D Square Consulting Pvt Ltd

**Name** : Badhrinarayan V

**Project** : Time Series Forecasting Project

**Position** : Data Science Intern

Time Series Forecasting Project – Internship at Dsquare Consulting

# 1. Introduction

As part of my internship at Dsquare Consulting, I undertook a time series forecasting project using the Electric Production dataset sourced from Kaggle. The objective was to explore various forecasting models and identify the most accurate one for predicting future electricity output. This project allowed me to apply and deepen my understanding of forecasting techniques like SARIMAX, Holt’s Winter method, and ensemble modeling.

# 2. Data Understanding and EDA

The dataset consisted of monthly electricity production figures. I started by converting the date column to datetime format and conducted exploratory data analysis (EDA) to visualize trends and seasonality. Time series decomposition was performed to break the data into trend, seasonal, and residual components.

# 3. Models Applied

## 3.1 SARIMAX Model

Stationarity was verified using the Augmented Dickey-Fuller (ADF) test, and lags were identified using ACF and PACF plots. A SARIMAX model was built by tuning the appropriate parameters. It captured both seasonality and trend effectively. SARIMAX gave the best performance on unseen data.

## 3.2 Holt-Winters Exponential Smoothing

Holt’s Winter model with additive trend and seasonality was applied to model the seasonal patterns. This method served as a simpler alternative to SARIMAX, offering reasonable accuracy.

## 3.3 Stacked Ensemble Model

To improve robustness, an ensemble model was built by combining SARIMAX and Holt’s forecasts using Linear Regression. While the ensemble provided consistent predictions, SARIMAX alone slightly outperformed it on test data.

# 4. Model Evaluation

Models were compared using both visual plots and metrics like MAE and RMSE. SARIMAX and the ensemble approach performed well, with SARIMAX being the most reliable in forecasting unknown values.

# 5. Key Learnings

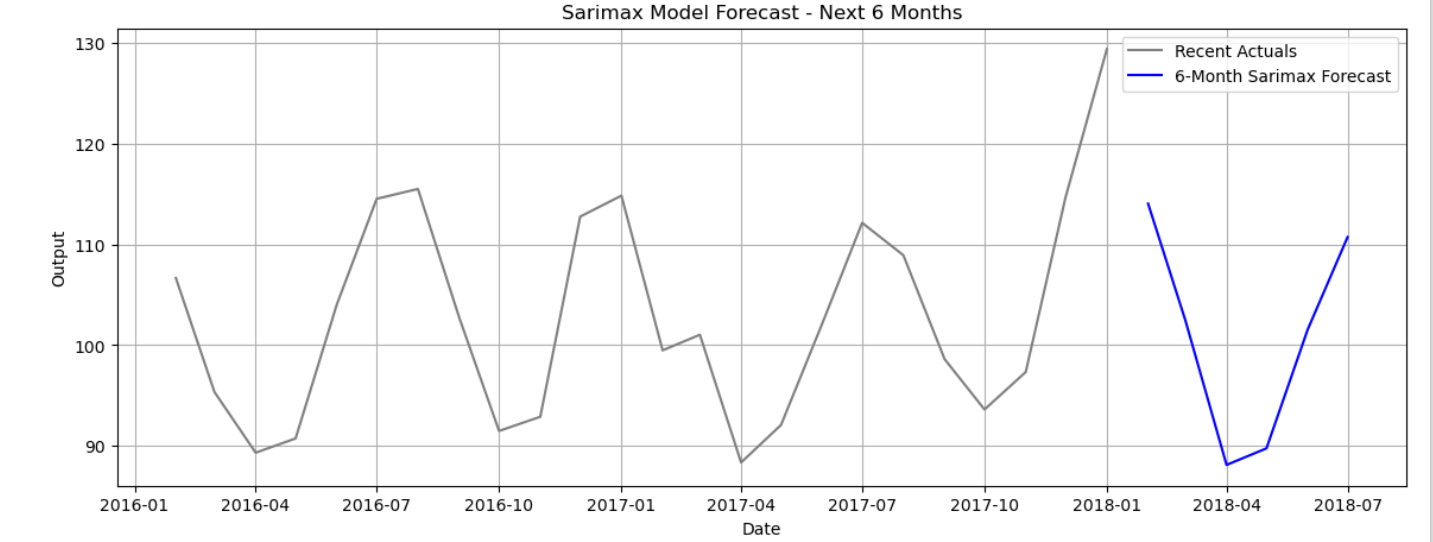
• Importance of stationarity, differencing, and lag identification in time series modeling.  
• Comparing models not just with metrics but also through visual insights.  
• Understanding how ensemble learning can combine strengths of individual models.  
• Real-world forecasting often requires model tuning and interpretability.

# 6. Conclusion

This internship project enhanced my understanding of time series forecasting in a practical setting. Working with SARIMAX, Holt’s model, and ensembles gave me a hands-on perspective on solving business problems using data science. I'm excited to build on this foundation in future analytics challenges.

# 7. Forecast Plots and Error Metrics

## 7.1 SARIMAX Model Forecast

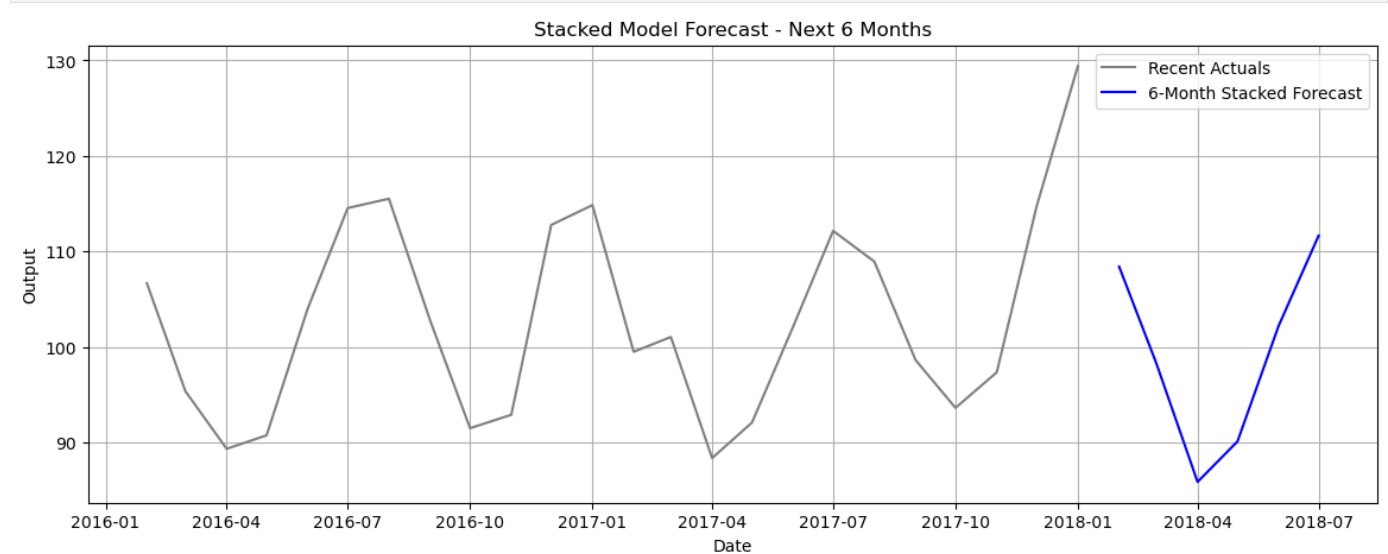


ADF p-value after first differencing: 4.07e-10

MAE: 4.10

RMSE: 5.83

## 7.2 Stacked Ensemble Forecast



MAE: 2.86

RMSE: 3.78

## 7.3 Holt-Winters Forecast

MAE: 3.27

RMSE: 4.23