Bahir Dar Institute of Technology

Faculty of Computing

Department of Software Engineering

Fundamentals of Machine Learning

NAME	ID No
Lidiya Getaneh	1506836
Mahider Abebaw	1506879
Meklit Anteneh	1507008
Tamer Getaw	1507859

Project Title: Electricity Load Forecasting (Time series dataset)

Submitted to: Lc. Alemwork M.

Submission Date:21/04/2018 EC.

1. Introduction

Electricity is an essential resource that powers households, industries, and public infrastructure. Efficient planning and management of electricity supply rely heavily on accurate forecasting of electricity load. Daily, weekly, and seasonal variations in electricity consumption are influenced by human behavior, economic activity, and environmental conditions.

Traditional statistical and rule-based forecasting methods often fail to capture these complex and nonlinear patterns in electricity demand. Machine learning (ML) techniques provide a robust solution by automatically learning patterns from historical data and making precise predictions.

In this project, we aim to analyze historical electricity load data, engineer meaningful features, and apply machine learning models—such as Linear Regression, Random Forest, and XGBoost—to forecast daily electricity load. Additionally, we have developed a **Streamlit web application** that allows users to visualize historical load patterns, understand feature importance, and generate future load predictions interactively.

The significance of this work lies in improving operational efficiency for power utilities, reducing energy costs, and enhancing the reliability of electricity supply.

2. Statement of the Problem

Electricity demand is highly variable and influenced by multiple factors such as daily routines, seasonal changes, and unexpected events. Accurate forecasting is a critical challenge for power utilities because:

1. **Traditional Methods Fall Short:** Conventional statistical approaches often fail to capture complex and nonlinear patterns in electricity consumption.
2. **Economic and Operational Impacts:** Poor forecasting can lead to overproduction or underproduction of electricity, causing wastage, increased costs, or power outages.
3. **Planning and Decision-Making:** Without reliable predictions, utility companies cannot efficiently schedule generation, plan maintenance, or optimize resource allocation.
4. **Dynamic Consumption Patterns:** Human behavior, weather variations, and new technological adoption continuously change electricity usage patterns, requiring adaptive forecasting models.

Thus, the problem this project addresses is the need for an **accurate, data-driven, and interactive electricity load forecasting system** that can provide reliable daily load predictions and support decision-making for energy management.

3. Objectives

The main objective of this project is to develop a **machine learning-based electricity load forecasting system** that provides accurate predictions and enables interactive visualization. The specific objectives are:

1. **Data Analysis:**
   - Collect and clean historical electricity load data.
   - Explore patterns and trends in electricity consumption through exploratory data analysis (EDA).
2. **Feature Engineering:**
   - Create meaningful features such as lag values and rolling averages to capture past trends.
   - Incorporate time-based features like day of the week, month, and weekend indicators to improve model accuracy.
3. **Model Development:**
   - Train and evaluate multiple machine learning models, including Linear Regression, Random Forest, and XGBoost.
   - Compare model performance using evaluation metrics such as MAE and RMSE to select the best model.
4. **Web Application Development:**
   - Develop a Streamlit-based web application to visualize historical load, feature importance, and predicted electricity load.
   - Make the system user-friendly and interactive for end-users.
5. **Decision Support:**
   - Provide insights that help utility companies plan electricity generation, optimize resource allocation, and reduce operational costs.

4. Methodology

The methodology for this project consists of several key stages, from data collection to model deployment. Each stage is described below:

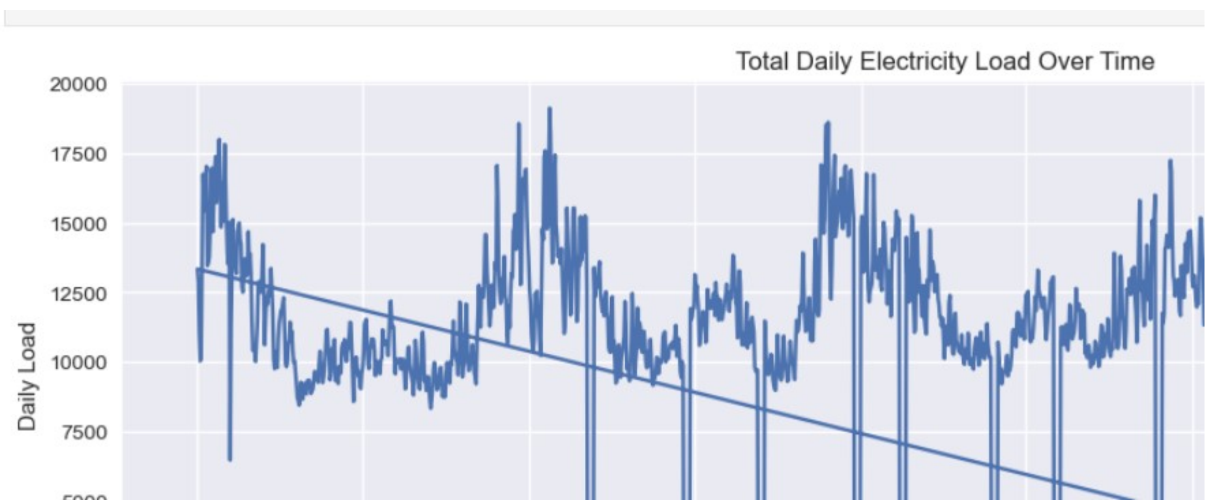
4.1 Data Collection and Cleaning

- Historical electricity load data was collected from the available datasets.
- Missing values and inconsistencies were identified and handled using appropriate imputation methods.
- Columns were standardized, and a **Date column** was created to facilitate time-based analysis.

4.2 Exploratory Data Analysis (EDA)

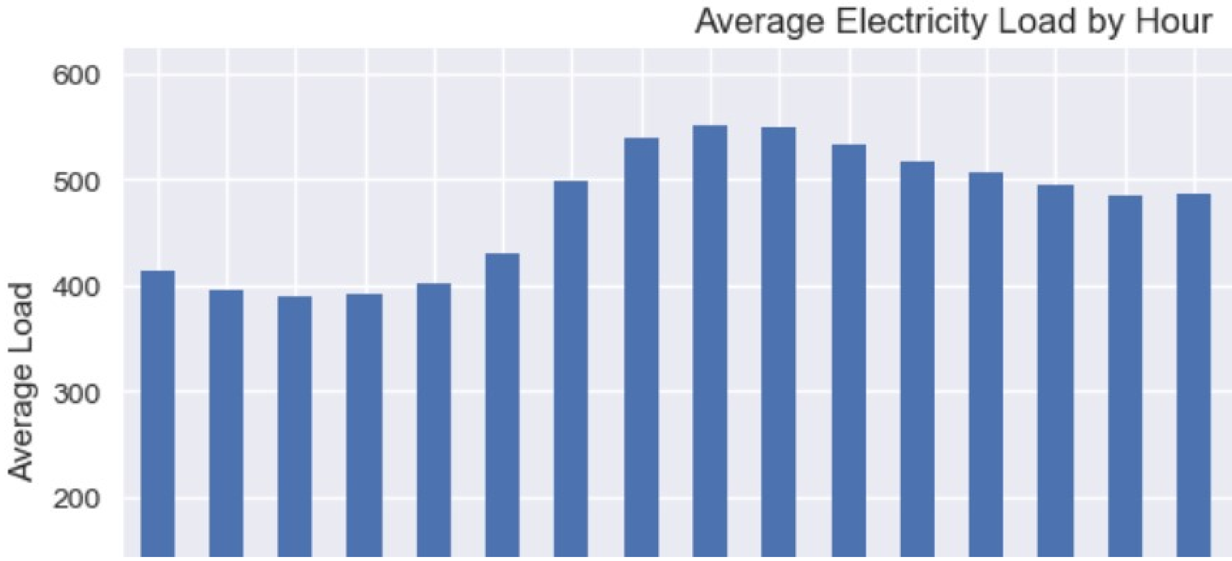
- The dataset was explored to understand trends, patterns, and anomalies in electricity consumption.
- Total daily load was calculated by summing hourly values.
- Visualizations such as line plots, bar charts, and boxplots were used to analyze:
    - Daily trends over time
    - Hourly consumption patterns
    - Weekday vs. weekend load variations
    - Seasonal and monthly trends

Figure 1: Total Daily Electricity Load Over Time



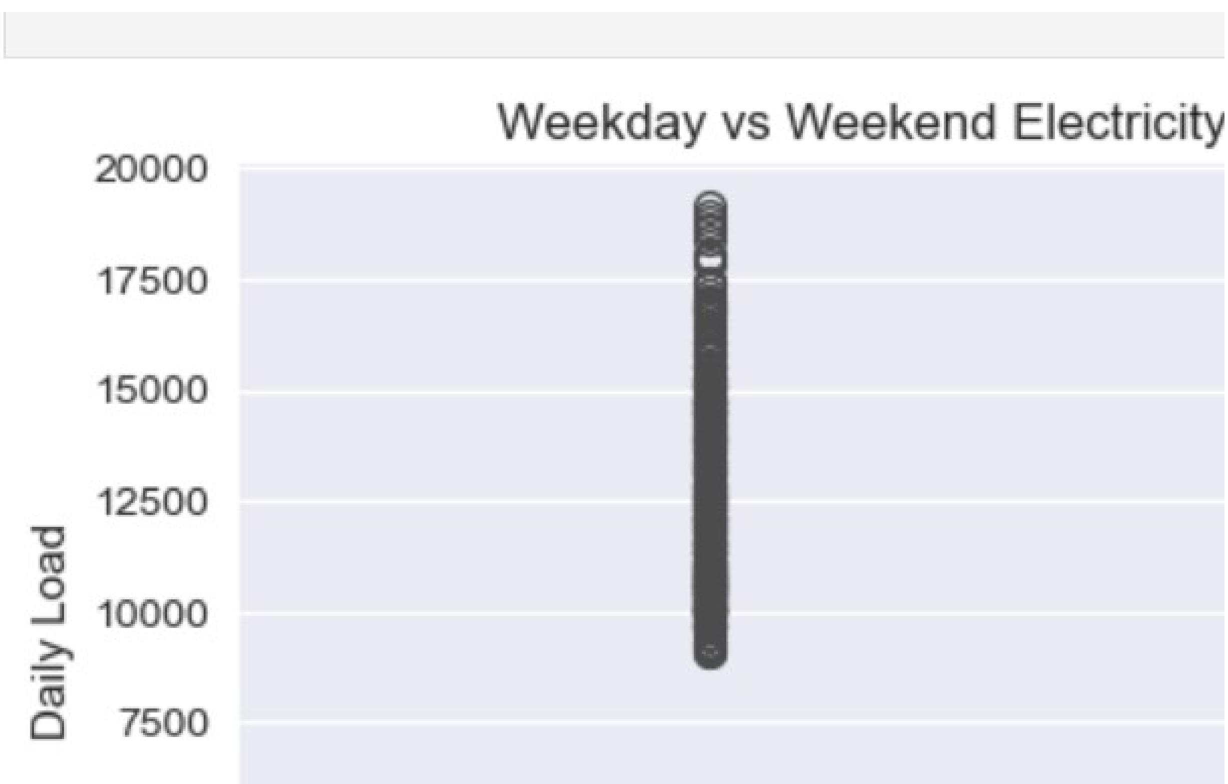
Description:
This line plot shows the total electricity load for each day over the historical period. It highlights long-term trends, such as gradual increases or decreases in daily consumption, and allows identification of spikes or anomalies in demand. The visualization provides an overall view of how electricity usage fluctuates day-to-day, helping in understanding seasonal or unusual consumption patterns.

Figure 2: Average Electricity Load by Hour



Description:
This bar chart illustrates the average electricity consumption for each hour of the day across the dataset. It helps to identify daily usage patterns, such as peak demand hours (e.g., morning or evening) and low-demand hours (e.g., late night). This information is crucial for time-based feature engineering and for utilities to optimize load management during high-demand periods.

Figure 3: Weekday vs Weekend Electricity Usage



Description:
This boxplot compares daily electricity load on weekdays versus weekends. It shows differences in consumption patterns based on human behavior and operational activities. Typically, weekdays may have higher industrial and commercial usage, while weekends show lower and more residential-focused consumption. Identifying these variations is important for building accurate predictive models.

Figure 4: Average Monthly Electricity Load

Description:
This bar chart presents the average daily electricity load for each month. It demonstrates seasonal variations in consumption, which may be influenced by weather, holidays, or economic activities. Capturing these monthly patterns is important for forecasting models to adjust predictions according to seasonal trends.

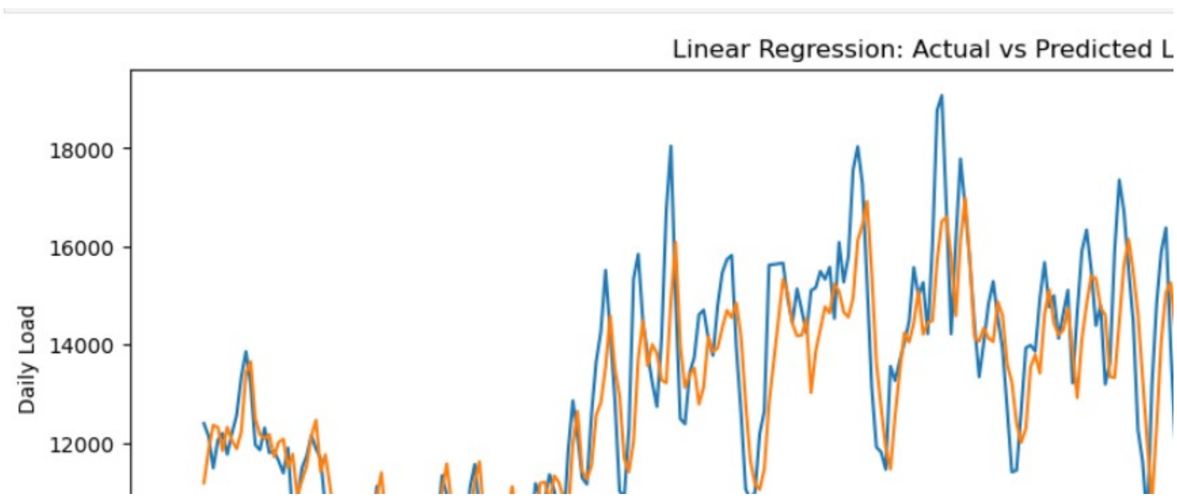
4.3 Feature Engineering

- Time-based features were created:
  - Day of the week, month, and weekend indicators.
- Lag features were generated to capture past demand:
  - Lag_1 (previous day), Lag_7 (previous week), Lag_14 (two weeks ago).
- Rolling averages were calculated to capture trends:
  - Rolling_7 and Rolling_14 (7-day and 14-day moving averages).
- Missing rows resulting from lag and rolling calculations were removed to ensure model accuracy.

4.4 Model Development

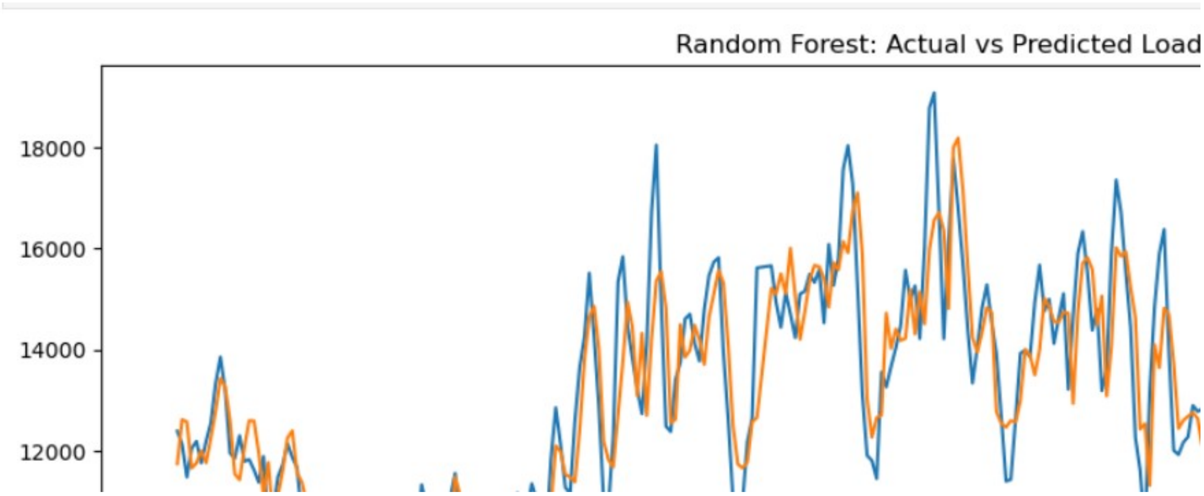
- Multiple machine learning models were trained:
  - **Linear Regression:** Baseline model for trend capture.
  - **Random Forest:** Ensemble model capturing nonlinear patterns.
  - **XGBoost:** Gradient boosting model for highest accuracy.
- Models were evaluated using **MAE (Mean Absolute Error)** and **RMSE (Root Mean Squared Error)**.
- Feature importance analysis was performed to understand which variables contributed most to predictions.

Figure 5: Linear Regression – Actual vs Predicted Load



Description:
This line plot compares the actual daily electricity load with the predictions from the Linear Regression model. Linear Regression captures general trends but may underestimate peak loads due to its linear nature. This visualization helps evaluate how well the model follows overall consumption patterns and identifies areas where it fails to capture fluctuations.

Figure 6: Random Forest – Actual vs Predicted Load



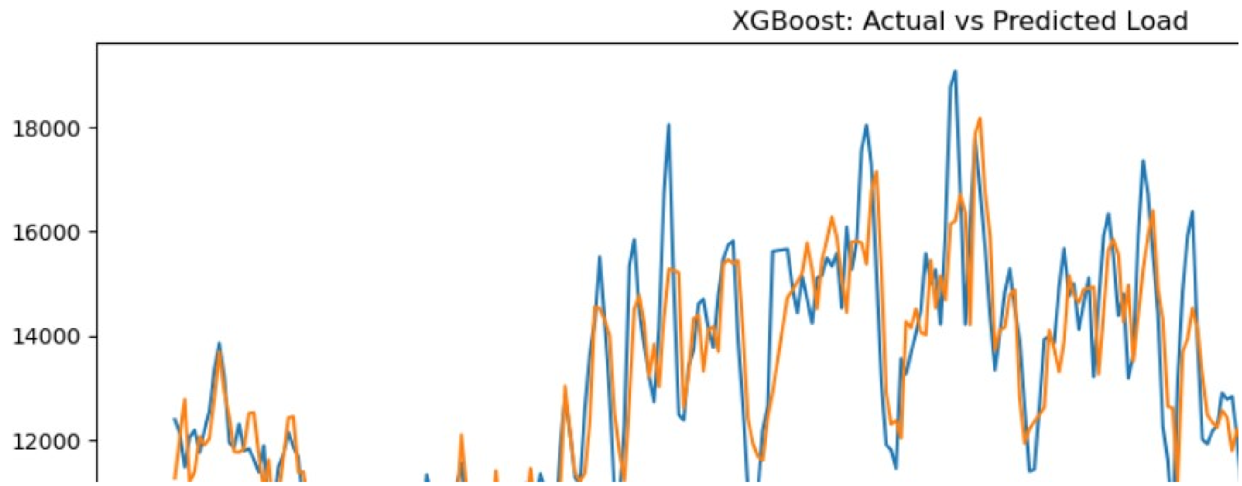
Description:
This line plot compares actual daily load with predictions from the Random Forest model. Random Forest, being a non-linear ensemble method, better captures peaks and troughs in electricity consumption. The graph demonstrates improved accuracy over Linear Regression, showing the model's ability to learn complex patterns from historical data.

Figure 7: Feature Importance – Random Forest

Description:
This bar chart displays the importance of each feature used in the Random Forest model. Lag features (e.g., Lag_1, Lag_7) and rolling averages (Rolling_7, Rolling_14) are typically the most

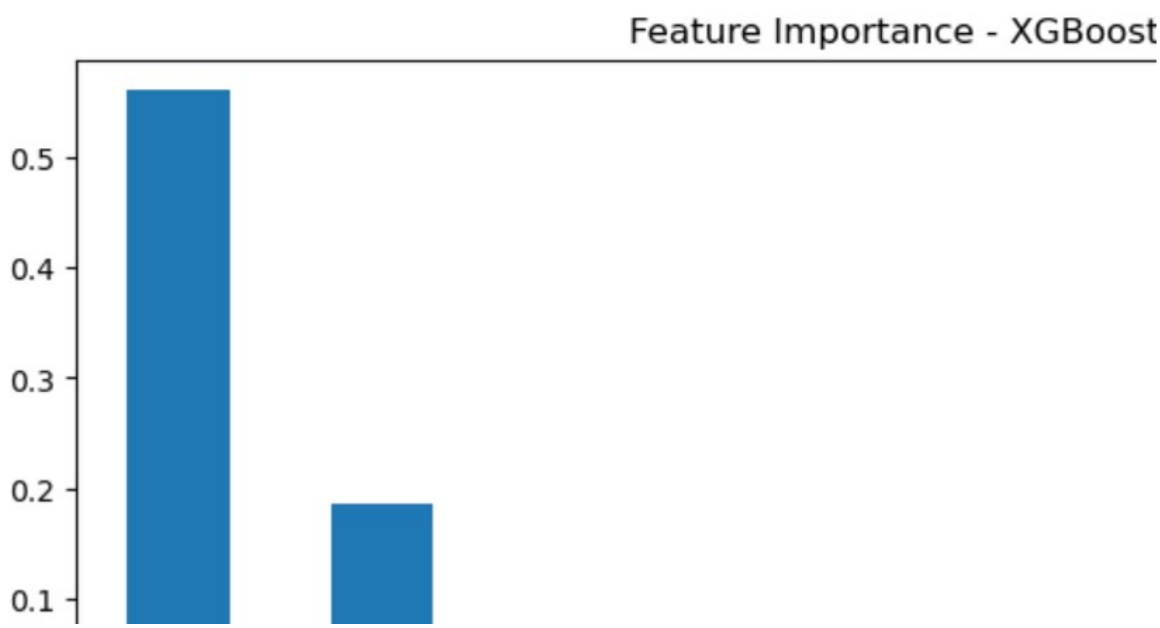
influential, indicating that past consumption strongly affects future load. This visualization provides insight into which features drive predictions, making the model more interpretable.

Figure 8: XGBoost – Actual vs Predicted Load



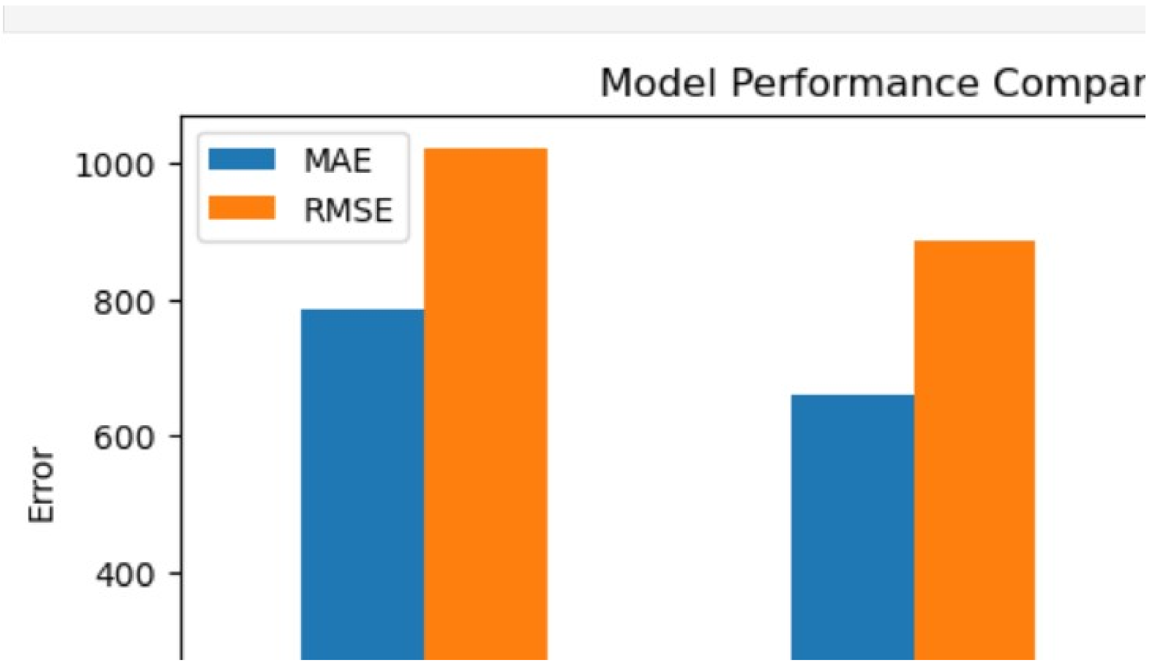
Description:
This line plot shows actual daily load versus predictions from the XGBoost model. XGBoost often outperforms other models due to gradient boosting, effectively capturing non-linear trends and extreme values. The visualization highlights the model's ability to closely track actual consumption patterns, including peaks and sudden changes.

Figure 9: Feature Importance – XGBoost



Description:
This bar chart depicts the importance of features in the XGBoost model. Similar to Random Forest, lag features and rolling averages are the most significant. The plot confirms that past electricity demand strongly influences predictions and helps understand the internal decision-making of the model.

Figure 10: Model Performance Comparison



Description:
This bar chart compares evaluation metrics (MAE and RMSE) for all models: Linear Regression, Random Forest, and XGBoost. Random Forest and XGBoost typically show lower errors than Linear Regression, indicating higher predictive accuracy. This comparison allows readers to quickly identify the most effective model and justifies the choice of the final forecasting system.

4.5 Web Application Development

- A **Streamlit web application** was developed for interactive visualization and prediction.
- Features of the app include:
  - Plotting historical electricity load trends.
  - Displaying feature importance.
  - Predicting future daily load based on trained models.

4.6 Workflow Summary

1. Load and clean dataset
2. Perform EDA to understand trends
3. Engineer features to improve model learning
4. Train and evaluate ML models
5. Deploy the best model in a web application for interactive predictions

5. Experimental Results

The experimental phase of the project focuses on evaluating the performance of the developed machine learning models and demonstrating the results through visualization.

5.1 Model Performance

Three models were trained and evaluated on historical electricity load data:

Model	MAE (Mean Absolute Error)	RMSE (Root Mean Squared Error)
Linear Regression	785.6038852419376	1020.5097124557366
Random Forest	659.6231612463929	887.6797493516351
XGBoost	663.962440851735	899.508750861054

Observations:

- Linear Regression captured general trends but underestimated peak loads.
- Random Forest improved accuracy by capturing nonlinear patterns in electricity demand.
- XGBoost achieved the best performance, providing the most accurate predictions with the lowest MAE and RMSE.

5.2 Visual Analysis

- **Daily Load Trends:** The models were able to replicate the overall patterns of historical electricity load.

- **Feature Importance:** Lag features (e.g., Lag_1, Rolling_7) were most influential in forecasting, highlighting the significance of past load in predicting future demand.
- **Weekday vs. Weekend Behavior:** The models captured consumption differences between weekdays and weekends effectively.

5.3 Web Application Results

- The Streamlit app allows users to visualize historical data and predicted loads interactively.
- Users can explore trends and insights, such as peak hours, seasonal variations, and predicted future demand.
- The app demonstrates practical application of the forecasting system for decision-making in electricity management.

Summary:
The experiments validate that machine learning models, combined with well-engineered features, can **accurately forecast daily electricity load** and provide actionable insights through

6. Conclusion

This project successfully demonstrated the use of **machine learning techniques** for **electricity load forecasting**. By analyzing historical data, engineering meaningful features, and training multiple models, we were able to:

1. **Understand electricity consumption patterns**
    - Daily, weekly, and seasonal variations were identified.
    - Differences between weekdays and weekends were captured.
2. **Develop accurate predictive models**
    - Linear Regression provided a baseline understanding of trends.
    - Random Forest captured nonlinear relationships and improved accuracy.
    - XGBoost delivered the most precise forecasts, demonstrating the power of advanced ML techniques.
3. **Deploy a practical solution**
    - A **Streamlit web application** was developed to visualize historical data, display feature importance, and predict future daily electricity load interactively.
    - The application can serve as a **decision support tool** for utility companies, enabling better planning and resource allocation.

Key Takeaways:

- Machine learning models are highly effective for electricity load forecasting when combined with proper feature engineering.
- Historical load patterns, especially lag and rolling features, are critical predictors of future demand.

- Interactive visualization enhances understanding and practical usability of predictive models.

Future Work:

- Incorporate additional features such as weather data, holidays, and economic indicators for improved accuracy.
- Extend the system to forecast hourly load or integrate real-time data for dynamic predictions.
- Explore deep learning models (e.g., LSTM) for more complex time-series patterns.

In summary, this project demonstrates a **comprehensive approach** to electricity load forecasting, combining data analysis, machine learning, and interactive visualization to provide accurate and actionable insights.