From Data to Decisions: Predicting Power Consumption in Indian State

Hardik Raina[1] and Vartika Puri[2[0000-0002-0751-9759]]

Department of Computer Science & Engineering and Information Technology Jaypee Institute of Information Technology, Noida, India

Abstract. Electricity is a foundation of modern society, serving as a vital energy source powering industries, homes, and essential services. Its importance goes beyond practicality; it is the cornerstone of society's progress and economic growth. In today's interconnected world, electricity fuels innovation and facilitates communication, transportation, healthcare, and various other sectors. Almost every facet of contemporary civilization depends on electricity, from owning industrial processes to facilitating digital communication. It is important to to accurately predict the electricity consumption so that the stability and efficiency of power systems and markets can be ensured. The objective of the paper is to predict the energy usage in different parts of India. Authors have used machine learning models XGBoost and Prophet for the same. Two metrices Mean Absolute Percentage Error and Root-Mean Square Error have been used to evaluate the performance of models. The results show that XGBoost outperforms Prophet.

Keywords: Power consumption, XGBoost, Prophet, forecasting, energy consumption.

1 Introduction

The enforcement of controlled use of electricity in India is very complex because of the existence of many interrelated problems, such as rapidly expanding population, urbanization, industry, and limited infrastructure. The demands for electricity are escalating which puts immense pressure on the existing power generation and distribution infrastructure. This results in frequent power outages, voltage fluctuations, and reliability issues. Furthermore, the country's heavy reliance on non-renewable energy sources such as coal aggravates environmental degradation and contributes to global climate change. India's shift towards a sustainable and resilient energy environment is hampered by inefficiencies in energy consumption, antiquated technologies, and insufficient investment in renewable energy infrastructure. Addressing these challenges requires a comprehensive approach that encompasses policy reform, investment in modern infrastructure, promotion of renewable energy adoption, enhancement of energy efficiency measures, and public awareness campaigns.

When we think about modern civilization, powering industries, homes, and essential services the electricity consumption will be a fundamental requirement.

As of October 2024, India emerging as a global economic powerhouse and the third-largest consumer and producer of electricity. It is utmost important to understand and manage the energy consumption patterns for sustainable development and economic growth.

The existing machine learning algorithms XGBoost [1] and Prophet [2] have been used in this paper to enhance the accuracy of prediction of energy consumption, uncover hidden patterns, and inform policy decisions, promote renewable energy adoption, and mitigate environmental impact. Machine learning techniques can help to uncover hidden patterns and correlations within the data, facilitating targeted interventions and tailored strategies for each region.

The rest of the paper is organized as follows: Section 2 explains the related work followed by the dataset used in Section 3. Section 4 demonstrates the implementation details, results, and analysis presented in Section 5 and Section 6 concludes the paper with some future research directions.

2 Related Work

This section discusses the previous studies to predict electrical energy consumption [4]. We elaborate on the various models and algorithms used for accurate

forecasting, highlighting their effectiveness in analyzing historical data and identifying patterns in previous research work [3]. The research combines Random forest and XGBoost model for the estimation of the consumption of electricity in Tripura, India, which can accurately predict the next 24 hours of load [7]. The model is trained on data from Tripura, India, which may not generalize well to other regions with different electricity consumption patterns, weather conditions, or infrastructure.

Another study compared logistic regression, Decision tree, Random forest, Gradient Boosting and Bagging classifier for predicting grid stability. The study aims to prevent grid failure by predicting load demand. Bagging classifier performed better as compared to other ML models [8]. The study used artificially generated dataset, it may not fully capture real-world grid fluctuations and uncertainties [8]. The research applies machine learning models (Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), Linear & Quadratic Discriminant Analysis, Naïve Bayes, K-Nearest Neighbors) for the prediction of the stability of the grid under the dynamically changing consumer requirement, they identified that SVM performed better than other models [9].

Gorzaczany et al. [10] proposed a fuzzy inference-based prediction system for predicting the smart grid control stability in decentralized smart grids. A grid stability assessment method based on the Bayesian optimized Light GBM optimization approach is presented by Wang et al. [11]. The simulated dataset has been used to check the accuracy of models, no validation has been done on real-world power grids with actual disturbances.

A load forecasting model based on real-time data collected from smart meters on the NIT Patna campus in India was proposed by Rai and De [12]. They discovered that the support vector regression model performs best in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The dataset is limited to certain geographical area and cannot be generalized for different geographical areas. The effects of upgrading operations and local wind farm power fluctuation on grid voltage stability, as well as the upgrading implications of grid voltage stability with different reactive power compensation strategies, were assessed by researchers using a simulated model [13].

Gupta et al. explored the application of XGBoost and Long Short-Term Memory (LSTM) models for time series forecasting of energy consumption. The study compares the performance of XGBoost and LSTM and demonstrates that XGBoost outperforms LSTM in energy forecasting [14]. The study focuses on energy consumption forecasting in a specific region (PJM East, 2002-2018). The model's effectiveness may vary in different geographic locations or industries.

This paper considers the dataset which belongs to all over India and have fluctuations in energy demands. Authors have compared XGBoost whose performance is better than other models for real world dataset with Prophet model.

The next section describes the dataset used in the paper.

3 Dataset

The Power System Operation Corporation Limited (POSOCO) provides invaluable insights into India's energy landscape through its comprehensive dataset, spanning 17 months from January 2019 to May 2020. This dataset, extracted from POSOCO's weekly energy reports [4], offers a detailed time series analysis of energy consumption at the state level, shedding light on regional variations, trends, and dynamics. For example, Figure 1 shows the trends of energy consumption monthly and weekly. The dataset provides a granular view of energy consumption at the state level, enabling policymakers, energy analysts, and stakeholders to gain valuable insights into regional consumption patterns, trends, and dynamics. For each state, data is split into training (80%) and testing sets (20%).

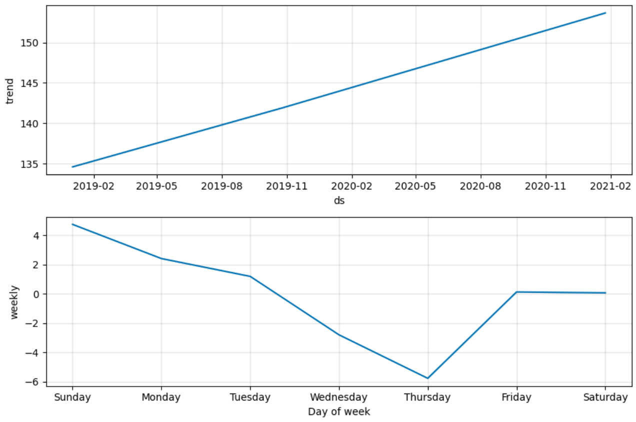

Fig. 1. Trends (month-wise and week-wise) in Punjab.

4 Implementation Details

We have implemented two models to understand the power utilization by the country Prophet and XGBoost. The models were implemented using python libraries like pandas, seaborn, matplotlib, XGBoost. The first step is to preprocess the data before providing to the XGBoost model. The individual model instances have been created for individual states to handle the variability in the scale of Power Consumed. The stages of predictions are reading, preprocessing, modeling, evaluating, and visualizing the energy usage data for different states using both the models.

5 Results and Analysis

The result section describes the performance of XGBoost and Prophet models and provides a comparative analysis to predict power usage across multiple states of India. We have used Mean Absolute Percentage Error (MAPE) [5] and Root Mean Square Error (RMSE) [6] parameters for each state to assess the model performance shown in Figure 2, Figure 3, Figure 4, and Figure 5.

MAPE is defined as the sum of the differences between actual value and predicted value which is represented as absolute errors shown as $|e_t|$ divided by the actual value shown as d_t as shown in Equation 1.

$$MAPE = \frac{1}{n}\sum \frac{|e_t|}{d_t} \quad (1)$$

RMSE is defined as the square root of the average squared error |e_t | as shown in Equation 2.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum e_t^2}$$

(2)

After applying both the models on individual states, XGBoost showed consistently better performance than Prophet even after fine-tune Prophet's hyperparameters for improved alignment with observed values. XGBoost can identify complex relationships in power consumption that may not be clear from the numerical time-series data alone.

In case of finding the seasonal and holiday influences on power usage Prophet is more accurate, but still it is not performing better than XGBoost for the regions like Dadra and Nagar Haveli (DNH), Punjab, and New Delhi.

The power consumption of DNH and Tripura is graphically compared in Figure 6, highlighting the significant fluctuations observed in DNH on specific dates: 2019-01-30, 2019-03-15, 2019-07-08, 2019-08-08, 2019-11-01, 2020-03-22, and 2020-05-03. Although both states experience changes in power consumption over time, the data for DNH demonstrates remarkably drastic variations in contrast to the relatively stable power consumption in Tripura.

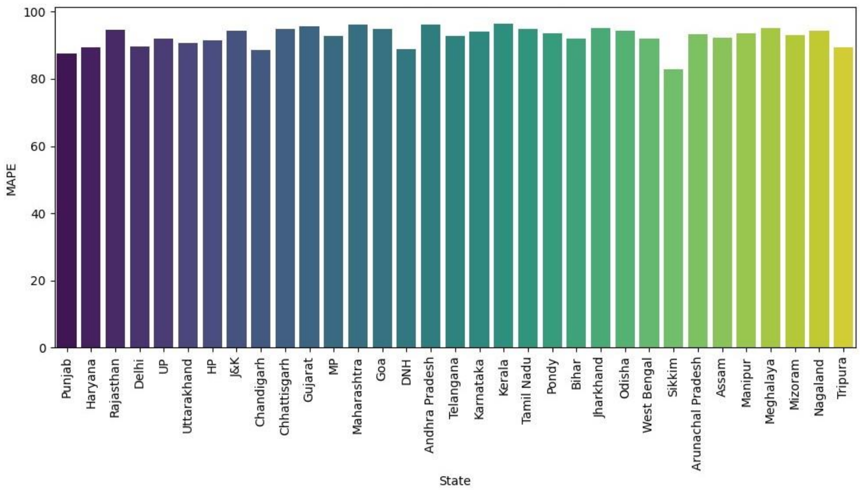

Fig. 2. MAPE on using XGBoost.

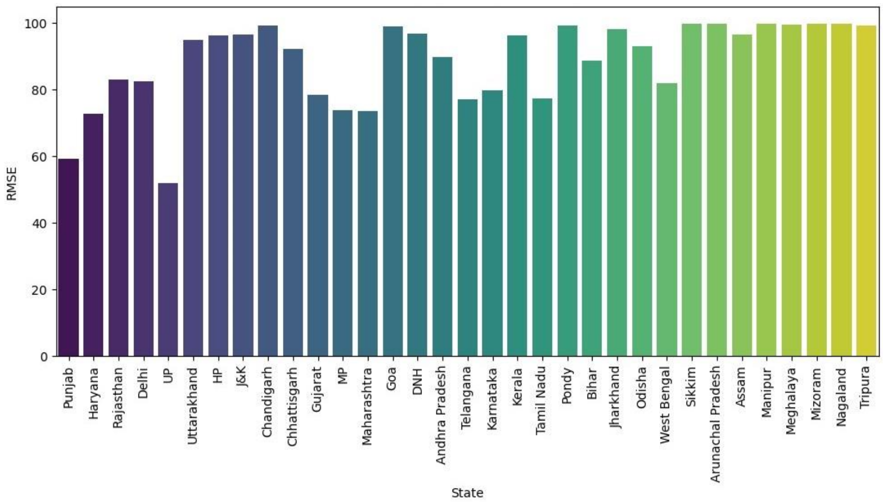

Fig. 3. RMSE on using XGBoost.

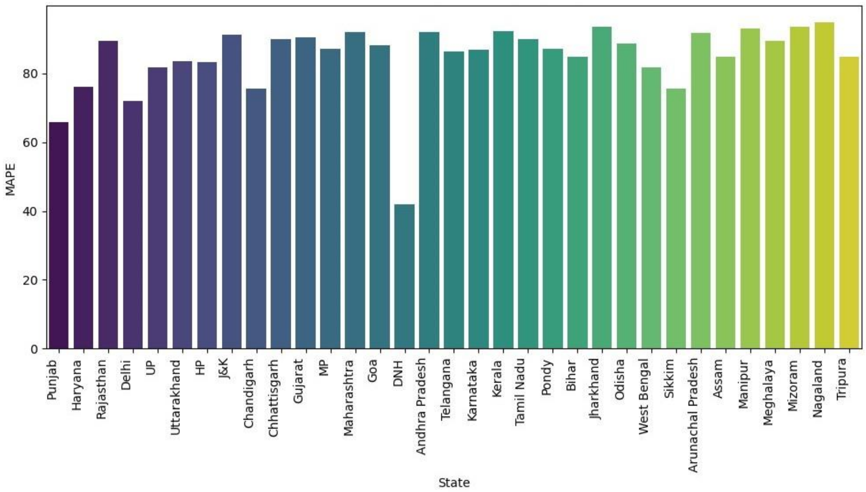

Fig. 4. MAPE on using Prophet.

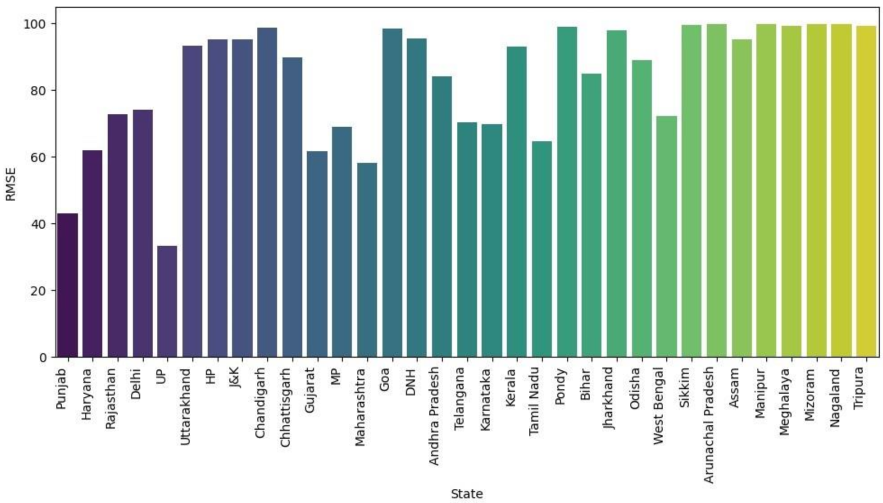

Fig. 5. RMSE on using Prophet.

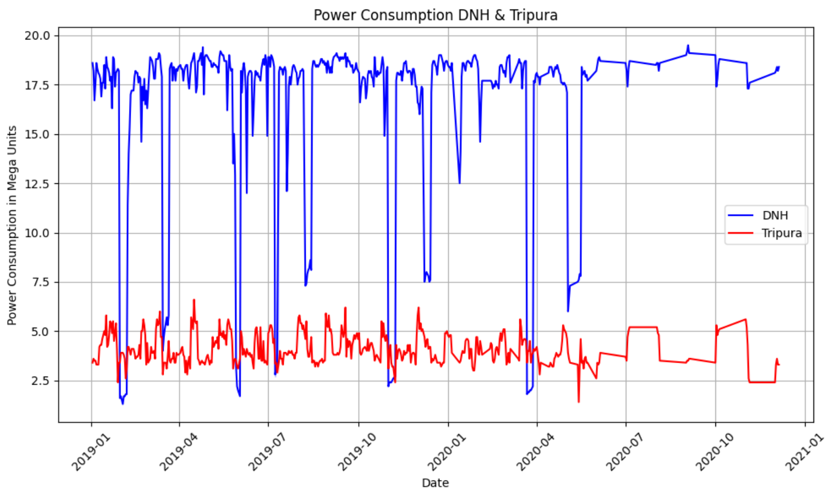

Fig. 6. Power Consumption Variability and Comparison.

Table 1. Performance of XGBoost and Prophet.

Model	MAPE (%)	RMSE (%)
XGBoost	7.38	11.32
Prophet	15.18	16.40

Table 1 shows the performance of both the algorithms. The following reasons we identified for better performance of XGBoost:

5.1 XGBoost's Superior Robustness

In contrast to Prophet, XGBoost demonstrated superior robustness in handling outlier points within the DNH power consumption data. XGBoost's flexibility and adaptability as observed from the gradient boosting algorithm, enabled it to effectively accommodate these abrupt deviations, thus enhancing its predictive accuracy compared to Prophet.

5.2 Prophet's Expectation of Smoother Data

Prophet, designed primarily for time series forecasting, operates under the assumption of relatively smoother data patterns and has been observed to struggle to appropriately capture abrupt fluctuations such as those observed in the DNH and Punjab power consumption data. As a result, Prophet's predictions tend to smooth out the impact of outlier points, leading to discrepancies between predicted and actual consumption levels, particularly during periods of significant deviation. The same has been observed to be made significantly more deviant on the limited scope of observations.

5.3 Implications for Predictive Modeling

The differential performance of XGBoost and Prophet in handling outlier points underscores the importance of selecting appropriate modeling techniques that align with the characteristics of the data under analysis. In scenarios where data exhibits abrupt fluctuations or outlier points, as observed in the case of DNH power consumption, the robustness and adaptability of algorithms like XGBoost prove advantageous in ensuring accurate predictions.

6 Conclusion & Future Work

The paper has demonstrated the use of XGBoost and Prophet models for the prediction of power consumption in different states of India. Our analysis concludes that XGBoost surpasses Prophet in predicting power consumption. Additionally, XGBoost able to handle outlier points, which we can see in DNH which is having sudden drops in power consumption. This shows that the model should be selected considering the characteristics of the data. Combining analysis of data belonging to different states highlights the power consumption patterns region wise. More research can detect the spatial links and correlations among nearby areas along with the influence of geographical elements on consumption trends. Combining geographic information systems (GIS) and spatial clustering methods can help in finding consistent consumption patterns in specific locations as well as improve the precision of forecasting models.

7 REFERENCES

1. Chen T. and Guestrin C.: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. [Online]. Available: http://dx.doi.org/10.1145/2939672.2939785 (2016)
2. Taylor S.J. and Letham, B.: Forecasting at scale. The American Statistician, 72(1), pp.37-45 (2018)
3. Sardar I., Akbar M.A., Leiva V., Alsanad A. and Mishra P.: Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: Methodology, evaluation, and case study in SAARC countries. Stochastic Environmental Research and Risk Assessment, 37(1), pp.345-359 (2023)
4. Power system operation corporation limited-NLDC- weekly Reports, [online] Available: https://nrldc.in/reports/weekly-reports/weekly-reports-2019-20/?lang=en.
5. De Myttenaere A., Golden B., Le Grand B. and Rossi F.: Mean absolute percentage error for regression models. Neurocomputing, 192, pp.38-48 (2016)
6. Hodson T.O.: Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions, pp.1-10 (2022)
7. Banik R., et al.: Prediction of electrical energy consumption based on machine learning technique. Electrical engineering 103.2, pp. 909-920 (2021)
8. Tiwari S., Jain A., Yadav K., and Ramadan R.: Machine learning-based model for prediction of power consumption in smart grid. Int. Arab J. Inf. Technol., 19(3), 323-329 (2022)
9. Tiwari S., Jain A., Ahmed N. M. O. S., Charu Alkwai L. M., Dafhalla A. K. Y. and Hamad S. A. S.: Machine learning-based model for prediction of power consumption in smart grid-smart way towards smart city. Expert Systems, 39(5), e12832. (2022)
10. Gorzałczany M., Piekoszewski J., and Rudziński F.: A Modern Data-Mining Approach Based on Genetically Optimized Fuzzy Systems for Interpretable and Accurate Smart-Grid Stability Prediction, Energies, vol. 13, no. 10, pp. 25-59 (2020).
11. Wang R., Liu Y., Ye X., Tang Q., Gou J., Huang, M., and Wen Y.: Power System Transient Stability Assessment Based on Bayesian Optimized Lightgbm, In Proceedings of IEEE 3 rd Conference on Energy Internet and Energy System Integration (EI2), Changsha, pp. 263-268 (2019)
12. Rai S. and De M.: Analysis of Classical and Machine Learning Based Short-Term and Mid-Term Load Forecasting for Smart Grid, International Journal of Sustainable Energy, vol. 40, no. 9, pp. 821-839 (2021)
13. Qinyu B., Lin Y., Jiayi M., Tianqi L., Xiaotian Z., and Youyin W.: Analysis of Influence with Connected wind Farm Power Changing and Improvement Strategies on Grid Voltage Stability, In Proceedings of China International Conference on Electricity Distribution, Tianjin, pp. 2029-2033 (2018)
14. Gupta R., Yadav A. K., Jha S. K., and Pathak P. K.: Time Series Forecasting For Energy Consumption Using XGBoost and LSTM. In proceeedings of 2024 Asia Pacific Conference on Innovation in Technology (APCIT) Mysore, India. Jul 26-27 (2024)