**PHASE 2**

**PREDICTING ENERGY CONSUMPTION PATTERNS USING TIME SERIES FORECASTING FOR SMART GRIDS**

**STUDENT NAME: S.MADHUMITHA**

**REGISTER NUMBER: 611823243033**

**INSTITUTION: P.S.V COLLEGE OF ENGINEERING AND TECHNOLOGY**

**DEPARTMENT:B.TECH ARTIFICAL INTELLIGENCE AND DATA SCIENCE**

**DATE OF SUBMISSION:10/05/2025**

**GITHUB REPOSITORY LINK:** [**github.com/Madhushree120**](https://github.com/Madhushree120)

**1. Problem Statement**

* *Sustainable energy management. This data-driven initiative addresses the operational challenge of accurately forecasting hourly, daily, and weekly energy demand to enable efficient load balancing, The goal is to predict energy consumption patterns in smart grids using advanced time series forecasting techniques to optimize energy distribution, enhance grid reliability, and support peak demand management, and integration of renewable energy sources. The analysis is critical for real-world decision-making, empowering utility companies to reduce operational costs, minimize energy waste, and improve grid resilience against demand fluctuations. This project employs predictive analytics, leveraging historical consumption data and external factors (e.g., weather, holidays) to develop robust forecasting models. The insights will guide strategic planning, such as scheduling maintenance, optimizing renewable energy storage, and informing policy for energy-efficient urban development.*

**2. Project Objective**

**Goals**:

* + *Build and validate time series forecasting models to predict energy consumption with high accuracy.*
  + *Identify key drivers of energy demand to inform grid management strategies.*
  + *Provide actionable recommendations for optimizing energy allocation and reducing peak load costs.*

**Key Questions**:

* + *What are the dominant temporal patterns (e.g., hourly, daily, seasonal) in energy consumption?*
  + *How do external factors like temperature, humidity, and holidays impact consumption?*
  + *Which forecasting models (e.g., ARIMA, Prophet, LSTM) perform best for short- and long-term predictions?*
  + *How can predictive insights improve load balancing and renewable energy integration?*

**Deliverables**:

* + *A suite of time series models with performance metrics (e.g., RMSE, MAE, MAPE).*
  + *Interactive visualizations of consumption trends, seasonality, and forecast accuracy.*
  + *A comprehensive report with insights, trends, and recommendations for grid operators.*
  + *A dashboard prototype for real-time consumption monitoring and forecasting.*

**Objective Evolution**:

* *Initially, the focus was on basic trend analysis and simple forecasting models. After deeper data exploration revealed complex seasonality and external influences, objectives expanded to include multivariate models, feature engineering, and real-time dashboard development to enhance practical utility.*

**3. Flowchart of the Project Workflow**

graph TD

*A[Data Collection] --> B[Data Cleaning]*

*B --> C[Feature Engineering]*

*C --> D[Exploratory Data Analysis]*

*D --> E[Model Development]*

*E --> F[Model Evaluation]*

*F --> G[Visualization & Dashboard]*

*G --> H[Reporting & Recommendations]*

*H --> I[Deployment Preparation]*

**4. Data Description**

* **Dataset Name and Source**:
  + *Primary: Smart Grid Energy Consumption Dataset (UCI Machine Learning Repository).*
  + *Secondary: Weather Data (NOAA Weather API for temperature and humidity).*
* **Data Type**:
* *Structured (time series data in tabular format, merged with external weather data).*
* **Number of Rows and Columns**:
  + Primary Dataset: ~150,000 rows, 10 columns.
  + *Merged Dataset: ~150,000 rows, 12 columns (after adding weather features).*
* **Static or Dynamic**:
* *Dynamic, with real-time updates possible through API integration for ongoing data collection.*

**Key Fields**:

* *timestamp: Date and time of measurement (YYYY-MM-DD HH:MM).*
* *consumption\_kWh: Energy consumption in kilowatt-hours.*
* *temperature: Ambient temperature (°C).*
* *humidity: Relative humidity (%).*
* *is\_holiday: Binary indicator for holidays (0 or 1).*
* *day\_of\_week: Categorical variable for weekday/weekend effects.*
* *wind\_speed: Wind speed (km/h, from weather data).*
* *solar\_radiation: Solar radiation intensity (W/m², for renewable energy context).*

**5. Data Preprocessing**

**Handling Missing Values**:

* + *Imputed missing consumption\_kWh using linear interpolation to maintain time series continuity.*
  + *Filled missing temperature and humidity with forward-fill for short gaps or mean values for longer gaps, based on seasonal patterns.*
  + *Dropped rows with missing timestamp to ensure chronological integrity.*

**Removing Duplicates**:

* + *Identified and removed duplicate rows based on timestamp and consumption\_kWh to avoid biased modeling*.

**Formatting and Parsing Data**:

* + *Standardized timestamp to datetime format (UTC) for consistency.*
  + *Extracted temporal features: hour, day, month, year, and season for pattern analysis.*
  + *Converted day\_of\_week to categorical labels (e.g., Monday, Tuesday).*

**Encoding Categorical Variables**:

* + *One-hot encoded is\_holiday and day\_of\_week for model compatibility.*
  + *Label-encoded season (e.g., Winter=1, Spring=2) for certain models.*

**Identifying and Treating Outliers**:

* *Applied IQR method to detect outliers in consumption\_kWh and temperature.*
* *Capped outliers at the 1st and 99th percentiles to minimize distortion while preserving data variability.*

**Transformations**:

* *Normalized consumption\_kWh, temperature, and humidity using Min-Max scaling for neural network models.*
* *Applied log transformation to consumption\_kWh to reduce skewness.*
* *Created lagged variables (e.g., consumption\_kWh\_lag1, consumption\_kWh\_lag24) for time series modeling.*

**for Transformations Reasons**:

* *Transformations ensured data compatibility with statistical and machine learning models, preserved time series properties, and enhanced model robustness by addressing skewness and outliers.*

**6. Exploratory Data Analysis (EDA)**

* **Univariate Analysis**:
* *Histogram of consumption\_kWh: Right-skewed distribution, with most values between 50–200 kWh.*
* *Time series plot of consumption\_kWh: Revealed daily peaks (evenings) and weekly cycles (lower on weekends).*
* *Boxplot of temperature: Showed seasonal variation, with extremes in summer and winter.*

**Bivariate/Multivariate Analysis**:

* *Correlation heatmap: Strong negative correlation (-0.65) between temperature and consumption\_kWh, indicating higher heating demand in colder weather.*
* *Scatterplot of consumption\_kWh vs. humidity: Weak positive correlation, suggesting minor influence.*
* *Pairplot of consumption\_kWh, temperature, wind\_speed, and solar\_radiation: Identified non-linear relationships, especially with renewable energy factors.*
* *Grouped bar plot: consumption\_kWh by day\_of\_week showed lower consumption on weekends; is\_holiday confirmed reduced demand on holidays.*
* *Seasonal decomposition: Isolated trend, seasonality, and residual components, confirming strong daily and yearly seasonality.*

**Key Metrics**:

* *Average hourly consumption: 145 kWh.*
* *Peak consumption: 350 kWh (winter evenings, 6–9 PM).*
* *Seasonal variation: Winter consumption 30% higher than summer.*
* *Holiday impact: 15% lower consumption on holidays.*

**Insights**:

* *Daily and seasonal patterns are driven by residential and industrial activity.*
* *Temperature is the strongest external predictor, followed by day-of-week effects.*
* *Renewable energy factors (e.g., solar\_radiation) show potential for hybrid grid modeling.*
* *Anomalies during extreme weather events suggest the need for robust outlier handling.*

**7. Tools and Technologies Used**

* **Programming Language**: *Python*
* **Notebook/IDE**: *Jupyter Notebook, Google Colab (for GPU support with neural networks)*
* **Libraries**:
  + *Data Manipulation: pandas, numpy*
  + *Visualization: matplotlib, seaborn, plotly, dash (for dashboard prototype)*
  + *Time Series Modeling: statsmodels (ARIMA), prophet, tensorflow (LSTM), pmdarima (auto-ARIMA)*
  + *Evaluation Metrics: scikit-learn*
  + *Feature Engineering: tsfresh (for automated time series feature extraction)*
* **Optional Automation Tools**:
  + *pandas-profiling for rapid EDA reports.*
  + *ydata-profiling for advanced data quality checks.*
* **Other Tools**:
  + *Git for version control.*
  + *Docker for reproducible environments.*
  + *API integration (NOAA Weather API) for dynamic data updates.*

**8. Team Members and Contributions**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | S.NO | NAMES | ROLES | | 1. | S.MAHALAKSHMI | DATA SCIENTIST | | 2. | B.LITHIGA | MACHINE LEARNING ENGINEER | | 3. | S.MADHUMITHA | SMART GRID ANALYST | | 4. | I.MADEEHA IRAM | DEMAND FORECASTING ANALYST | |  |  |
|  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |