Report: Time Series Analysis and Forecasting of Electric Production Data

Introduction

This report presents a comprehensive analysis of the electric production dataset, focusing on time series analysis and forecasting. The dataset, sourced from the file "Electric_Production.csv," contains monthly electric production values from January 1985 to December 2017. The primary objective is to model the data, identify trends and seasonality, and forecast future production values using various statistical methods.

Data Preparation and Initial Analysis

The dataset was loaded into a Pandas DataFrame and preprocessed to ensure proper date indexing. The initial data inspection revealed a clear seasonal pattern and a general trend over time. The first few rows of the dataset are as follows:

Table

DATE	IPG2211A2N
1985-01-01	72.5052
1985-02-01	70.6720
1985-03-01	62.4502
1985-04-01	57.4714
1985-05-01	55.3151

A line plot of the data showed a clear seasonal pattern with some variability over time. The Mann-Kendall trend test indicated a significant increasing trend in the data (p-value < 0.05).

Seasonal Decomposition

Seasonal decomposition was performed using both multiplicative and additive models. The results showed distinct seasonal patterns and trends. The multiplicative model highlighted the seasonal fluctuations, while the additive model provided a clearer view of the underlying trend.

Exponential Smoothing

An Exponential Smoothing model with trend and seasonal components was fitted to the data. The model was trained on 70% of the dataset, and the remaining 30% was used for testing. The forecasted values were compared to the actual data, resulting in a Mean Absolute Percentage Error (MAPE) of 0.0945 for the test set.

Stationarity Tests

The Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests were conducted to assess the stationarity of the data. The ADF test suggested that the data was non-stationary (p-value = 0.1862), while the KPSS test indicated that the data was trend stationary (p-value = 0.01). After differencing the data, both tests confirmed stationarity.

SARIMAX Modeling

Two SARIMAX models were fitted to the data:

1. **Model 1**: SARIMAX(1,1,1)x(1,1,1,12)
2. **Model 2**: SARIMAX(1,1,2)x(1,0,1,12)

Model 1 achieved a MAPE of 0.0930 for the test set, while Model 2 achieved a MAPE of 0.0862. The diagnostics plots for both models showed that the residuals were normally distributed and exhibited no significant autocorrelation.

Forecasting

The models were used to forecast future electric production values. The forecasts extended 90 months beyond the test set. The forecasted values for the next few months are as follows:

Table

DATE	Forecasted Value (Model 1)	Forecasted Value (Model 2)
2018-01-01	131.233972	129.028578
2018-02-01	126.187633	124.299441
2018-03-01	119.036998	117.562077
2018-04-01	109.462493	108.577938

DATE	Forecasted Value (Model 1)	Forecasted Value (Model 2)
2018-05-01	110.648377	109.695474

Conclusion

The analysis and forecasting of the electric production dataset were successfully conducted using Exponential Smoothing and SARIMAX models. The SARIMAX model with parameters (1,1,2)x(1,0,1,12) provided the most accurate forecasts, achieving a MAPE of 0.0862. The models were able to capture the underlying trends and seasonal patterns in the data, providing reliable forecasts for future electric production values.