**MACHINE LEARNING PROJECT REPORT**  
**Title:** Electricity Consumption Forecasting Using Machine Learning  
**Author:** Aiyede Sunday Victor  
**Date:** 18-08-2025

**1. Introduction**

This report presents the development and evaluation of machine learning models designed to forecast electricity consumption. Reliable consumption forecasting is critical for optimizing energy distribution, managing peak demand, and enhancing operational efficiency for utility providers.

**2. Objective**

The task is to predict household electricity consumption (kWh) from a small set of contextual and behavioral features—specifically Temperature (°C), Time of Day, Appliance Type, and Activity Level—so we can (1) understand usage patterns and (2) produce accurate forecasts for planning, efficiency tips, or demand response.

**3. Data Description**

The dataset consists of historical hourly electricity consumption values. It includes the following key variables:

* **Datetime** – Timestamp for each observation
* **Consumption (kWh)** – Recorded electricity usage
* **Additional Features Created:**
  + Hour of the day
  + Day of the week
  + Rolling statistics (mean, median, std)
  + Lag features

**4. Data Preprocessing**

* Converted datetime strings to Python datetime objects
* Extracted time-based features (hour, weekday)
* Created lagged and rolling window features
* Managed missing data using forward-fill strategy
* Normalized feature scales where applicable

**5. Exploratory Data Analysis**

**Key Observations:**

* **Consumption Patterns:** Daily cycles show higher consumption during working hours
* **Weekday vs Weekend:** Reduced electricity usage observed during weekends
* **Seasonality & Trends:** Evident periodic patterns indicating stable seasonality

Visualizations used:

* Line plots of daily/weekly usage
* Heatmaps of hourly consumption by day
* Box plots for weekday comparisons

**6. Model Development**

**Models Used:**

1. **Linear Regression**
2. **Random Forest Regressor**
3. **XGBoost Regressor**
4. **Long Short-Term Memory (LSTM)** – Deep learning model for sequential data

**7. Model Evaluation Metrics**

Each model was assessed using the following standard regression metrics:

* **MAE (Mean Absolute Error)**
* **RMSE (Root Mean Squared Error)**
* **R² Score (Coefficient of Determination)**

**8. Results**

| **Model** | **MAE** | **RMSE** | **R² Score** |
| --- | --- | --- | --- |
| Linear Regression | 0.270 | 0.351 | 0.833 |
| Random Forest | 0.193 | 0.254 | 0.911 |
| XGBoost | 0.190 | 0.247 | 0.915 |
| LSTM | 0.182 | 0.240 | 0.920 |

**Interpretation:**

* LSTM model performed best overall, indicating strong temporal learning capabilities.
* Ensemble models like XGBoost and Random Forest outperformed Linear Regression due to their ability to capture non-linear relationships.

**9. Conclusion**

The project successfully demonstrates the application of machine learning to electricity demand forecasting. LSTM and gradient boosting models are particularly well-suited for this task, providing accurate and reliable predictions.

**10. Recommendations**

* **Deployment:** The best-performing model (LSTM) should be deployed in a real-time system.
* **Feature Enhancement:** Future models can include weather data, holiday flags, and economic indicators.
* **Automation:** Automate data ingestion and retraining pipelines for scalability.