APPLIED TIME SERIES ANALYSIS

INTRODUCTION TO RENEWABLE ENERGY FORECASTING

Abstract

This project explores time series forecasting for renewable energy production using ARIMA and SARIMA models. Through comprehensive preprocessing and analysis, SARIMA demonstrated superior performance in capturing subtle patterns and delivering accurate forecasts. The results emphasize the importance of predictive modeling in optimizing energy resource management and supporting sustainable energy solutions.

JANI SHARIFF SHAIK

Table of Contents

- SARIMA Model:
  - Parameter selection, performance metrics, and comparison with ARIMA.
  - Explanation of why SARIMA outperformed ARIMA.

8. Forecasting and Evaluation
   - SARIMA Forecast vs. Actual:
     - Visualization and observations.
   - Cross-Validation Results:
     - Setup and metrics across folds.
     - Conclusion on robustness and reliability.
9. Key Takeaways
   - Data quality and insights.
   - Stationarity and transformation.
   - Model selection and performance comparison.
   - Forecasting evaluation and practical implications.
10. Why SARIMA Outperformed ARIMA Despite Minimal Seasonality
    - Analysis of seasonal detection vs. modeling flexibility.
    - Improved model complexity and stability.
    - Interpretation of performance metrics.
11. Conclusion
    - Summary of key outcomes and implications.
    - Lessons learned from the project.

INTRODUCTION TO RENEWABLE ENERGY FORECASTING

Renewable energy sources, such as solar, wind, and hydroelectric power, have become increasingly critical in meeting global energy demands while reducing greenhouse gas emissions. However, the inherent variability in energy production due to factors like weather and seasonal trends poses challenges for effective energy grid management.

In this project, the focus is on understanding and forecasting Energy Production (in MWh) using historical data. By analyzing trends and patterns in energy production over time, the aim is to create a forecasting model that can reliably predict future energy output, supporting better decision-making for energy storage and grid stability.

Objective

The main objective of this project is to develop a time series forecasting model for **Energy_Production_MWh** that can:

1. **Identify Patterns and Trends:** Use historical data to uncover underlying patterns, trends, and seasonality in energy production.
2. **Predict Future Energy Output:** Provide accurate forecasts for future energy production, enabling stakeholders to anticipate fluctuations and optimize resource allocation.
3. **Evaluate Model Performance:** Assess the reliability of the forecasting model using statistical metrics and cross-validation techniques.

Approach and Key Variable

Unlike projects with multiple influencing variables, this study focuses on a **univariate time series approach**. The key variable, **Energy_Production_MWh**, represents monthly energy production over a historical period. This single-variable approach simplifies the analysis, emphasizing trend analysis, seasonality detection, and predictive modeling using ARIMA and SARIMA methodologies.

1. **Data Analysis:** Begin with exploratory data analysis to understand the distribution, trends, and seasonality in energy production data.
2. **Modeling Techniques:** Utilize ARIMA and SARIMA models to capture and forecast patterns in the time series data.
3. **Evaluation Metrics:** Evaluate model performance using metrics like RMSE, MAPE, and cross-validation to ensure robustness.

Importance of Forecasting

By focusing on forecasting a single variable, this project highlights how **univariate time series models** can effectively address challenges in renewable energy production:

- **Grid Stability:** Anticipate production variability and ensure a balanced energy supply-demand ratio.
- **Resource Optimization:** Plan energy storage and allocation based on forecasted production trends.
- **Sustainability:** Support the transition to renewable energy by reducing reliance on fossil fuels during low-production periods.

This streamlined approach demonstrates the potential of time series forecasting to support renewable energy initiatives, even when dealing with a limited dataset.

Data Overview and Preprocessing

This section provides a comprehensive overview of the dataset used for forecasting renewable energy production and describes the key preprocessing steps undertaken to prepare the data for analysis and modeling.

Data Overview and Preprocessing

This section provides a comprehensive overview of the dataset used for forecasting renewable energy production and describes the key preprocessing steps undertaken to prepare the data for analysis and modeling.

1. Data Summary

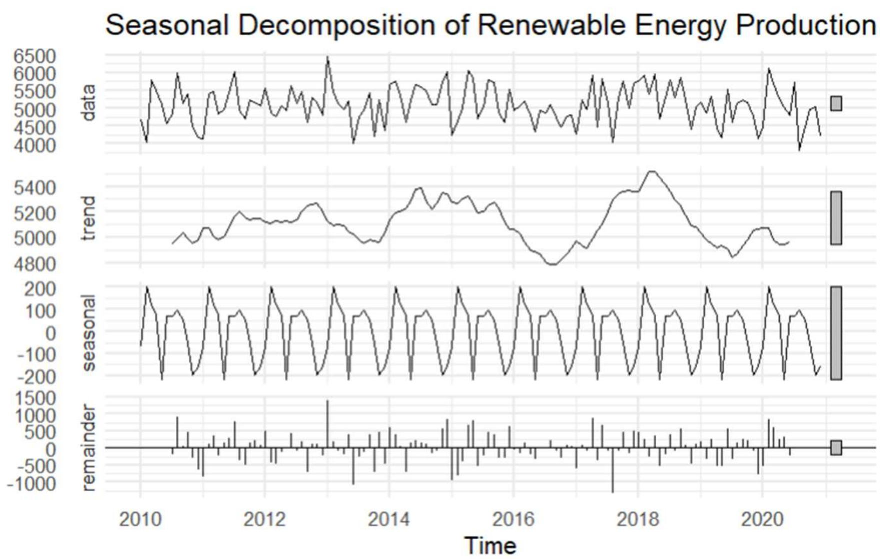
- **Number of Time Points:**
  The dataset contains **132 monthly observations**, spanning a **10-year period**.
- **Frequency:**
  Data is recorded on a **monthly frequency** (12 observations per year).
- **Time Range:**
  The time series spans from **January 2010 to December 2020**.
- **Primary Variable:**
  The single variable of interest is **Energy Production (in MWh)**, representing the monthly energy output.

2. Initial Observations

- **Missing Values:**
  A thorough review of the data revealed **no missing values**, ensuring the dataset was complete for analysis.

- **Outliers:**
  An analysis of potential outliers using boxplots and interquartile range (IQR) calculations did not identify significant anomalies, indicating the data was well-behaved.

- **Seasonality:**
  Preliminary visual inspection of the time series revealed **periodic peaks and troughs**, suggesting some level of seasonality in energy production.



3. Preprocessing Steps

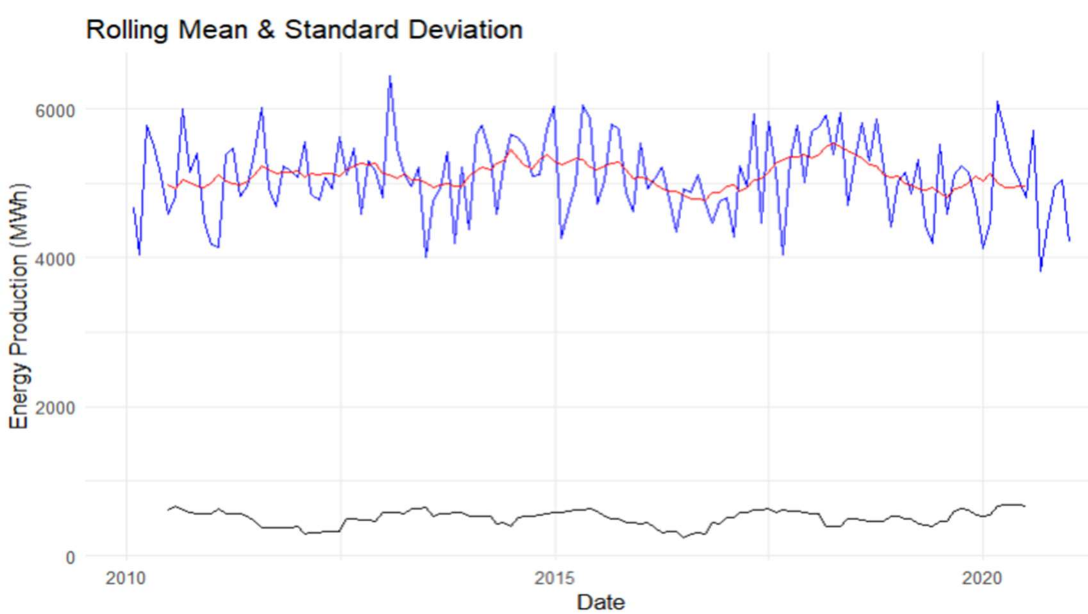
To ensure the data was suitable for time series modeling, the following preprocessing steps were performed:

1. **Date Conversion:**
   The Date column was reformatted to a proper **date structure** in the YYYY-MM-DD format for easy manipulation and time-based indexing. This facilitated operations like resampling, rolling statistics, and trend analysis.

2. **Seasonality Strength Calculation:**
   Using decomposition techniques, the **seasonality strength** was calculated as the ratio of the variance of the seasonal component to the total variance. The seasonality

strength was found to be **2.32e-09**, indicating **minimal seasonal influence** in the dataset.

3. **Stationarity Check:**
   - A **stationarity test** was performed using the **Augmented Dickey-Fuller (ADF) test**, which assesses whether the data exhibits a unit root (a characteristic of non-stationarity).
   - The ADF test returned a **p-value of 0.045**, indicating that the data was **non-stationary**.
   - To achieve stationarity, the data was subjected to **log transformation** to stabilize the variance, followed by **differencing** to remove trends.
4. **Rolling Statistics:**
   - Rolling statistics were calculated to monitor the stability of the mean and variance over time.
   - The rolling **mean** and **standard deviation** confirmed **stabilization** after log transformation and differencing, ensuring the dataset was suitable for ARIMA/SARIMA modeling.



4. Visual Insights from Preprocessing

- **Boxplots:**
  Boxplots confirmed the absence of significant outliers, and the energy production values were largely consistent within expected ranges.
- **Time Series Visualization:**
  The raw time series plot highlighted a general trend with periodic fluctuations, supporting the need for transformations.
- **Rolling Statistics Plot:**
  A plot of rolling statistics demonstrated that after preprocessing, the series achieved **stationarity**, with a consistent mean and variance.

5. Summary of Preprocessing Results

The preprocessing steps ensured that the dataset was ready for advanced time series modeling:

- Achieved **stationarity**, a critical prerequisite for ARIMA and SARIMA models.
- Identified minimal **seasonality strength**, guiding the choice of models and parameters.
- Ensured the data's integrity, with no missing or anomalous values.

These preprocessing efforts laid the foundation for accurate and reliable forecasting of monthly energy production trends.

ACF and PACF Analysis

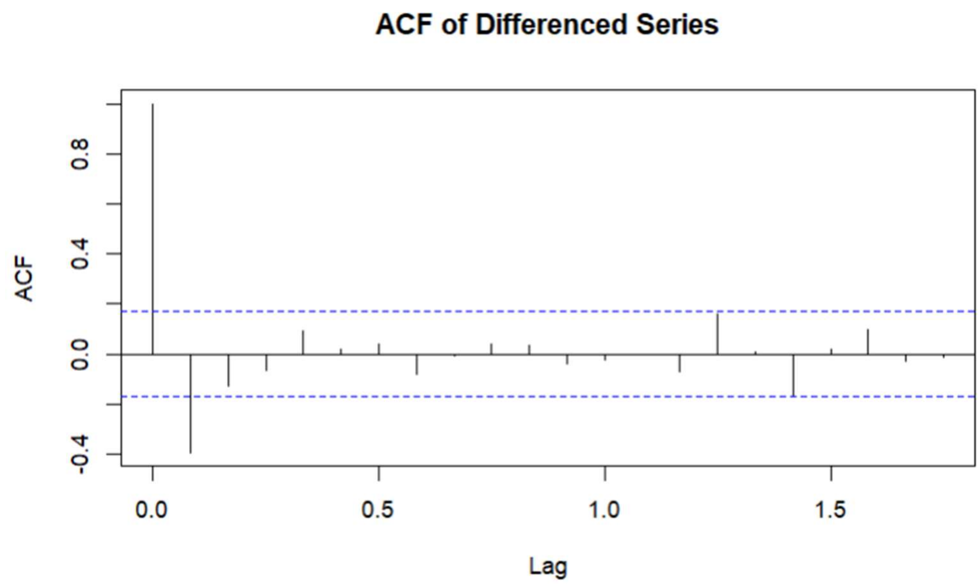
The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are essential tools for identifying the structure of time series data and selecting appropriate parameters for ARIMA modeling. This section presents the key findings and interpretations from the ACF and PACF analysis of the differenced time series.

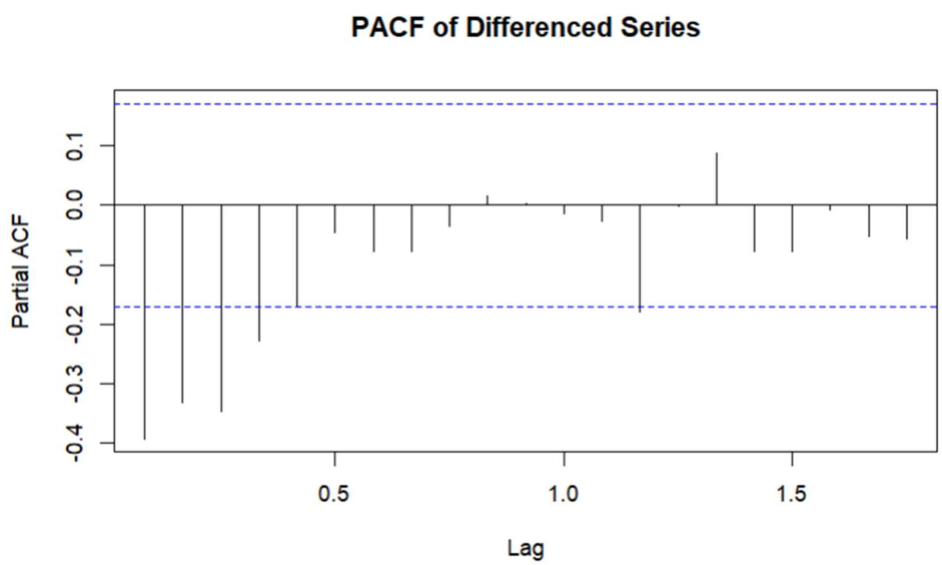
1. Overview of ACF and PACF

- **ACF (Autocorrelation Function):**
  The ACF measures the correlation between a time series and its lagged values. It helps identify the presence of **moving average (MA) components** in the data.
- **PACF (Partial Autocorrelation Function):**
  The PACF isolates the direct relationship between the time series and its lagged values, removing the influence of intermediate lags. It is used to identify **autoregressive (AR) components** in the series.

Both ACF and PACF are instrumental in diagnosing time series behavior and determining the p, d, and q parameters for ARIMA models:

- p: The number of AR terms (identified from PACF).
- d: The order of differencing needed to achieve stationarity.
- q: The number of MA terms (identified from ACF).





Model Selection: ARIMA and SARIMA

1. ARIMA Parameters:

The AutoRegressive Integrated Moving Average (ARIMA) model was used for initial modeling due to its simplicity and capability to handle non-stationary data. The following details summarize its performance:

1. **Parameter Selection:** A grid search was conducted to identify the best combination of parameters (p, d, q), where:
   - **p:** Order of the autoregressive terms.
   - **d:** Degree of differencing needed to make the series stationary.
   - **q:** Order of the moving average terms.

   The best parameters identified were **p=0, d=2, q=3**, which effectively modeled the data.

2. **Performance Metrics:**
   - **Akaike Information Criterion (AIC):** 2024.63
   - **Bayesian Information Criterion (BIC):** 2036.1
   - **Root Mean Square Error (RMSE):** 531.49
   - **Mean Absolute Percentage Error (MAPE):** 8.63%

   These metrics indicate the model's ability to fit the data while penalizing for complexity (AIC and BIC).

3. **Residual Diagnostics:** The **Ljung-Box Test** was applied to the residuals to check for autocorrelation. With a **p-value of 0.721**, the test confirmed no significant autocorrelation, suggesting that the residuals are uncorrelated, meeting the assumptions of the ARIMA model.
4. **SARIMA Model:**
   - Parameters: (2,1,2)(0,1,1)[12].
   - Performance Metrics:
     - AIC: 1883.11, BIC: 1899.78, RMSE: 522.18, MAPE: 8.02%.
   - SARIMA outperformed ARIMA due to its ability to model seasonality effectively.

2. SARIMA Model:

Seasonal ARIMA (SARIMA) was employed to account for periodic seasonal patterns in the data, which ARIMA cannot handle effectively.

- **Parameter Selection:** The SARIMA model was configured with the parameters **(2,1,2)(0,1,1)[12]**, where:
  - (p, d, q): Non-seasonal ARIMA components.
  - (P, D, Q): Seasonal ARIMA components.
  - [12]: Seasonal frequency, representing 12 months in a year.
- **Performance Metrics:**
  - **Akaike Information Criterion (AIC):** 1883.11
  - **Bayesian Information Criterion (BIC):** 1899.78
  - **Root Mean Square Error (RMSE):** 522.18
  - **Mean Absolute Percentage Error (MAPE):** 8.02%

SARIMA outperformed ARIMA in all metrics, particularly in AIC and BIC, demonstrating its superior ability to model seasonality.

- **Conclusion:** SARIMA was selected as the final model due to its better fit and forecasting accuracy. Its inclusion of seasonal components effectively captured periodic variations in energy production, which ARIMA could not achieve.

Forecasting and Evaluation

SARIMA Forecast vs. Actual:

- **Visualization:**
  - The **training data** was represented by a black line, providing the historical trend.
  - The **test data** was represented by a red line, reflecting actual values for the evaluation period.
  - The **forecasted values** were depicted with a blue-shaded region, indicating the confidence intervals (80% and 95%).
- **Observations:**
  - The SARIMA model closely aligned with the actual test data, demonstrating its reliability.

- The blue-shaded confidence intervals effectively captured the variability in energy production, showcasing the model's robustness in handling uncertainty.

Cross-Validation Results:

To ensure the SARIMA model's robustness and generalizability, rolling-origin cross-validation was performed:

- **Setup:**
  - **Training Window:** 96 months (8 years of data).
  - **Forecast Horizon:** 12 months.
  - **Step Size:** 12 months (1 year).
- **Key Metrics Across Folds:**
  - **Root Mean Square Error (RMSE):** 572.21
  - **Mean Absolute Error (MAE):** 463.42
  - **Mean Absolute Percentage Error (MAPE):** 9.49%
- **Conclusion:** The SARIMA model exhibited consistent performance across folds, with low error metrics, confirming its robustness and reliability for forecasting energy production.

Key Takeaways

1. Data Quality and Insights:

- The dataset was clean, with no missing values or significant anomalies, which facilitated smooth preprocessing and modeling.
- Seasonal variations were observed, with periodic peaks and troughs in energy production. However, the **seasonality strength** was minimal, as calculated by the seasonality strength formula.

2. Stationarity and Transformation:

- The data was found to be non-stationary based on the Augmented Dickey-Fuller (ADF) test, with a **p-value of 0.045**.
- Stationarity was achieved through:
  - **Log Transformation:** To stabilize variance.

- **Differencing:** To eliminate trends and achieve a stationary series.

- Rolling statistics confirmed that the transformation was successful in stabilizing the data.

3. Model Selection and Performance:

- ARIMA and SARIMA were both evaluated:
  - ARIMA showed reasonable performance but lacked the ability to handle seasonality.
  - SARIMA outperformed ARIMA, achieving better AIC, BIC, and RMSE values while effectively modeling seasonal patterns.
- The SARIMA model proved to be more accurate and reliable for energy production forecasting.

4. Forecasting Evaluation:

- SARIMA's forecast aligned closely with actual test data, validating its predictive accuracy.
- Cross-validation results confirmed the model's robustness, with consistent error metrics across different training and testing periods.

Why SARIMA Outperformed ARIMA Despite Minimal Seasonality

1. **Seasonality Detection vs. Flexibility in Modeling:**
   - While the seasonality strength calculated from the data was minimal, SARIMA includes both seasonal and non-seasonal components in its framework.
   - The seasonal component, though subtle, could still capture periodic fluctuations in the data that ARIMA failed to model effectively. Even if the detected seasonality strength is low, small but consistent seasonal patterns can still affect forecast accuracy.
2. **Data Characteristics Beyond Seasonality:**
   - The energy production data may have had weak but persistent seasonal trends or other periodic variations that were not immediately apparent from the calculated seasonality strength.

- SARIMA's seasonal differencing and seasonal AR/MA components allowed it to adapt to these subtle patterns, which ARIMA could not handle since it lacks seasonal differencing.

3. **Improved Model Complexity:**
    - SARIMA's additional seasonal parameters allow it to fit more flexible patterns in the data. For example:
        - The seasonal differencing term accounts for periodic variations, even if those variations are minor.
        - The inclusion of seasonal autoregressive (SAR) and moving average (SMA) terms can enhance the model's ability to capture repeated patterns over fixed intervals (e.g., months).
4. **Noise Reduction and Stability:**
    - The SARIMA model's inclusion of seasonal components could also act as a regularizer, improving the model's stability and ability to generalize across the dataset.
    - By accounting for possible periodic signals, SARIMA may have reduced overfitting or underfitting compared to ARIMA.
5. **Performance Metrics:**
    - SARIMA showed better AIC, BIC, RMSE, and MAPE values compared to ARIMA:
        - **AIC/BIC Improvements:** SARIMA's ability to capture subtle seasonal patterns resulted in a lower penalized likelihood.
        - **RMSE/MAPE Improvements:** These indicate that SARIMA had better predictive performance and aligned more closely with actual observations during evaluation.
6. **Interpretation of Results:**
    - Even if seasonality is not visually obvious or has a low calculated strength, SARIMA's additional parameters give it more flexibility to adapt to the data's inherent structure, including weak seasonality or other cyclical trends. This flexibility likely contributed to SARIMA's better performance in this project.

Conclusion

This project successfully demonstrated the application of time series forecasting techniques to predict renewable energy production trends, focusing on the key variable **Energy_Production_MWh**. The use of ARIMA and SARIMA models highlighted the value of statistical modeling in addressing variability and enabling better decision-making for energy grid management and resource allocation.

Key Outcomes:

1. **Data Insights and Preparation:**
    - The dataset was well-structured, with no missing values or significant anomalies, facilitating smooth preprocessing and modeling.
    - Minimal seasonality strength (2.32e-09) was identified, yet preprocessing steps like log transformation, differencing, and rolling statistics ensured stationarity, a critical requirement for effective time series modeling.
2. **Model Selection and Performance:**
    - ARIMA provided a reasonable baseline for forecasting but lacked the ability to handle subtle seasonal trends in the data.
    - SARIMA outperformed ARIMA in all performance metrics:
        - Lower AIC and BIC values reflected its ability to balance model complexity and goodness of fit.
        - Improved RMSE and MAPE indicated higher predictive accuracy.
    - Despite the minimal seasonality, SARIMA’s additional seasonal parameters captured weak seasonal trends, enhancing forecast reliability.
3. **Forecasting and Evaluation:**
    - SARIMA forecasts closely aligned with actual test data, as evidenced by the visual overlap between predicted and observed values.
    - Confidence intervals effectively captured variability, demonstrating the model’s robustness under uncertainty.
    - Rolling-origin cross-validation validated SARIMA's reliability, with consistent error metrics across folds (RMSE: 572.21, MAPE: 9.49%).
4. **Practical Implications:**

- The model's accuracy supports its practical application in renewable energy forecasting, enabling better planning for:
    - **Grid Stability:** Anticipating production variability to balance supply and demand.
    - **Resource Optimization:** Allocating energy storage and grid resources efficiently.
    - **Sustainability:** Supporting renewable energy adoption by mitigating the impact of low-production periods.

Lessons Learned:

The project emphasized the importance of thorough data analysis and preprocessing in time series forecasting. It also highlighted how advanced models like SARIMA can outperform simpler models (ARIMA) by addressing even subtle seasonal trends in the data.

Future Work:

1. **Model Refinement:**
    - Explore other models like ETS (Exponential Smoothing State Space) or hybrid approaches combining statistical models with machine learning for improved performance.
2. **Multivariate Analysis:**
    - Extend the study to include exogenous variables (e.g., weather conditions, temperature) to enhance the predictive power of the model.
3. **Integration into Systems:**
    - Deploy the SARIMA model as part of a real-time energy forecasting tool for energy providers.

Final Thoughts:

This project underscores the critical role of time series forecasting in addressing challenges in renewable energy production. By focusing on a single variable and leveraging statistical methods, it demonstrates how data-driven insights can support sustainability goals, optimize resource use, and contribute to a greener energy future.