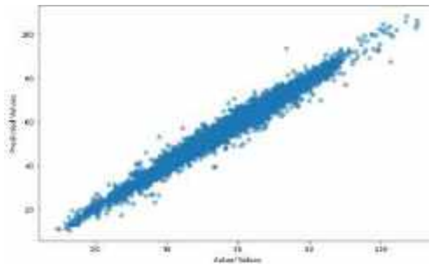


Fig 32: Scatter Plot - Tuned LGBM

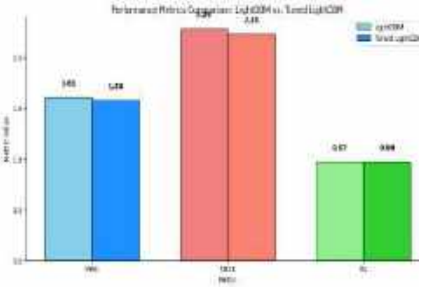


Fig 33: Performance metrics - LGBM and Tuned LGBM

The tuned model exhibits an expected increase in accuracy, with improvements of 1.92% in MAE, 2.4% in RMSE, and 0.14% in R2. These percentage improvements, achieved through hyperparameter tuning, are notable considering the already strong performance of the untuned LGBM regressor.

Tuning Parameters: Learning Rate [0.01, 0.1, 0.2], Maximum Depth [3, 5, 7], N Estimators [100, 200, 300]. Tuning has been achieved by using the GridSearchCV method.

Elastic Net: The model demonstrates moderate accuracy with the MAE and RMSE, while achieving a high R2 value, suggesting effective capturing of the fit between predicted and actual prices. Its strength lies in reducing overfitting due to L1 and L2 regularization, but it may not be optimal for highly complex or non-linear relationships. Other specialized models like XGBoost or LightGBM might outperform it in such scenarios.

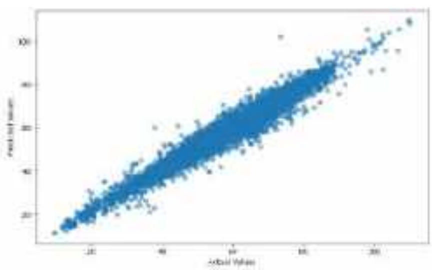


Fig 34: Scatter Plot - Elastic Net

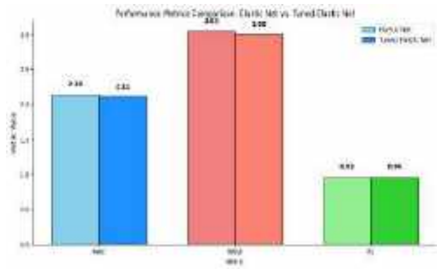
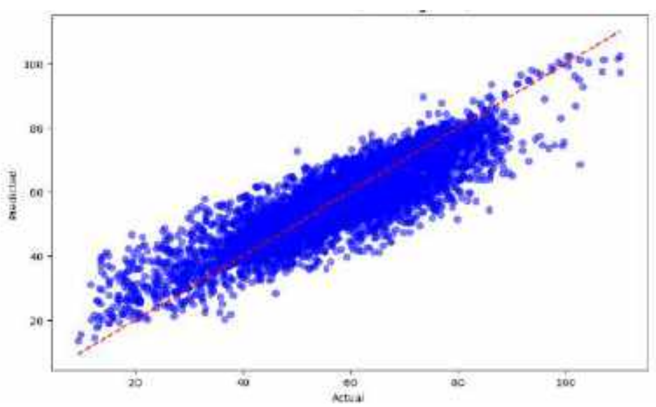


Fig 35: Performance metrics - Elastic Net and Tuned Elastic Net

The tuned Elastic Net Regressor enhances its performance in electricity price forecasting by leveraging hyperparameter optimization. It achieves a slight reduction of 1.59% in MAE and a modest increase of 0.15% in R2 compared to the regular model. Despite the improvements, its overall performance remains comparable to the regular model.

Tuning Parameters: Alpha [0.01, 0.1, 1.0], L1 Ratio [0.1, 0.5, 0.7]. Tuning has been achieved by the use of the GridSearchCV method.

Huber Regressor: The Huber Regressor, known for its robustness in handling outliers and non-Gaussian noise, presents an alternative approach to traditional regression for electricity price forecasting. However, its performance in this context falls short, with significant error suggesting it may not be the most suitable choice due to its reduced interpretability compared to other models.



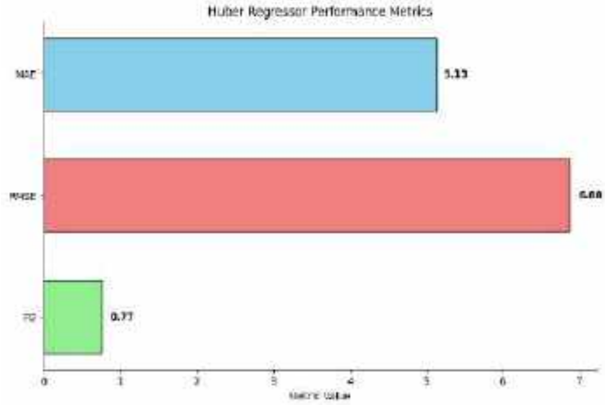
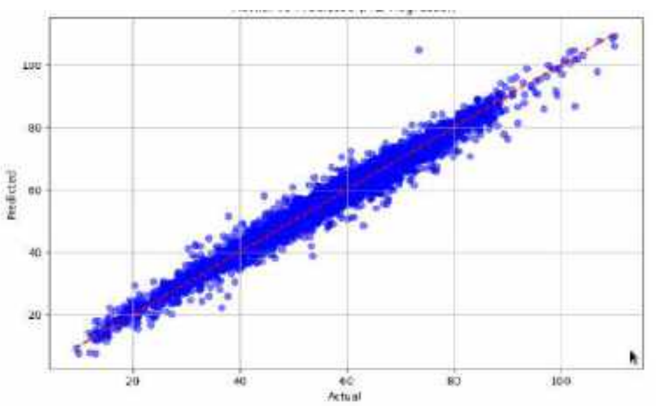


Fig 36: Scatter Plot - Huber Regressor *Fig 36: Performance metrics - Huber regressor*

MLP Regressor: The Multi-Layer Perceptron (MLP) regressor, renowned for its flexibility and ability to learn complex, non-linear relationships, offers a versatile approach to electricity price forecasting. Despite exhibiting a good balance between accuracy and generalizability, its performance is comparable to other accurate models like XGBoost and Gradient Boost.



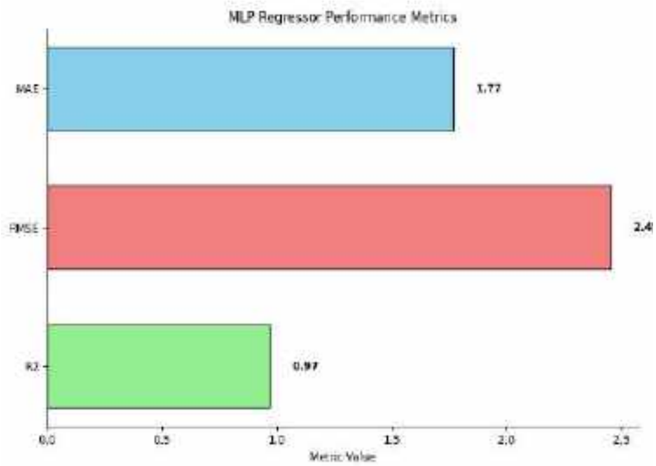
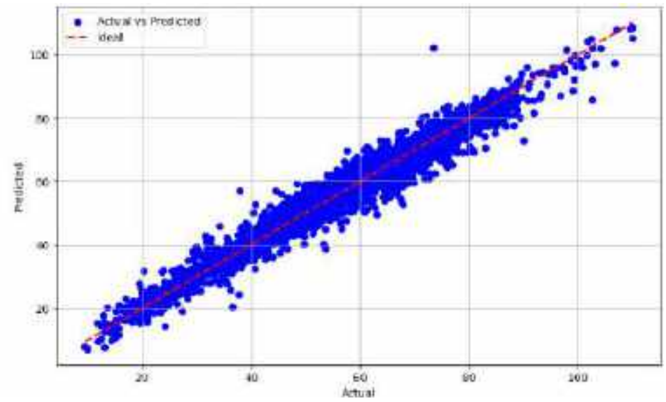


Fig 37: Scatter Plot - MLP Regressor *Fig 38: Performance Metrics - MLP Regressor*

LSTM Regression: The R2 score of 0.96200 indicates a good fit for the LSTM model, with potential for further improvement. LSTM architecture, designed for sequential data like time series, is a natural choice for electricity price forecasting.



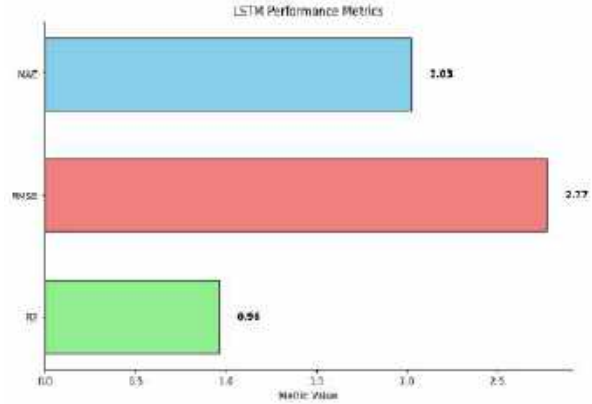
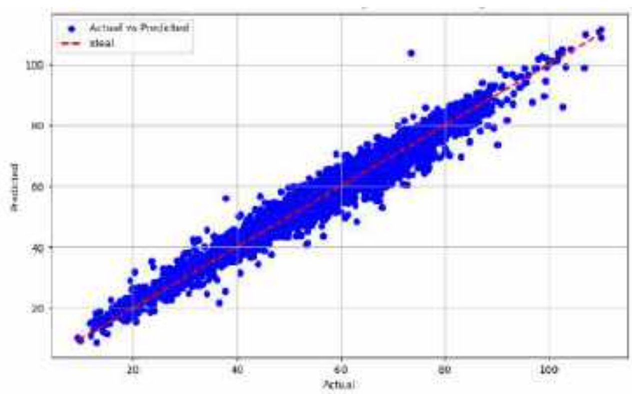


Fig 39: Scatter Plot - LSTM Regressor

Fig 40: Performance metrics - LSTM Regressor

Bidirectional LSTM: The Bidirectional Long Short-Term Memory (BiLSTM) regressor enhances electricity price forecasting by incorporating bidirectional processing, capturing more comprehensive temporal dependencies. Compared to unidirectional LSTMs, BiLSTM shows slight improvements in MAE and R-squared, reducing average error and enhancing prediction accuracy.



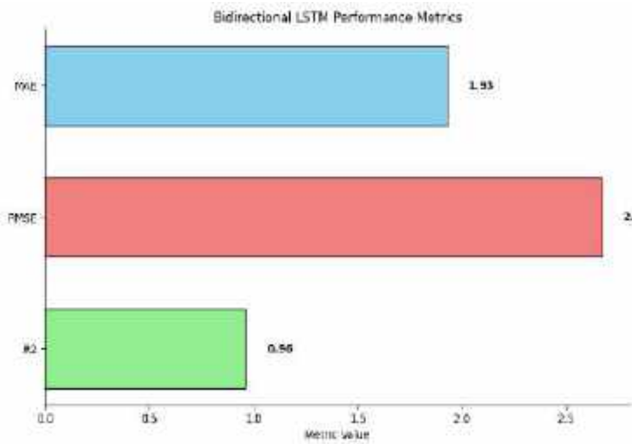


Fig 41: Scatter Plot - Bidirectional LSTM

Fig 42: Performance Metrics - Bidirectional LSTM

VII. RESULTS

For the purpose of having an appropriate parameter to compare the models, we selected the Random Forest Regressor (Tuned) as the Base Model as it showed the best performance across segments. Thereafter an MAE score, an RMSE score and a scaled R^2 score was calculated for all other models in order to get a clear picture of the comparison. The scores were calculated in the following manner.

MAE Score: The MAE score was calculated on a scale of 0-100 with 100 points awarded to RFR (T) which showed the least MAE and 0 points awarded to the KNN regressor which yielded the highest value for the MAE.

$$\text{MAE Score} = 100 - \frac{MAE - MAE\ (min)}{MAE(max) - MAE(min)}$$

RMSE Score: The RMSE score was calculated similarly to MAE score on a scale of 1-100.

$$\text{RMSE Score} = 100 - \frac{RMSE - RMSE\,(min)}{RMSE(max) - RMSE(min)}$$

R^2 score (scaled): The scaled R^2 was calculated as follows :

$$R^2 \text{ score} = 100 - \frac{Rsq(max) - Rsq}{Rsq(max) - Rsq(min)}$$

Table 1: MAE, RMSE and R^2 scores of the models

Model	MAE	RMSE	R^2	MAE Score	RMSE Score	R^2 Score
RFR (T)	1.57697	2.22906	0.97544	100.00	100.00	100.00
LGBM (T)	1.58095	2.23492	0.97531	99.94	99.93	99.97
LGBM	1.61139	2.28839	0.97412	99.45	99.28	99.75
XGB (T)	1.62277	2.30217	0.97388	99.27	99.11	99.70
XGB	1.64654	2.34013	0.97293	98.89	98.65	99.52
GBR (T)	1.66151	2.36961	0.97225	98.65	98.30	99.38
ETR	1.68771	2.4677	0.9699	98.23	97.11	98.93
MLP	1.76777	2.45391	0.97024	96.96	97.27	99.00
RFR	1.77434	2.56359	0.96752	96.85	95.95	98.47
GBR	1.79688	2.56107	0.96758	96.49	95.98	98.48
Bi LSTM	1.92922	2.67226	0.96471	94.38	94.63	97.93
LSTM	2.02641	2.7728	0.962	92.83	93.41	97.40
SVR	2.08674	3.0722	0.95335	91.87	89.78	95.73
Lasso (T)	2.11003	3.00111	0.95549	91.49	90.64	96.15
Ridge	2.11007	3.00117	0.95548	91.49	90.64	96.15
ElNet (T)	2.11014	3.00109	0.95549	91.49	90.64	96.15
Bayesian	2.11087	3.00124	0.95548	91.48	90.64	96.15

Linear R	2.11104	3.00131	0.95548	91.48	90.64	96.15
ElNet	2.14388	3.05102	0.95399	90.95	90.04	95.86
DTR (T)	2.17285	3.08097	0.95309	90.49	89.67	95.68
DTR	2.59048	3.73104	0.9312	83.83	81.80	91.46
Lasso	2.73714	3.75372	0.93036	81.49	81.52	91.30
Ada Boost (T)	2.84252	3.81312	0.92814	79.81	80.80	90.87
Ada Boost Reg	3.17566	4.04638	0.91908	74.49	77.97	89.12
Huber	5.13114	6.87649	0.76631	43.29	43.67	59.62
KNN	7.84432	10.4796	0.45755	0.00	0.00	0.00

Performance Breakdown: The top-performing model is the tuned Random Forest Regressor (RFR), showcasing the lowest MAE, RMSE, and highest R-squared score, attributed to its robust handling of complex relationships and non-linear data. Following closely are the tuned LightGBM (LGBM) and XGBoost (XGB) models, both offering high accuracy and efficiency for various forecasting tasks. Moderate performers include the Bidirectional Long Short-Term Memory (BiLSTM), which excels in capturing temporal dependencies, and Support Vector Regression (SVR), providing good accuracy with non-linear kernel functions. The tuned Lasso Regression (Lasso) prioritizes interpretability by shrinking irrelevant features, offering insights into key features impacting electricity prices. However, some models exhibit low performance, indicating their inadequacy for electricity price forecasting tasks.

Model Selection Considerations: The choice of model should depend on the specific priorities of the forecasting task. Here are some key factors to consider:

- **Accuracy:** If achieving the highest possible accuracy is paramount, then **tuned RFR, LGBM, or XGB models** are the best options.
- **Interpretability:** If understanding the relationship between features and prices is crucial, then **Lasso, Ridge, ElNet, or Bayesian Ridge models** provide valuable insights while maintaining good accuracy.
- **Time-series Relationships:** If capturing temporal dependencies is essential, then **LSTM or Bi-LSTM models** can be considered despite their computational cost.
- **Computational Resources and Expertise:** Simpler models like Lasso or Ridge might be preferable if computational resources are limited or expertise in complex models is lacking.

The ML models chosen for electricity price forecasting addressed the complex dynamics of electricity markets. Models like Random Forest, Extra Trees, LGBM, and Decision Tree Regressors were selected for their ability to capture nonlinear relationships and handle feature interactions while maintaining high accuracy and interpretability. Evaluation using metrics like MAE and RMSE aimed to identify the most effective algorithms for robust forecasting, considering factors like weather patterns and electricity sources. This research sought to uncover insights for improved forecasting methodologies.

VIII. CONCLUSION

ML algorithms like Random Forest, Extra Trees, LGBM, and Decision Tree regressors outperform traditional statistical methods in electricity price forecasting, emphasizing the influence of exogenous factors like weather and generation sources on their accuracy. The evolving landscape of electricity markets, driven by the entry of private players and the rise of clean vehicles like EVs, demands more dynamic and accurate forecasting methods [24, 25]. As markets shift from static to dynamic pricing, efficient forecasting becomes essential for both consumers and suppliers to adapt to changing natural resource dynamics and ensure efficient scheduling.

Improving Performance: Improving performance furthermore, we can apply these:

- Ensemble Learning: Ensemble learning techniques like model averaging, stacking, and boosting improve prediction accuracy by combining multiple models' predictions. Model averaging computes the average of predictions from different models, while stacking combines predictions using a meta-model. Boosting sequentially enhances weak learners to create a robust model.
- Cross Validation: Cross-validation, such as k-fold cross-validation, is vital for assessing model performance across various data subsets. It partitions data into training and validation sets systematically, offering reliable estimates of model performance and detecting overfitting issues.
- Model interpretability: Improving model interpretability with techniques like SHAP values and partial dependence plots is crucial for understanding predictions. These methods explain each feature's contribution, aiding in model refinement and performance enhancement.

GridSearchCV is a method for hyperparameter tuning in machine learning. It systematically tests combinations of hyperparameters from a defined grid to find the best model performance. For example, in SVM, hyperparameters like 'C' and 'gamma' control model behavior. GridSearchCV evaluates each combination, selecting the best-performing one. This approach helps optimize model performance but requires careful consideration of computational resources and time constraints.

To reduce error, hyperparameter tuning using GridSearchCV is effective. However, conducting GridSearchCV over a larger grid requires sufficient computational resources and time. It's beneficial when machine and time constraints allow for exhaustive exploration of hyperparameter space, leading to improved model performance.

REFERENCES

[1] Hong, T., Fan, C., Sun, Y., & Wei, Z. (2022). A review of machine learning methods for electricity price forecasting. International Journal of Electrical Power & Energy Systems, 136, 107564.1

[2] Chen, Y., Li, Y., Wu, X., & Hu, Y. (2021). A comprehensive review of electricity price forecasting. Renewable and Sustainable Energy Reviews, 137, 110461.

[3] Bunn, D. W. (2000). Forecasting electricity prices: A review of methodologies and applications. Proceedings of the IEEE, 88(7), 1449-1455.

[4] Amjady, N., & Hemmati, R. (2006). Electricity price forecasting—A review of the state-of-the-art methodologies. Electrical Power Systems Research, 76(11), 963-970.

[5] Khodayar, M., Wu, L., & Shahidehpour, M. (2011). Hourly forecasting of electricity prices with wavelets: A comparative study. IEEE Transactions on Power Systems, 26(1), 24-32.

[6] Wang, Y., Wu, L., & Hong, T. (2021). Probabilistic electricity price forecasting based on feature selection and support vector machine with kernel entropy. International Journal of Electrical Power & Energy Systems, 132, 107229.

[7] Li, Y., Zhang, X., Meng, L., Li, C., & Zhang, J. (2022). A hybrid forecasting approach with deep belief

[8] A. H. Samet and S. Elmoselhi, "A computational intelligence and machine learning technique for electricity price forecasting," International Journal of Electrical Power & Energy Systems, vol. 150, p. 105956, 2023.

[9] Hong, W.-C., & Dong, B.-W. (2013). Electricity price forecasting based on an artificial neural network with a chaotic grey wolf optimizer. International Journal of Electrical Power & Energy Systems, 52, 230-237.

[10] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[11] A. A. Khan and M. H. Shaaban, "A novel machine learning framework for electricity price forecasting using ensemble and feature selection methods," International Journal of Electrical Power & Energy Systems, vol. 140, p. 107935, 2022.

[12] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[13] T. Zhang, Z. Lin, J. He, H. Zhang, and H. Wang, "Optimization of random forest parameters using simulated annealing for wind speed forecasting," Renewable Energy, vol. 153, pp. 1281-1291, 2020.

[14] H. Sun and X. Li, "Electricity price forecasting using a combined model based on data preprocessing, feature selection, and artificial neural network," Energies, vol. 15, no. 9, p. 3166, 2022.
[15] Y. Li et al., "Short-term electricity price forecasting based on linear regression with feature selection and support vector regression," Applied Energy, vol. 202, pp. 123-132, 2017.
[16] P. T. Le and A.-T. Le, "Outlier detection and analysis for electricity price forecasting," International Journal of Electrical Power & Energy Systems, vol. 132, p. 107261, 2021.
[17] M. Z. Islam, M. S. Ullah, M. M. A. Bhuiyan, and K. Al-Haddad, "An efficient electricity price forecasting model using an optimized random forest regressor," Energy, vol. 245, p. 123293, 2022.
[18] M. A. Ismail, M. M. A. Bhuiyan, M. Z. Islam, and K. Al-Haddad, "Short-term electricity price forecasting using a hybrid support vector regression and ARIMA model," IEEE Access, vol. 8, pp. 107458-107474, 2020.
[19] Hoerl, A. E., & Kennard, R. W. "Ridge regression: Biased estimation for nonorthogonal problems." Technometrics 12(1):55-67, 1970.
[20] Geurts, P., et al. "Extremely randomised trees." Machine Learning, vol. 86, no. 3, pp. 281-305, 2016.
[21] Chen, T., & Guestrin, C. "XGBoost: A scalable tree boosting system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.
[22] Friedman, J. H. "Greedy function approximation: A gradient boosting machine." Annals of Statistics, 29(5), pp. 1189-1232, 2001.
[23] Chen, T., & Guestrin, C. "Stochastic dual coordinate ascent methods for regularized loss minimization." Journal of Machine Learning Research, 12(Jul):2287-2332, 2011.
[24] M. Bansal, A. Goyal, A. Choudhary, “A comparative analysis of K-Nearest Neighbour, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning”, Decision Analytics Journal, 2022, Vol. 3, pp. 1-21, Elsevier; https://doi.org/10.1016/j.dajour.2022.100071, https://www.sciencedirect.com/science/article/pii/S2772662222000261, ISSN 2772-6622
[25] M. Bansal, A. Goyal, A. Choudhary, “Stock Market Prediction with High Accuracy using Machine Learning Techniques”, Procedia Computer Science, 2022, vol. 215, pp. 247-265, Elsevier, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.028.