Received Date:10th December, 2024
Revision Date: 28th December, 2024
Accepted Date:10th January, 2025

Electricity Consumption Analysis and Prediction

Aayushma Adhikari[1*], Kumud Shrestha[2], Nistha Bhandari[3], Pawesha Shrestha[4]
[1]*Department of Computer Engineering, Kathmandu Engineering College. E-Mail: aayushma.adk59@gmail.com*
[2]*Department of Computer Engineering, Kathmandu Engineering College. E-Mail: shresthakumud5@gmail.com*
[3]*Department of Computer Engineering, Kathmandu Engineering College. E-Mail: bhandarinistha7@gmail.com*
[4]*Department of Computer Engineering, Kathmandu Engineering College. E-Mail: pawesha24@gmail.com*

Abstract — In our rapidly evolving world, the demand for electricity has surged due to its vital role in driving national development across various sectors. Accurate forecasting of electricity demand is crucial for effective energy resource planning and management. The Electricity Consumption Analysis and Prediction System utilized a comprehensive dataset and employed KNN, linear regression, and XGBoost algorithms to predict electricity consumption. The data for our project has been collected from the Grid Office in Baneshwor. After meticulous data cleaning, feature engineering, and integration, XGBoost emerged as the optimal model, showcasing superior accuracy with Mean Squared Error (MSE) of 0.03, Root Mean Squared Error (RMSE) of 0.18, and an impressive R2-score of 0.95. The system effectively visualizes analyzed data and model predictions using charts and graphs, providing users with intuitive insights. The system generates predictions for electricity consumption on a monthly, daily, and yearly basis for the Baneshwor Area.

Keywords: Electricity demand, Machine Learning, Linear Regression, KNN, XGBoost

Introduction

Background Theory

Electricity is important in our modern world. It is the main source of energy we use. We rely on electricity or lots of things we use every day, like turning on lights, using appliances, and making our computers work. Whether it is for watching TV, charging our phones, or keeping our houses warm or cold, we depend on electricity to make our lives simpler and more comfortable. These industries rely heavily on electricity to meet their needs. That is why forecasting electricity demand is crucial. It helps in planning and managing energy resources more effectively. By accurately predicting how much electricity will be needed, we can ensure a stable and reliable supply of energy for industries, supporting their operations and contributing to overall economic growth.

** Corresponding Author*

Accurate predictions of electricity demand are crucial for making smart decisions in the power industry. Whether it's managing day-to-day operations or planning for the future, these forecasts guide actions like balancing energy production and usage, scheduling loads efficiently, and deciding which power plants to use. Reliable forecasts are the key to minimizing risks and ensuring that power systems work smoothly and effectively, both in the short term and for long-term investments.

Based on the lead time involved in forecasting, it can be categorized into short-term, medium-term, and long-term forecasting. The choice of prediction timeframe influences model selection, methodologies, and the inclusion of external parameters. For instance, long-term forecasting models may incorporate socio-economic and population growth factors, while excluding them in favor of atmospheric, seasonal, and short-term dependencies in medium-term models. Our project, concentrates on developing a medium-term electricity demand model to forecast electricity demand in Baneshwor area.

In the case of Nepal, where electricity demand forecasting has been relatively underexplored, our project aims to fill this research gap by constructing a robust middle-term electricity demand forecasting model. By analyzing historical data and incorporating relevant factors such as temperature, hour of the day, day of the week, etc the model aims to provide accurate predictions tailored to the specific context of Baneshwor.

"Electricity Consumption Analysis and Prediction" plays a crucial role in analyzing and comprehending electricity consumption patterns. This system provides insights into the patterns and trends of electricity usage, enabling decision makers to make informed decisions regarding resource allocation and energy management. Using the capabilities of this system, it is possible to plan ahead for future energy requirements, obtain a better understanding of the dynamics of power consumption, and create efforts that encourage the effective and sustainable use of electricity.

In the realm of machine learning, various algorithms are employed for forecasting tasks, each with its strengths and weaknesses. K-Nearest Neighbors (KNN) is a simple yet effective algorithm used for regression tasks, relying on the majority class or the average of the k-nearest data points in the feature space for predictions. Linear Regression, on the other hand, fits a linear model to the data, assuming a linear relationship between the independent and dependent variables, making it simple and interpretable but limited to modeling linear relationships. XGBoost (Extreme Gradient Boosting), a powerful ensemble learning algorithm, sequentially combines multiple weak learners to create a strong learner, offering advantages such as handling missing values, preventing over- fitting, and capturing nonlinear relationships and interactions between features, making it widely used in various machine learning competitions and real-world applications, including electricity demand forecasting.

In this project, XGBoost, a machine learning technique, has played a central role in forecasting electricity demand. The model's training utilized two years' worth of historical input data, enabling it to establish and utilize relevant relationships for generating predictions. Subsequently, these predictions underwent rigorous validation against actual data to ensure their accuracy. Additionally, acknowledging the significance of high-quality datasets for system efficiency, direct data collection visits were conducted to the Nepal Electricity Authority (NEA). This ensured that our system obtained reliable and up-to-date information crucial for accurate electrical demand forecasting.

Problem Statement

In Nepal, current demand prediction practices primarily focus on long-term planning, with limited attention given to short-term and medium-term load forecasting. The absence of accurate medium-term forecasts contributes to imbalances between power generation and demand, resulting in issues such as load switching problems and energy leakage. This unpredictability also leads to instances where electricity production either surpasses or falls short of the actual demand, highlighting the critical need for improved forecasting methods to ensure a more balanced and efficient energy distribution system.

Objectives

To analyze and deliver an accurate prediction for electricity consumption on a daily, monthly and yearly basis for the Baneshwor area.

Literature Review

Researchers have extensively explored the dynamics of the retail electricity market, investigating topics like business models, pricing strategies, risk management, and electricity price forecasting. They employ various methods such as regression analysis, time series analysis, and soft computing techniques like fuzzy logic and neural networks. Recent studies have integrated artificial intelligence and econometric approaches for more precise analysis.

In Chile, Palacios and Saavedra advocate for partial liberalization of the residential electricity market, emphasizing welfare gains [1]. Streimikiene and Siksnelyte's work assesses the sustainability of electricity market models, considering economic, social, and environmental factors [2].

The volatility of wholesale electricity markets poses significant risks to retailers, leading to the need for detailed risk analysis and mitigation strategies. Bartelj et al. propose a model for evaluating sales contract offer maturity risk in the electricity retail business [3]. The Markov chain model finds widespread application in energy markets, offering insights into system status transitions and market volatility prediction [4]. Moreover, electricity consumption forecasting plays a crucial role in local and national planning and trading on electricity markets [5].

In addition, Mocanu et al. categorize electricity demand forecasting into short-term, medium-term, and long-term horizons, highlighting its strategic importance for local and national planning and trading on electricity markets [6].

Despite extensive studies on short-term load forecasting, Nepal has not utilized these methods. Existing forecasts, such as the MAED-based approach, focus on long-term projections driven by socio-economic factors, technology, and demography, indicating a significant gap in short-term forecasting practices [7].

Furthermore, analysis and forecasting of electricity consumption play a pivotal role at various levels, regardless of forecasting granularity. Such forecasts are crucial for planning and trading on electricity markets [8] Regarding forecasting efforts in Nepal, the Water and Energy Commission Secretariat, Government of Nepal, has conducted studies such as the Electricity Demand Forecast and the Energy Demand Projection 2030: A MAED Based Approach [9]. These studies utilize scenario-based planning tools like the MAED model to project long-term energy demand based on socio-economic, technological, and demographic factors, indicating a reliance on long-term projections for planning purposes [10].

Moreover, the Markov chain model has been extensively applied in various energy market contexts, including biogas production, wind power forecasting, crude oil import analysis, and energy supply and demand modeling. Its ability to provide quantitative analysis of system status transitions and market volatility has made it a valuable tool in energy sector decision-making [11].

In conclusion, the research landscape surrounding the retail electricity market is diverse and multifaceted, encompassing a wide range of topics and methodologies. From policy recommendations to risk analysis and forecasting techniques, scholars are continually striving to improve our understanding of this critical sector. These insights are essential for informing decision-makers and stakeholders, ultimately contributing to the efficiency, sustainability, and resilience of electricity markets worldwide.

Methodology

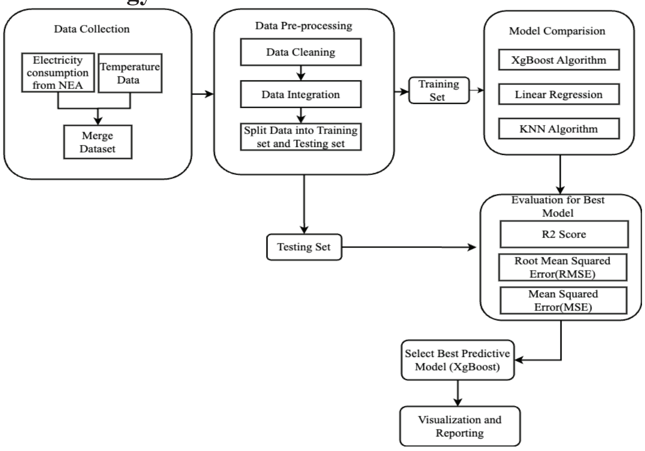


Fig. 1 System block diagram

The system block diagram illustrates the basic operation of the model utilizing machine learning algorithms. The prediction of electricity demand is influenced by various factors, which are determined through the analysis of historical data. Some of these factors that affect electricity demand include the hour of the day, day of the week, temperature, distinction between weekends and weekdays, and holidays. These factors serve as inputs for the trained model during prediction.

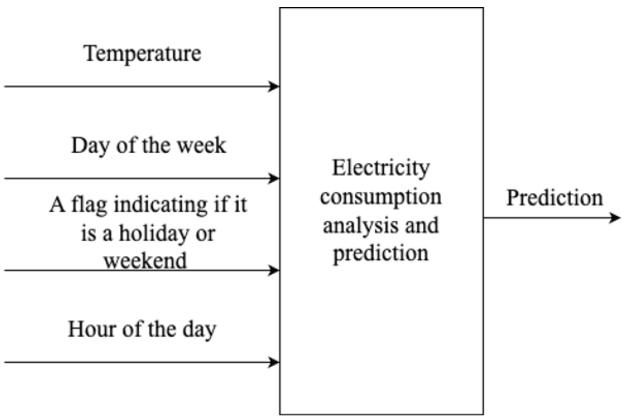


Fig. 2 Influencing factors for electricity prediction

1.Data Collection

The data collection and dataset utilized in the development of the Electricity Consumption Analysis and Prediction played a pivotal role in ensuring the functionality of the system. Diverse data sources were seamlessly integrated to ensure the Electricity Consumption Analysis and Prediction System had access to precise and up-to-date information.

Our project used hourly electrical consumption and temperature data for the Baneshwor area.

- Historical Electrical Consumption Data: The hourly electrical data was collected from the Kathmandu Grid situated at Baneshwor. The data was manually recorded into files and wasn't available readily in digital format. We manually collected data from the year 2078-2080.
- Weather Data: The hourly temperature data for the Baneshwor area was obtained from the official NASA POWER website and recorded in Excel format.

2. Data Processing

- Data Cleaning: In data cleaning, the raw hourly electrical data was converted to Excel format, addressing missing values and entry errors. Cubic spline interpolation handled missing and inconsistent values, ensuring final values were consistent by rounding off to two decimal points. This process enhanced the dataset's quality for subsequent analysis and modelling.
- Feature Engineering: Additional features were created to enhance the dataset's predictive capabilities. This involved deriving new variables or transforming existing ones based on the requirements of the predictive models.
- Data Integration: Dataset made up of 2.6 years of hourly data equivalent to 23449 rows of data was generated by combining individual data of electric load and temperature. Electric data was found to be in B.S. date while temperature was found in AD format. Since NEA used B.S. format, all were converted to that format and combined.

YEAR	MONTH	DAY	HOUR	DAY_OF_THE_WEEK	IS_HOLIDAY	TEMPERATURE	ELECTRICITY
2,078	1	1	1	4	0	20.35	0.80
2,078	1	1	2	4	0	19.79	0.80
2,078	1	1	3	4	0	19.30	0.80
2,078	1	1	4	4	0	19.01	0.80
2,078	1	1	5	4	0	18.93	2
2,078	1	1	6	4	0	19.22	1.40
2,078	1	1	7	4	0	22.08	1.80
2,078	1	1	8	4	0	25.87	2.10
2,078	1	1	9	4	0	28.02	2
2,078	1	1	10	4	0	29.45	1.90
2,078	1	1	11	4	0	30.46	1.70
2,078	1	1	12	4	0	30.79	1.60
2,078	1	1	13	4	0	30.89	1.60
2,078	1	1	14	4	0	30.64	1.50
2,078	1	1	15	4	0	29.95	1.50
2,078	1	1	16	4	0	28.81	1.50
2,078	1	1	17	4	0	26.69	1.60
2,078	1	1	18	4	0	23.98	1.80
2,078	1	1	19	4	0	23.73	2.30
2,078	1	1	20	4	0	23.57	2.10
2,078	1	1	21	4	0	22.21	1.50
2,078	1	1	22	4	0	21.01	1.40
2,078	1	1	23	4	0	19.93	1.20

Fig. 3 Final Dataset

- Data splitting: The dataset was divided into training and testing sets to facilitate model training and evaluation. 80% of the total data was used for training the model whereas 20% of the data was used for testing the model.

3. Model Training

The created dataset was fed to the prediction system which utilized K-nearest neighbors' algorithm, linear regression algorithm, XGBoost algorithm. For all of the algorithms, the model was trained with 80% of the dataset and 20% used for testing.

4. Model Selection

Table 1
Comparison of evaluation metrics

Metric	linear_regression	knn	xgboost
MSE	0.54	0.08	0.03
RMSE	0.73	0.28	0.18
r2_score	0.21	0.88	0.95

In terms of MSE and RMSE, lower values were preferred, indicating better accuracy. XGBoost outperformed both Linear Regression and KNN in these metrics, showcasing its ability to make more accurate predictions. Similarly, in the r2_score, XGBoost excelled with the highest value, indicating better explanation of data variations. Therefore, based on these evaluation metrics, XGBoost emerged as the optimal model for predicting electricity consumption, demonstrating superior accuracy compared to the other models.

B. System Architecture

The Electricity Consumption Analysis and Prediction System integrates Python for backend development, Streamlit for frontend interface creation, and PHPMyAdmin for database management. Streamlit facilitates intuitive web application development directly from Python scripts, offering features like prediction by month, year, and day, as well as electricity consumption analysis. The backend, implemented in Python using Flask and machine learning libraries, preprocesses data from an Excel file and trains an XGBoost regressor model. PHPMyAdmin serves as the database platform, managing user authentication, login, and registration functionalities securely. The system workflow involves user interaction through a web browser, authentication managed by PHPMyAdmin, access to Streamlit dashboard for functionality, routing requests to Flask backend for processing, and data storage and retrieval through PHPMyAdmin. Real-time updates ensure users receive accurate and up-to-date information on electricity consumption and prediction.

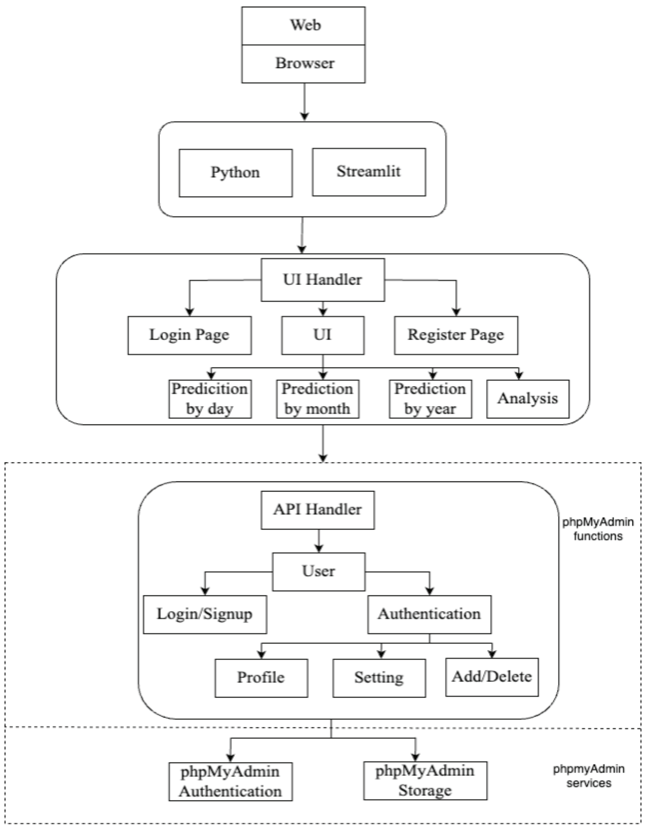


Fig. 4 System Architecture

C. Algorithms Used

1. XGBoost Algorithm:

XGBoost, an ensemble learning algorithm known for its exceptional performance in various machine learning tasks, utilizes gradient boosting to combine predictions from multiple weak learners, resulting in a strong predictive model. Key parameters such as max_depth, eta, gamma, and subsample are crucial for optimizing its performance. In our project, XGBoost was employed for electricity demand forecasting due to its efficiency, scalability, and ability to handle complex datasets. The XGBRegressor class from the xgboost library facilitated model instantiation and training, leading to robust predictive models. Evaluation metrics like MSE, RMSE, and R-squared were utilized to assess model performance, ensuring accuracy and reliability. Finally, the trained XGBoost model was deployed in a Flask web application for practical use, contributing to the project's success in delivering impactful results.

- Input Data Set Acquisition: The project started with acquiring the input data set comprising features such as 'year', 'month', 'day', 'hour', 'Day of the Week', 'is holiday', and 'temperature', alongside the target variable 'electricity'.
- Model Training:The XGBoost algorithm was trained using the acquired data set, sequentially constructing decision trees to minimize residuals and enhance predictive accuracy.
- Residuals Calculation and Prediction:Predictions were made on .the test data set, and residuals were computed using metrics like MSE, RMSE, and R-squared.
- Model Evaluation: The performance of the trained XGBoost model was assessed using various evaluation metrics.
- Model Deployment: The final trained XGBoost model was integrated into a Flask web application for practical deployment, enabling predictions based on input parameters such as date, time, and temperature.

The inclusion of XGBoost in the project's modeling pipeline enhanced prediction accuracy and contributed to achieving project objectives effectively.

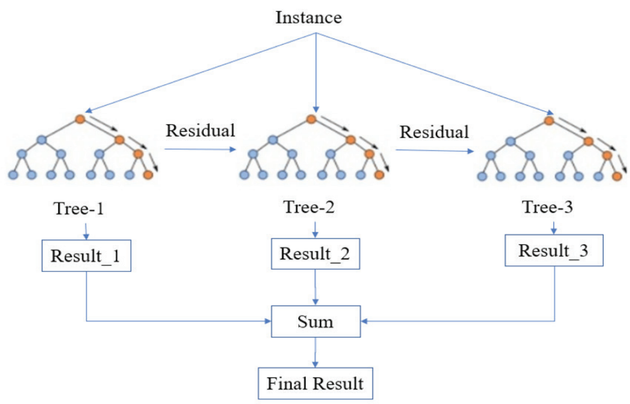


Fig. 5 XGBoost Algorithm

2. Cubic Spline Interpolation

Cubic spline interpolation offers a robust method for filling missing data points within a dataset, ensuring continuity and smoothness in the interpolated curve. In our project, we utilized cubic spline interpolation to address missing elements specifically within the 'electricity consumption' column. Here's a concise overview of our approach:

i. *Data Preprocessing:* We initially loaded the dataset from an Excel file using the pandas library and removed any duplicate entries to maintain data integrity.

ii. *Missing Data Imputation:* Identification of missing values within the 'electricity consumption' column enabled us to segregate them from non-missing entries, preparing them for the interpolation process.

iii. *Interpolation:* Leveraging the CubicSpline class from the scipy library, we fitted a cubic spline function to the non-missing data points. This function was then utilized to interpolate the missing values seamlessly.

iv. *Data Saving:* The interpolated values were rounded to two decimal places and seamlessly integrated back into the DataFrame. Finally, the updated DataFrame was saved to an Excel file for subsequent analysis.

By employing cubic spline interpolation, we ensured the reliability and robustness of our dataset, laying a solid foundation for further analysis and modeling.

3. Inter-Quartile Range

In data-driven models, outliers can significantly impact the integrity of analysis, potentially skewing conclusions. Detecting and addressing outliers is crucial for accuracy. One effective method is the Inter-Quartile Range (IQR) approach, which involves these steps: calculating the IQR as the difference between the third (Q3) and first (Q1) quartiles, determining a threshold typically multiplied by a constant (e.g., 1.5 or 3), flagging data points beyond this threshold as potential outliers, and deciding whether to retain or remove them based on their impact. Following these steps systematically ensures robust analysis. In the procedure, we:

i. Find Q1.

ii. Find Q3.

iii. Calculate IQR (Q3-Q1).

iv. Define the normal data range: Q1–1.5*IQR to Q3+1.5*IQR.

v. Remove any data point outside this range as an outlier.

In a boxplot, this method sets the maximum point (end of high whisker) at Q3+1.5*IQR and the minimum point (start of low whisker) at Q1-1.5*I

Result and Analysis

In conclusion, the visualizations reveal significant patterns in electricity consumption. Saturdays consistently emerge as the days with the lowest average consumption, while weekdays exhibit the highest. Poush and Magh stand out with the highest consumption levels, attributed to the winter season, whereas Shrawan records the lowest consumption, possibly due to summer conditions. Additionally, autumn and spring register the lowest overall consumption levels. The provided KPIs, encompassing Total Consumption, Average Consumption, Maximum Consumption, Minimum Consumption, and Peak Hour Range (18:00 to 22:00), offer valuable insights into consumption trends, facilitating comprehensive analysis and informed decision-making.

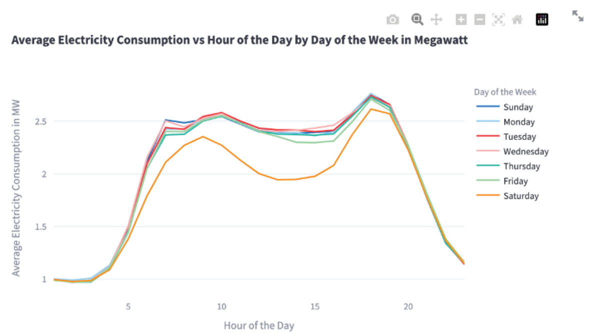


Fig. 6 Line Chart - Average Electricity Consumption by Hour of the Day for Each Day of the Week

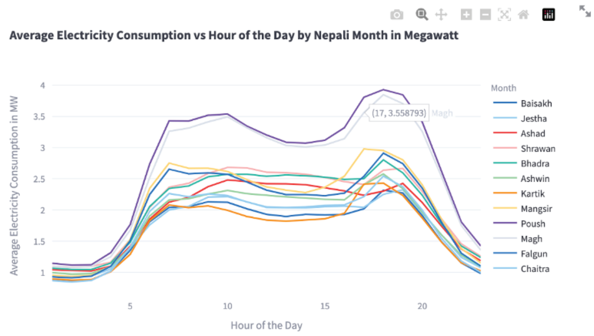


Fig. 7 Line Chart - Average Electricity Consumption by Hour of the Day for Each Nepali Month

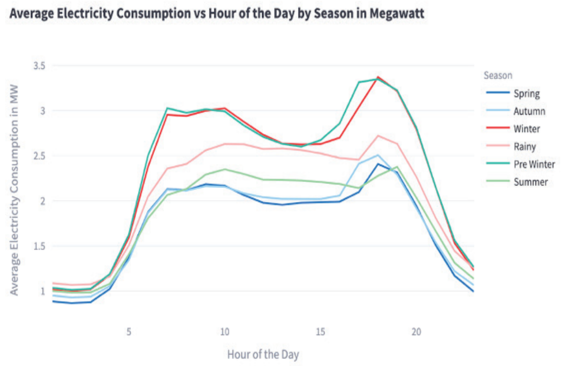


Fig. 8 Line Chart - Average Electricity Consumption by Hour of the Day, Categorized by Seasons

Major Key Performance indicators (KPIs)

Total Consumption: 44847.84 megawatt

Average Consumption: 2.00 megawatt

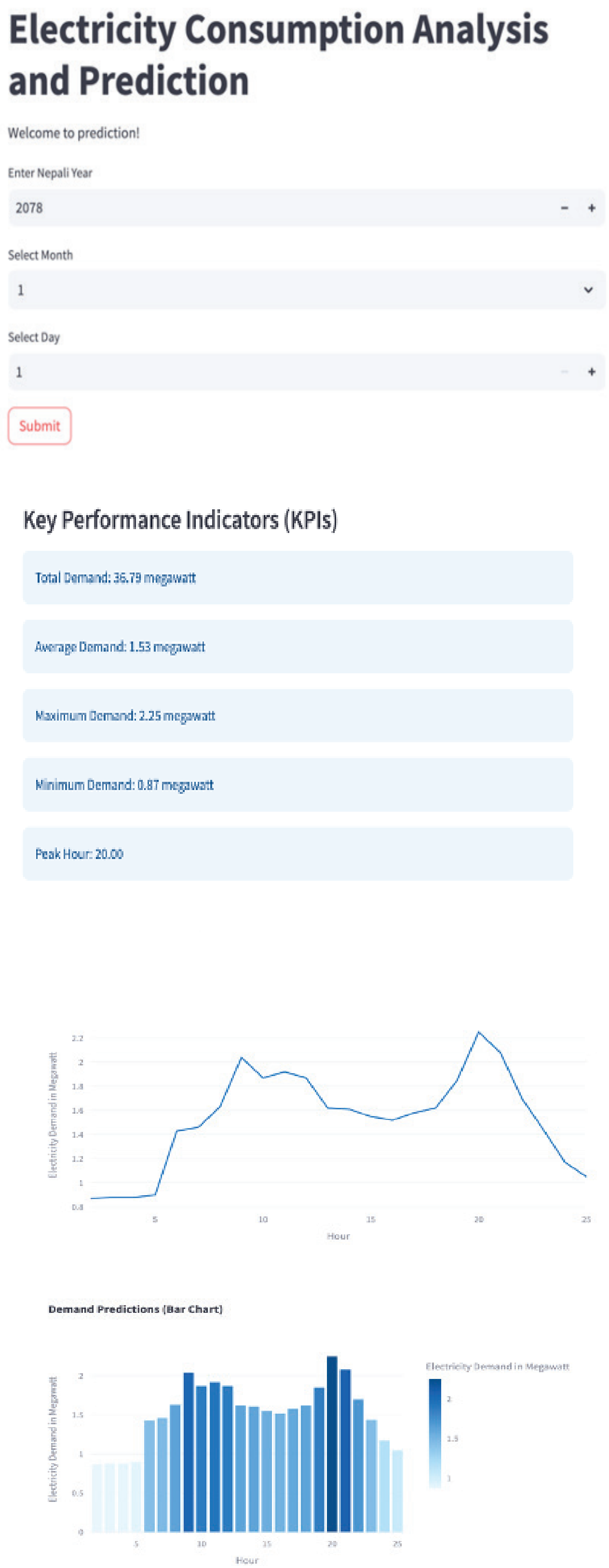
Maximum Consumption: 5.60 megawatt

Minimum Consumption: 0.10 megawatt

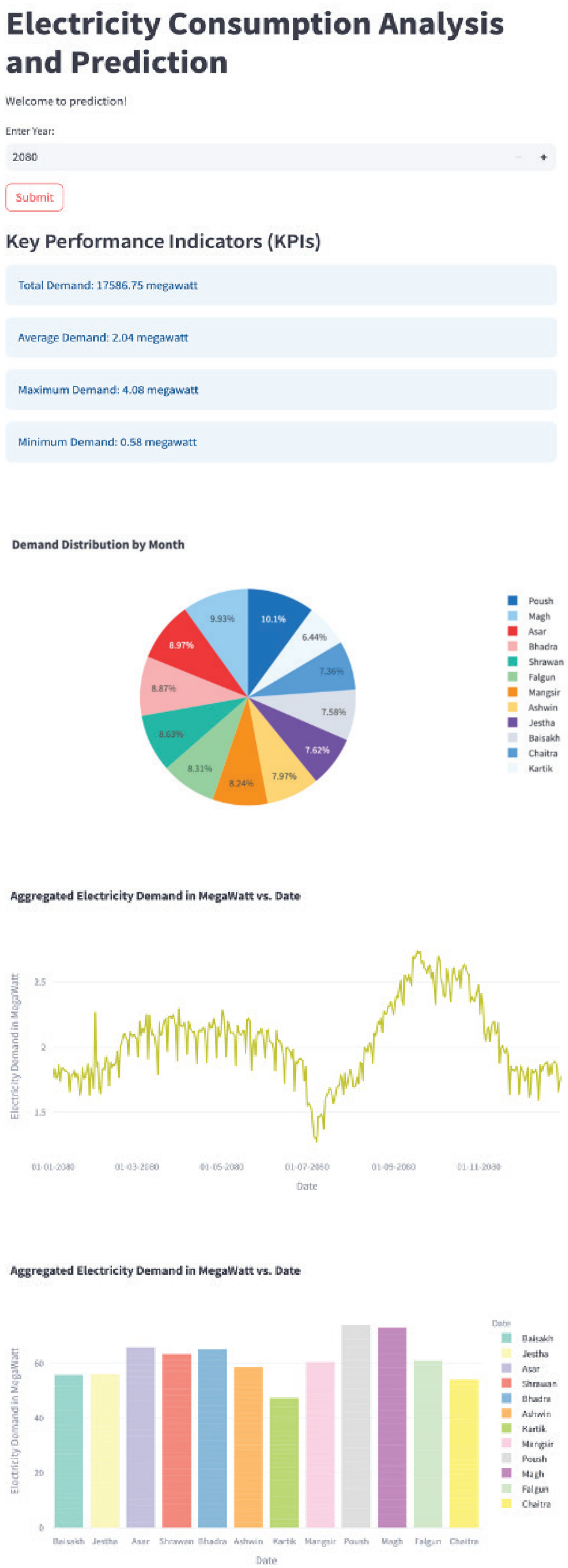
Peak Hour Range: 18:00 - 22:00

Fig. 9 Average KPIs of the electricity consumption from year 2079-2080

In result, the insightful visualization of electricity consumption patterns provides actionable predictions across various time scales, including monthly, daily, and yearly intervals. These visualizations allow for a deeper understanding of consumption trends, facilitating informed decision-making and resource management strategies.


Electricity Consumption Analysis and Prediction
Welcome to prediction!
Enter Nepali Year
2078
Select Month
1
Select Day
1
Submit
Key Performance Indicators (KPIs)
Total Demand: 36.79 megawatt
Average Demand: 1.53 megawatt
Maximum Demand: 2.25 megawatt
Minimum Demand: 0.87 megawatt
Peak Hour: 20.00
Hour
Demand Predictions (Bar Chart)
Electricity Demand in Megawatt
Hour


Electricity Consumption Analysis and Prediction
Welcome to prediction!
Enter Year:
2080
Enter Month:
1
Submit
Key Performance Indicators (KPIs)
Total Demand: 1333.72 megawatt
Average Demand: 1.79 megawatt
Maximum Demand: 4.08 megawatt
Minimum Demand: 0.76 megawatt
Aggregated Electricity Demand in MegaWatt vs Date
Electricity Demand in MegaWatt
date
Aggregated Electricity Demand in MegaWatt vs Date
Electricity Demand in MegaWatt
is_holiday
False
True
date



References

[1] P. M. Sebastián and S. P. Eduardo, “Alternative policies for the liberalization of retail electricity markets in Chile,” Utilities Policy, 2017.

[2] D. Streimikiene and I. Siksnelyte, “Sustainability assessment of electricity market models in selected developed world countries,” Renewable & Sustainable Energy Reviews, vol. 57, pp. 72–82, 2016.

[3] L. Bartelj, D. Paravan, A. F. Gubina, and R. Golob, “Valuating risk from sales contract offer maturity in electricity market,” International Journal of Electrical Power & Energy Systems, vol. 32, no. 2, pp. 147–155, 2010.

[4] Q. Zhu and T. Prieto-Rumeau, “Bias and overtaking optimality for continuous-time jump Markov decision processes in Polish spaces,” Journal of Applied Probability, vol. 45, no. 2, pp. 417–429, 2008.

[5] S. Altug, B. Tan, and G. Gencer, “Cyclical dynamics of industrial production and employment: Markov chain-based estimates and tests,” Journal of Economic Dynamics & Control, vol. 36, no. 10, pp. 1534–1550, 2012.

[6] L. Min, H. Jiang, Y. Huang, and X. Song, “Combination Load Forecast with Time-varying Weights Based on Markov Chain,” in Proceedings of the Chinese Society of Universities for Electric Power System & Its Automation, vol. 23, pp. 131–134, 2011.

[7] W. Sun and Y. Xu, “Research on China’s energy supply and demand using an improved Grey-Markov chain model based on wavelet transform,” Energy, vol. 118, pp. 969–984, 2017.

[8] E. Mocanu, P.H. Nguyen, M. Gibescu, W.L. “Kling Deep learning for estimating building energy consumption Sustainable Energy, Grids and Networks,” 6 (2016), pp. 91-99.

[9] Y. Goude, R. Nedellec, N. Kong “Local short and middle term electricity load forecasting with semi-parametric additive models IEEE Trans Smart Grid,” 5 (1) (2013), pp. 440-446.

[10] Government of Nepal, “Electricity Demand Forecast Report”, Jan 2017

[11] Government of Nepal, “Energy Demand Projection 2030: A MAED Based Approach”, 2016