**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.

|  |  |  |
| --- | --- | --- |
| Model | MAE | RMSE |
| Naïve Forecasting | 22.15 | 30.50 |
| Exponential Smoothing | 15.10 | 20.87 |
| ARIMA | 13.22 | 18.33 |
| SARIMA | 12.05 | 16.10 |
| Linear Regression | 14.50 | 19.45 |

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.

**4.4 Seasonal Patterns in Energy Demand**

1. **Monthly Energy Consumption Patterns**

A graph with numbers and lines

Description automatically generated with medium confidence

**Figure 1.1** Monthly Energy Consumption

* + **Purpose**: Demonstrates seasonal variations and is sufficient to infer yearly, monthly, and daily patterns.

**4.5 Model Comparison**

1. **Forecast vs Actual for Best-Performing Model (SARIMA)**

A graph with blue and orange lines

Description automatically generated

**Figure 2.1** Forecast vs Actual Values for SARIMA Model

* + **Purpose**: Highlights the accuracy of the best model without needing to show all models' results.

1. **Error Distribution Across Models**

A graph with a line and numbers

Description automatically generated

**Fig 3.1** AR Model

A graph with a line and numbers

Description automatically generated

**Fig 3.2** ARIMA Model

A graph with lines and numbers

Description automatically generated

**Fig 3.3** Holt-Winters Model

A graph with lines and numbers

Description automatically generated

**Fig 3.4** Linear Regression Model

A graph with a line and numbers

Description automatically generated

**Fig 3.5** MA Model

A graph with blue and orange lines

Description automatically generated

**Fig 3.6** SARIMA Model

* + **Purpose**: Allows direct comparison of model performance in a single image.

**4.3 Error Analysis and Grid Management**

1. **Grid Load Before and After Optimization (SARIMA)**

A graph of a graph

Description automatically generated

**Fig 4.3** Forecast vs Actual Values

* + **Purpose**: Demonstrates practical implications of the project for real-time energy grid management.

**5 Discussion**

**5.1 Caveats**

* **Data Completeness:**  
  The dataset used for forecasting energy demand is restricted to the DUQ zone within the PJM Interconnection and lacks information from other regions, limiting the generalizability of the results. Additionally, gaps in the time-series data due to missing timestamps pose challenges for accurate trend analysis.
* **Model Assumptions:**  
  Models like ARIMA and SARIMA assume stationarity and linear relationships, which might not fully capture complex energy consumption patterns influenced by non-linear and external factors like weather or economic activity.
* **Seasonal and External Factors:**  
  While strong daily and monthly seasonality patterns were identified, external influences such as holidays, extreme weather events, or industrial demands are not explicitly included in the dataset, potentially limiting forecasting precision.
* **Forecasting Limitations:**  
  Simpler models like Naïve and Exponential Smoothing provide quick insights but fail to capture intricate seasonal patterns, leading to higher error rates. Advanced models like SARIMA require careful tuning, which can be computationally expensive and time-intensive.

**5.2 Overcoming the Errors**

* **Enhance Data Completeness**:
  + Interpolate missing data points to maintain continuity in the time series.
  + Collaborate with PJM or similar data providers to obtain datasets with fewer missing fields and higher granularity.
* **Incorporate External Features**:
  + Integrate weather data, industrial output, and holiday calendars to capture external factors influencing energy demand.
  + Use feature engineering to create additional attributes, such as moving averages and lagged values, to improve model performance.
* **Improve Model Validation**:
  + Use cross-validation techniques to assess model robustness across multiple subsets of the dataset.
  + Automate hyperparameter tuning with tools like Grid Search or Bayesian Optimization to identify optimal model configurations.
* **Handle Computational Constraints**:
  + Employ distributed computing frameworks like Apache Spark for efficient processing of larger datasets and complex models.
  + Test lightweight forecasting methods for faster real-time predictions.

**5.3 Future Scope**

* **Scalability and Dataset Expansion**:
  + Extend the analysis to include multiple zones within the PJM Interconnection for broader insights.
  + Expand the dataset to include longer timeframes and additional external features, such as renewable energy integration and market demand data.
* **Advanced Forecasting Techniques**:
  + Experiment with deep learning models like LSTMs and Transformers to capture long-term dependencies and non-linear relationships.
  + Develop ensemble models that combine the strengths of ARIMA, SARIMA, and machine learning approaches for improved accuracy.
* **Dynamic Forecasting and Optimization**:
  + Implement adaptive forecasting systems that adjust predictions based on real-time grid load data.
  + Explore optimization algorithms for scheduling energy distribution based on predicted demand and renewable energy availability.
* **Policy and Sustainability Applications**:
  + Simulate the impact of government policies, such as subsidies or renewable energy mandates, on energy consumption patterns.
  + Incorporate carbon footprint calculations to evaluate and reduce the environmental impact of energy grid operations.
* **User Behaviour Analysis**:
  + Conduct segmentation of energy users to identify usage patterns and offer personalized recommendations for demand-side management.
  + Use longitudinal studies to track changes in consumption behaviour and adapt forecasting strategies accordingly.

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 expresses gratitude to the faculty at Wentworth Institute of Technology for their guidance and support throughout this project. Special thanks to peers for constructive feedback and collaboration.

**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