Energy Forecasting Using Various Machine Learning Algorithms and its Comparative Analysis

Premalatha K
Professor
Electrical and Electronics Engineering
Kumaraguru College of Technology
Coimbatore, India

Bhuvaneswari S
Undergraduate Student
Electrical and Electronics Engineering
Kumaraguru College of Technology
Coimbatore, India

Janaranjani R
Undergraduate Student
Electrical and Electronics Engineering
Kumaraguru College of Technology
Coimbatore, India

Karthik M
Undergraduate Student
Electrical and Electronics Engineering
Kumaraguru College of Technology
Coimbatore, India

ABSTRACT:
India's economy heavily relies on energy, and ensuring a stable and secure energy supply is crucial for its growth and development. Accurate forecasting helps in planning for future energy needs, ensuring security and reliability in the energy supply chain. India's power grid faces challenges related to intermittency and variability due to the increasing integration of renewable energy sources like solar and wind power. Forecasting helps grid operators anticipate fluctuations in supply and demand, thus improving grid stability and reliability. This research explores the use of machine learning (ML) algorithms in energy forecasting, focusing on historical energy consumption and production data, weather patterns, and economic indicators. The study evaluates various ML models, including Linear Regression, Gradient Boost Regression, Decision trees, Random forest Regression, and Support vector Regression, for their predictive accuracy. Data pre-processing and Data mining are used to refine and prepare the dataset for model training. The results reveal distinct performance traits among ML algorithms, providing valuable insights into the application of ML in energy forecasting and guiding future research directions in meeting energy needs. We delve into various types of energy sources, including but not limited to solar, wind, fossil fuels, and emerging technologies, and discuss how machine learning models can enhance accuracy and adaptability in predicting load demands.

Keywords − Energy forecasting, Machine learning algorithms, Renewable energy integration, Grid stability,

---

I.   INTRODUCTION

In today's rapidly evolving energy landscape, effective load forecasting is critical for ensuring the stability and efficiency of power systems. With the integration of diverse energy sources such as renewable, non-renewable, and distributed energy generation, the need for accurate load forecasting has become more pronounced. Traditional methods often struggle to adapt to the dynamic nature of these sources, making ML, an increasingly attractive solution for optimizing energy management. Load forecasting involves predicting future energy consumption patterns, taking into account various factors like weather conditions, economic indicators, and technological advancements. Machine learning algorithms excel in handling complex and nonlinear relationships within large datasets, making them well-suited for the multifaceted nature of load forecasting across different energy sources. This paper explores the application of machine learning techniques in load forecasting, addressing the unique challenges posed by the diverse energy landscape. The earlier study related to this compares ARIMA, SVM, and ANN for daily solar energy generation forecasting. SVM outperforms ARIMA, while ANN's underperformance is acknowledged [1]. Popular accuracy metrics include Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Weather and economic parameters heavily influence forecasts, with a preference for Very Short-Term to Long-Term forecasting periods [2]. The integration of Big Data and machine learning into the MERIDA system for energy efficiency management uses deep neural networks and Random Forest to create predictive models for energy consumption in public buildings [5]. The ensemble, which includes Support Vector Machines, Random Forests, Deep Learning, and ARIMA, outperforms existing models and individual experts focusing on PV power production and electricity demand [3]. The use of smart meter data for improved short-term energy forecasting, highlighting the benefits of customer segmentation and deep learning techniques, demonstrates the segmentation and feature extraction can improve single-day predictions by up to 12%, reducing blackouts and enabling smart grid control [4].

II.   PROPOSED IDEA

Previous studies have focused on energy forecasting individually for different energy types, resulting in various datasets and forecasting models. While effective to some extent, this approach lacks synergy and fails to capture potential correlations and interactions among different energy sources. Each energy type is forecasted using separate models, leading to suboptimal predictions and inefficiencies in resource allocation.

Our proposed approach involves combining datasets from various energy sources, such as solar, wind, and hydroelectric power, into a unified framework within the generation-side transformer. By integrating multi-energy datasets, we aimed to create a comprehensive forecasting model that considers the interdependencies and synergies among different energy sources. Our dataset also includes environmental factors such as wind, pressure, temperature, humidity etc..., This holistic approach enables more accurate predictions by leveraging the collective information from diverse energy inputs. We have aggregated data from different energy sources into a unified dataset, incorporating relevant features such as weather conditions, geographical factors, and historical energy production. The integrated approach facilitates enhanced forecasting accuracy by considering the combined effects of multiple energy sources on overall generation dynamics. The generation-side transformer leverages the synergies among energy inputs to make more informed and precise predictions, leading to improved resource allocation and energy management strategies.

III.   METHODOLOGY

We have used an open source Python Platform named "Google Colab" for our work. Google Colab is a free cloud-based platform that enables Python code writing and execution in a collaborative environment. It is popular among data scientists, machine learning practitioners, and researchers due to its ease of use and integration with popular libraries and frameworks. Key features include Jupyter Notebooks, free GPU and TPU resources, seamless integration with Google Drive, pre-installed libraries like TensorFlow, PyTorch, Scikit-learn, Matplotlib, and Pandas, real-time collaboration features, code snippets and examples, a flexible environment, and easy publishing. Colab is a crucial tool for data science and machine learning professionals, offering cloud-based computing resources and collaborative features.

(i)Data pre-processing:
Handle missing numbers, outliers, and inconsistent data to make the data cleaner. To prepare the data for model training, this process may also include feature engineering, feature scaling, and data normalization.

(ii)Exploratory Data analysis
To learn more about the distribution, connections, and trends of the data, analyse and visualize it. EDA facilitates comprehension of the data's properties and the identification of possible model features.

Outlier detection is performed for all the feature of our dataset. The Box-Plot of one of the features is shown below.
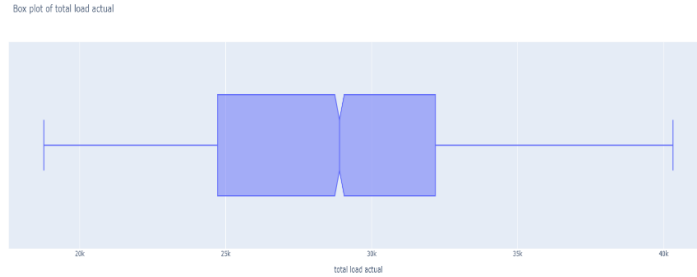
Figure 1: Box-Plot of Actual load to detect Outlier

(iii)Feature Selection

Feature selection is a critical step in machine learning that involves choosing the most relevant and informative features from the original dataset while removing irrelevant or redundant ones. Effective feature selection can improve model performance, reduce overfitting, and speed up training and inference. Here are some common methods for feature selection in machine learning

Our dataset contains nearly 40 different features which includes generation from various sources like, solar, hydro power, wind, pressure, temperature, humidity etc...,

(iv)Model Selection

Select the machine learning approach or algorithms that best fit the type of objective (classification, regression, clustering, etc.) and the features of the data. Neural networks, decision trees, random forests, support vector machines, logistic regression, k-nearest neighbours, and support vector machines are examples of common machine learning methods.

(v)Data training:

Divide the data into sets for training and validation. Using the training data, train the chosen model(s), modifying model parameters to maximize performance while preventing either underfitting (model failing to capture the underlying patterns) or overfitting (model memorizing the training data).

(vi)Model Evaluation:

Utilizing the validation set, evaluate the trained model(s) for performance and predictability. Depending on the kind of problem, common evaluation metrics include mean squared error, accuracy, precision, recall, R2-score and F1-score. Here, we've used R_2 Score to determine the accuracy of our model.

(vii) Hyperparameter tuning: Fine-tune the model's hyperparameters (parameters that control the learning process) using techniques such as grid search, random search, or Bayesian optimization to improve performance further.
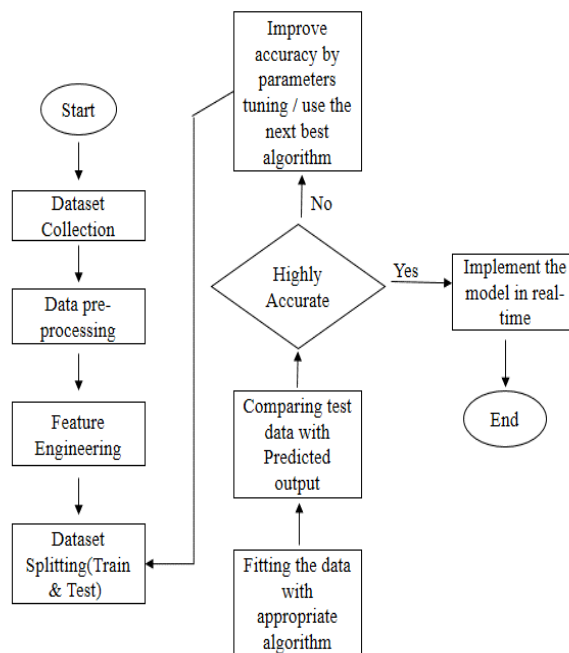
IV.WORKFLOW

Figure 2: Workflow of the proposed work

IV.   RESULTS AND DISCUSSIONS

The results of our comparative study, as depicted in the bar graph below, highlight the performance of each algorithm across the specified metrics. Notably, the Random Forest Regression algorithm consistently outperformed the other algorithms in terms of accuracy, achieving higher R-squared scores, lower MAPE, and lower MAE values.

The superior performance of the Random Forest Regression algorithm can be attributed to its ability to handle non-linear relationships and high-dimensional datasets effectively.
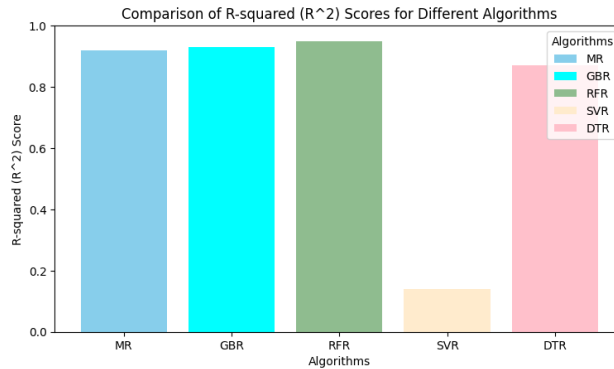

Figure 3: Graphical representation of comparing different algorithms

The above figure is the result of the proposed system that compared five different algorithms.

Figure 4: Testing the dataset

Figure 5: Predicted values using Random Forest Algorithm

From the above image, we can see that the predicted value is almost similar to the values in test data. Hence, the R_2 Score and different errors for Random Forest algorithm is

Figure 6: Accuracy determining parameters for our best algorithm.

Similarly, for different algorithms, these four parameters are analysed based on their predictions.

ALGORITHM	R^2 SCORE	RMSE	MAE	MAPE
Gradient Boost Regression	0.93	1208.03	918.65	0.03
Decision Tree	0.87	1616.78	1137.78	0.04
Random Forest	0.96	978.54	723.84	0.03
Support Vector Regression	0.14	4201.95	1040.15	0.13
Linear Regression	0.92	1281.66	997.51	0.04

Figure 7: Comparative analysis of selected 5 algorithms with four different parameters.

The performance of each algorithm was evaluated using four key metrics: R-squared score, Root Mean Square Error (RMSE). Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). The real-time dataset was fed into training and testing sets to ensure unbiased evaluation.

V. CONCLUSION

In summary, our research underscores the critical role of energy forecasting in strategic planning, resource allocation, and grid management. Through the utilization of machine learning algorithms and comprehensive data analysis, we aimed to enhance predictive accuracy in forecasting energy trends.

Our study involved a thorough examination of various machine learning models, including Linear Regression, Gradient Boost Regression, Decision Trees, Random Forest Regression, and Support Vector Regression. By evaluating their performance metrics such as R_2 Score, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared, we were able to compare their effectiveness in forecasting future energy consumption and production patterns.

Furthermore, our research involved meticulous data preprocessing techniques such as cleaning, normalization, and feature engineering to ensure the reliability and accuracy of the models. The division of data into training, validation, and test sets facilitated a rigorous assessment of model performance.

Our findings revealed distinct performance traits among the machine learning algorithms, shedding light on their respective strengths and weaknesses in energy forecasting. This comparative analysis provides valuable insights for stakeholders in the energy sector, enabling them to make informed decisions and allocate resources more efficiently.

Overall, this research contributes to the advancement of energy forecasting methodologies and offers practical guidance for future research and applications in this critical domain. We believe that our findings will not only enhance predictive accuracy but also foster innovation and improvement in energy-related predictive modeling techniques.

REFERENCES

[1]. Dominique Smith1, Kristen Jaskie1, JohnCadigan2, Joseph Marvin2, and Andreas Spanias1 "Machine learning for fast short-term energy load forecasting" in SenSIP Center, School of ECEE, Arizona State University, USA2 Prime Solutions Group (PSG), Goodyear, Arizona, USA.

[2]. Isaac Kof Ntil*, Moses Teimeh, Owusu Nyarko-Boateng and Adebayo Felix Adekoya "Electricity load forecasting: a systematic review" Department of Computer Science, Sunyani Technical, University, Box 206, Sunyani, Ghana.

[3]. Marijana Zekić-Sušac a, Saša Mitrović a, Adela Has b "Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities" International Journal of Information Management.

[4]. Gonçalo Luís R&D Neste João Esteves Nuno Pinho da Silva "Energy Forecasting Using an Ensamble of Machine Learning Methods Trained Only with Electricity Data" International Journal of Electrical Power & Energy Systems [4]

[5]. Sharif Atique, Subrina Noureen, Vishwajit Roy, Stephen Bayne, Joshua Macfie "Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques" 2020 IEEE Green Technologies Conference (GreenTech)

[6]. B. R. SAIDMURODOV, A. M. ZYUZEV AND E. A. LYUKHANOV, "MODELS OF OPERATIONAL FORECASTING OF ARC FURNACE ENERGY CONSUMPTION USING MACHINE LEARNING METHODS," 2023 RUSSIAN WORKSHOP ON POWER ENGINEERING AND AUTOMATION OF METALLURGY INDUSTRY: RESEARCH & PRACTICE (PEAMI), MAGNITOGORSK, RUSSIAN FEDERATION, 2023, PP. 205-210, DOI: 10.1109/PEAMI58441.2023.10299891.

[7]. B. R. Saidmurodov, A. M. Zyuzev and E. A. Lyukhanov, "Models of Operational Forecasting of Arc Furnace Energy Consumption Using Machine Learning Methods," 2023 Russian Workshop on Power Engineering and Automation of Metallurgy Industry: Research & Practice (PEAMI), Magnitogorsk, Russian Federation, 2023, pp. 205-210, doi: 10.1109/PEAMI58441.2023.10299891.

[8]. A. T. Eseye and M. Lehtonen, "Short-Term Forecasting of Heat Demand of Buildings for Efficient and Optimal Energy Management Based on Integrated Machine Learning Models," in IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7743-7755, Dec. 2020, doi: 10.1109/TII.2020.2970165.

[9]. R. Shah, H. Shah, S. Bhim, L. Heistrene and V. Pandya, "Short-term Electricity Price Forecasting using Ensemble Machine Learning Technique," 2021 1st International Conference in Information and Computing Research (iCORE), Manila, Philippines, 2021, pp. 145-150, doi: 10.1109/iCORE54267.2021.00045.

[10]. F. Jawaid and K. NazirJunejo, "Predicting daily mean solar power using machine learning regression techniques," 2016 Sixth International Conference on Innovative Computing Technology (INTECH), Dublin, Ireland, 2016, pp. 355-360, doi: 10.1109/INTECH.2016.7845051.

[11]. Y. Hu, L. Zhu, J. Li, Y. Zeng, L. Zheng and Y. Hou, "Short-term Wind Power Prediction Based on Error Estimation," 2023 8th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 2023, pp.1347-1351, doi: 10.1109/ICPRE59655.2023.10353699.

[12]. A. Dridi, H. Afifi, H. Moungla and J. Badosa, "A Novel Deep Reinforcement Approach for IIoT Micro grid Energy Management Systems," in IEEE Transactions on Green Communications and Networking, vol. 6, no. 1, pp. 148-159, March 2022, doi: 10.1109/TGCN.2021.3112043.

[13]. G. F. D. Apolinario, Y. -Y. Hong, Y. -D. Lee, J. -L. Jiang and S. -S. Wang, "Comparative Studies on Use of Discrete Wavelet Transform-based Feature Extraction for Peak Load Forecasting Using LSTM," 2021 IEEE 4th International Conference on Power and Energy Applications (ICPEA), Busan, Korea, Republic of, 2021, pp. 88-92, doi: 10.1109/ICPEA52760.2021.9639283.

[14]. M. Nutakki and S. Mandava, "A Comparative Assessment of Machine Learning and Deep Learning Techniques for Enhanced Home Energy Management Systems," 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2023, pp. 1-6, doi: 10.1109/i-PACT58649.2023.10434666.

[15]. A. P. Biswal, K. Roy and A. Ghosh, "AI-Enabled Prediction of Wind Speed and Temperature with Multiverse Optimization for Efficient Control of Standalone Hybrid System," 2023 IEEE 3rd International Conference on Sustainable Energy and Future Electric Transportation (SEFET), Bhubaneswar, India, 2023, pp. 1-6, doi: 10.1109/SeFeT57834.2023.10244917.