**Energy Demand Forecasting for PJM Interconnection**

Time-Series Analysis and Machine Learning Approaches

**Annanahmed Shaikh**  
 School of Computing and Data Science  
 Wentworth Institute of Technology  
 Massachusetts, Boston, USA  
 [shaikha4@wit.edu](mailto:shaikha4@wit.edu)

ABSTRACT

Energy demand forecasting is crucial for efficient energy management and operational planning. This project leverages time-series modeling techniques, including Naïve, Exponential Smoothing, ARIMA, SARIMA, and Linear Regression, to forecast 24-hour energy consumption for the PJM Interconnection. Using hourly data from the DUQ zone, this study identifies seasonal trends, evaluates model performance, and explores the implications of forecast errors. The SARIMA model achieved the highest accuracy, highlighting its suitability for seasonal data. These findings contribute to optimizing grid operations and minimizing costs.

∗

KEYWORDS

Energy Forecasting, Time-Series Analysis, Machine Learning, PJM Interconnection, SARIMA

1 Introduction

Energy forecasting plays a pivotal role in balancing supply and demand, reducing operational costs, and ensuring grid stability. Accurate predictions enable grid operators to allocate resources efficiently and avoid disruptions. The PJM Interconnection dataset provides granular hourly energy consumption data, making it an ideal choice for evaluating time-series forecasting models. This project addresses key questions, including identifying seasonal patterns, comparing forecasting methods, and assessing model accuracy. The study contributes to the existing literature by combining traditional time-series models with machine learning approaches.

2 Data

**2.1 Source of Dataset**

The dataset, **DUQ\_hourly.csv**, was sourced from a publicly available Kaggle repository. It contains hourly energy consumption data for the DUQ zone, part of the PJM Interconnection. The dataset's credibility lies in its detailed and granular time-series structure, generated using utility records.

**2.2 Characteristics of the Dataset**

* **Format:** CSV file
* **Size:** ~500,000 rows
* **Columns:**
  + **Datetime:** Timestamp of energy consumption (hourly granularity)
  + **DUQ\_MW:** Energy consumption in megawatts (MW)
* **Preprocessing Steps:**
  + Missing values were handled by forward-filling techniques.
  + Time-based features (Year, Month, Day, Hour) and flags (Weekday, Weekend) were created to capture seasonal patterns.
* **Visualization:** Seasonal trends were identified using boxplots and ACF/PACF plots.

3 Methodology

**3.1 Naïve Forecasting**

* **Description:** Uses the last observed value for future predictions.
* **Advantage:** Simplicity and low computational cost.
* **Disadvantage:** Ignores trends and seasonality**.**

**3.2 Exponential Smoothing**

* **Description:** Captures level and trend with additive seasonality.
* **Advantage:** Suitable for short-term forecasting with minimal data requirements.
* **Disadvantage:** Limited in capturing complex seasonal patterns.

**3.3 ARIMA**

* **Description:** Combines autoregressive and moving average terms for univariate time series.
* **Advantage:** Effective for short-term trends.
* **Disadvantage:** Struggles with seasonality.

**3.4 SARIMA**

* **Description:** Extends ARIMA with seasonal components.
* **Advantage:** Handles complex seasonal data efficiently.
* **Disadvantage:** Requires careful parameter tuning.

**3.5 Linear Regression**

* **Description:** Reduces features to lagged variables and uses regression for prediction**.**
* **Advantage:** Explores feature importance.
* **Disadvantage:** Less effective for high-frequency seasonality.

4 Results

**4.1 Model Performance**

* **Naïve Model:** MAE = 22.15, RMSE = 30.50
* **Exponential Smoothing:** MAE = 15.10, RMSE = 20.87
* **ARIMA:** MAE = 13.22, RMSE = 18.33
* **SARIMA:** MAE = 12.05, RMSE = 16.10
* **Linear Regression:** MAE = 14.50, RMSE = 19.45

**4.2 Seasonal Trends**

* Strong 24-hour and monthly seasonality patterns were observed.
* Winter months exhibited higher energy consumption compared to summer.

**4.3 Graphical Insights**

* Forecast vs Actual plots highlighted the SARIMA model's accuracy.
* Error distribution plots revealed the limitations of simpler models like Naïve forecasting.

**5 Discussion**

While SARIMA provided the most accurate forecasts, it requires extensive parameter tuning, which can be computationally expensive. Linear regression, although interpretable, lacked accuracy for high-frequency seasonal data. Future work could integrate external factors like weather and economic indicators to improve predictions. Ensemble models combining multiple techniques may also yield better results.

6 Conclusion

This project successfully demonstrated the effectiveness of SARIMA for forecasting energy demand in the PJM Interconnection. The findings underscore the importance of capturing seasonality in time-series data. Accurate forecasts can optimize grid operations, reduce costs, and ensure sustainable energy management.

ACKNOWLEDGMENTS

The author acknowledges the guidance of Wentworth Institute of Technology **faculty,** and the resources provided for completing this project.

**REFERENCES**

1. Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
2. Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. (2015). *Time Series Analysis: Forecasting and Control*. Wiley.Conference Name:ACM Woodstock conference
3. Analytics Vidhya. (2023, August 15). *Time Series Forecasting Simplified: ARIMA, SARIMA, and Beyond*. [YouTube video.](https://youtu.be/7NXCdfzr5d8?si=hVF5T0-06cGkfGhL" \t "_new)

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00