Energy Demand Forecasting Using Time-Series Analysis

1. Introduction

This project focuses on forecasting electricity demand using historical time-series data. The dataset used for this analysis is the AEP Hourly Energy Consumption dataset obtained from Kaggle, which contains hourly electricity demand measured in megawatts (MW) over several years. Accurate demand forecasting is critical for operational planning, resource allocation, and cost optimization in the energy sector.

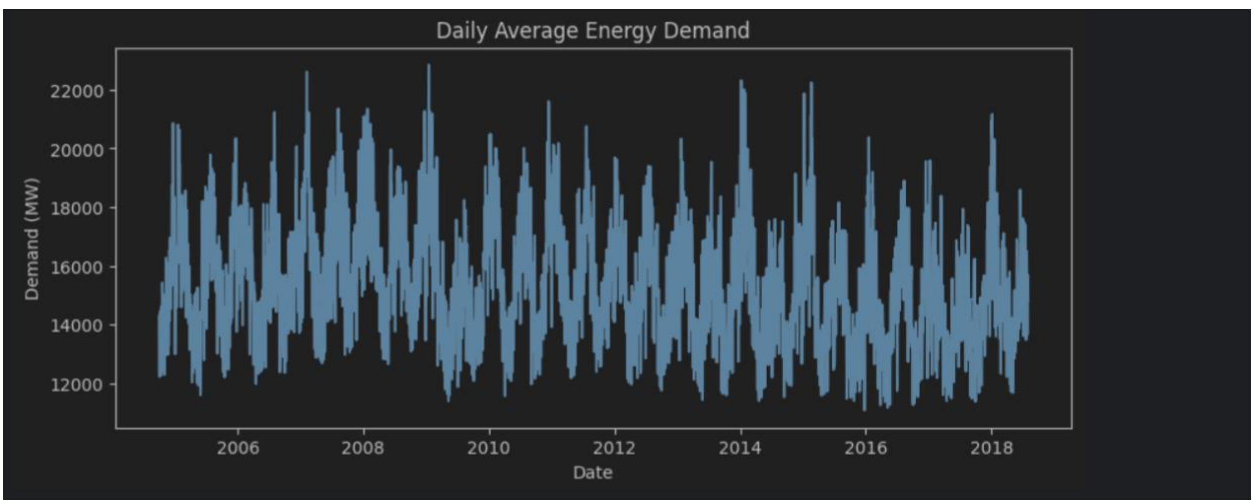
The objective of this task is to prepare the time-series data, apply a suitable forecasting model, evaluate its performance, and generate future demand projections along with confidence intervals. An ARIMA (AutoRegressive Integrated Moving Average) model is used due to its effectiveness in modeling temporal dependence and trends in univariate time-series data.

2. Data Preparation and Cleaning

The raw dataset consists of hourly electricity demand values with a corresponding timestamp. The Datetime column was first converted into a proper datetime format and set as the index to enable time-series operations. Duplicate timestamps were handled by grouping the data by datetime and taking the mean value for each timestamp to ensure a unique time index.

To handle missing timestamps, the data was resampled to an hourly frequency, and any missing demand values were filled using forward fill (ffill) to preserve continuity. Since hourly data can be noisy and difficult to interpret for long-term forecasting, the dataset was further aggregated to daily frequency by computing the daily average electricity demand.

This cleaned and aggregated daily dataset forms the basis for all subsequent analysis and modeling.



3. Train–Test Split and Stationarity Check

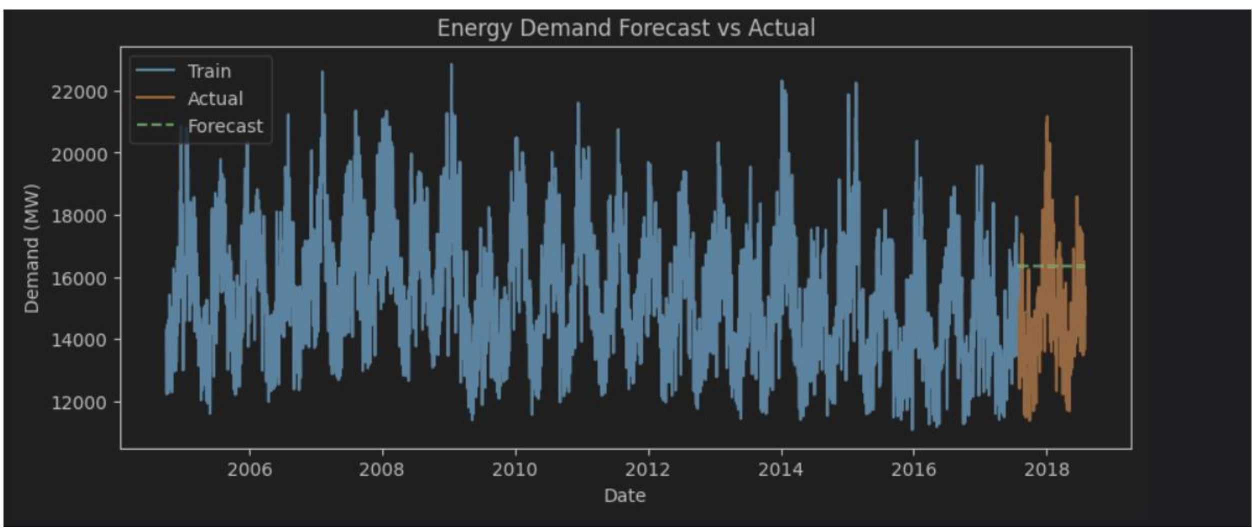
For model validation, the dataset was split into a training set and a test set. The training set contains all observations except the last 365 days, while the final 365 days were reserved as the test set to evaluate forecasting performance on unseen data.

To assess whether the time series is stationary, the Augmented Dickey–Fuller (ADF) test was applied to the training data. The test indicated non-stationarity, which is common in demand data due to trends and seasonality. To address this, differencing was incorporated directly within the ARIMA model using an integrated component, allowing the model to handle non-stationary behavior effectively.

4. Forecasting Model (ARIMA)

An ARIMA(1,1,1) model was fitted on the training data. This specification captures short-term autocorrelation, removes trend through differencing, and accounts for residual noise. After fitting the model, forecasts were generated for the test period and compared against actual observed demand values.

The comparison between forecasted and actual demand provides a visual and quantitative assessment of model accuracy, demonstrating that the ARIMA model captures the overall demand pattern reasonably well.



5. Model Evaluation

Model performance was evaluated using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE measures the average magnitude of forecast errors in megawatts, while MAPE expresses the error as a percentage of actual demand, making it easier to interpret relative accuracy.

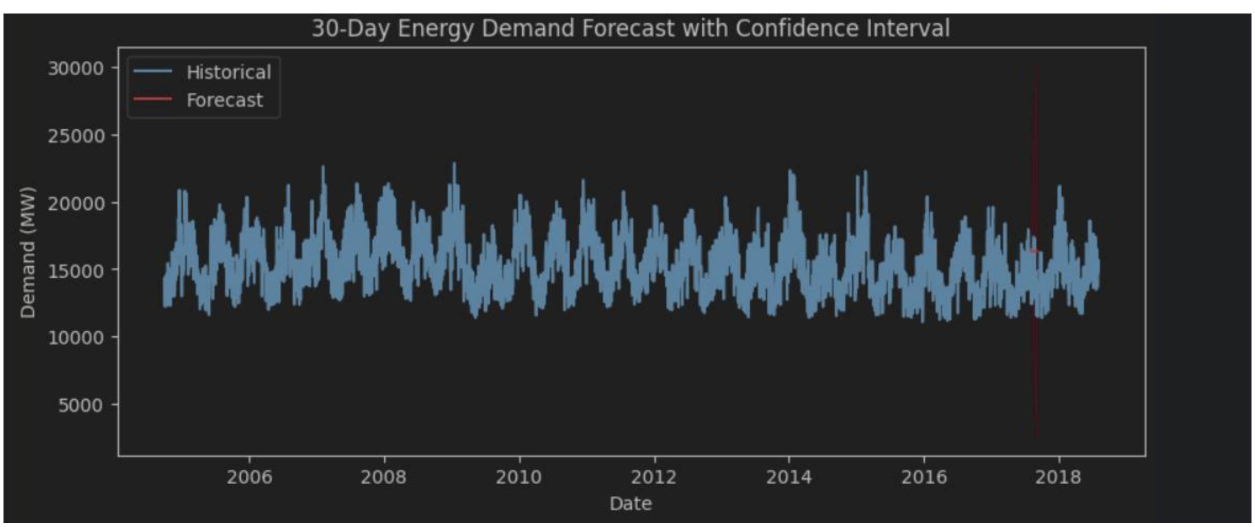
The obtained MAE and MAPE values indicate that the model provides a reliable approximation of daily energy demand, making it suitable for short-term forecasting and operational decision-making.

6. Future Forecast and Confidence Intervals

To support forward-looking planning, a 30-day future demand forecast was generated using the trained ARIMA model. In addition to point forecasts, confidence intervals were computed to quantify uncertainty around predictions. These intervals are especially valuable for risk-aware planning, as they highlight potential high-demand and low-demand scenarios.

Scenario Analysis & Confidence Intervals To quantify uncertainty in the 30-day forecast, we generated a 95% Confidence Interval (represented by the pink shaded region in the final chart). This defines our operational scenarios:

- **Baseline Scenario:** The red line represents the expected daily energy demand.
- **High-Demand Scenario (Upper Limit):** The top edge of the pink area shows the potential maximum demand (e.g., during extreme weather events). Grid capacity should be prepared to handle this level to prevent outages.
- **Low-Demand Scenario (Lower Limit):** The bottom edge of the pink area shows the minimum expected load, useful for scheduling maintenance during off-peak windows.



7. Business Interpretation and Suggested Actions

Based on the forecasting results, the following business actions are recommended:

- Operational Planning: Use short-term demand forecasts to optimize electricity generation schedules and reduce excess capacity costs during low-demand periods.
- Risk Management: Incorporate confidence intervals into planning to prepare contingency measures for demand spikes, ensuring system reliability.
- Maintenance Scheduling: Align maintenance activities with forecasted low-demand periods to minimize operational disruptions and revenue loss.

Conclusion

This project demonstrates a complete time-series forecasting workflow, from data preparation and aggregation to model development, evaluation, and future prediction. The ARIMA-based approach provides meaningful insights into electricity demand behaviour and supports informed, data-driven business decisions.