Time Series Forecasting Report: An Ensemble Approach

Introduction:

The primary objective of this study was to predict future trends for various sectors based on a historical data span of 500 days. Our methodology utilized two significant models: SARIMA and XGBoost, coupled with the inclusion of weather-based features to bolster the predictive capability.

Data:

The dataset encompassed six diverse sectors. Each dataset underwent rigorous preprocessing to ensure its readiness for modeling. Weather-derived features were incorporated, given their critical influence on time series forecasting, particularly for sectors sensitive to environmental fluctuations.

Modeling:

SARIMA:

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model, adept at managing time series data with inherent trends, seasonality, and autocorrelation, was our first choice. We undertook an exhaustive search for optimal hyperparameters using the Akaike Information Criterion (AIC) as the guiding metric.

XGBoost:

Subsequently, we engaged the XGBoost algorithm, a gradient-boosted model celebrated for its efficacy and speed. This model was not limited to the time series data but also made use of the previously introduced weather-centric features. A systematic hyperparameter tuning was executed using GridSearchCV to ensure the model's peak performance.

Time Series Forecasting Report: An Ensemble Approach

Ensemble Methodology:

To capitalize on the strengths of both models, a weighted ensemble technique was employed. Predictions from both SARIMA and XGBoost were amalgamated, with the weightage determined by the error metrics of individual models across datasets. Specifically, the combination formula was:

Ensemble Predictions = 0.65 x SARIMA Predictions + 0.35 x XGBoost Predictions

The above ratio was fine-tuned for each dataset to optimize the ensemble's output. This approach facilitated the capture of intricate patterns, enhancing forecast precision.

Performance Metrics and Evaluation:

The models' predictive accuracy was scrutinized using the following metrics:

- Mean Absolute Percentage Error (MAPE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)

Each sector dataset was evaluated individually. Subsequently, a weighted average of the errors was computed, leveraging the mean of the actual test data values, ensuring a holistic assessment of model performance.

Concluding Remarks:

While our ensemble strategy, combining SARIMA and XGBoost, delivered robust projections for the datasets in question, the ever-evolving domain of modeling offers further avenues for exploration.

With an abundance of data and computational might, cutting-edge models like Long Short-Term Memory (LSTM) networks, Transformer architectures, or Facebook's Prophet could be harnessed for potentially superior forecasting accuracy.

In summation, this study's ensemble methodology, enriched with weather-based feature engineering, showcases the synergy of SARIMA and XGBoost. The realm of data science and forecasting perpetually beckons with opportunities for refinement and innovation. Given adequate resources and data, future endeavors might pivot to more advanced models and feature engineering methodologies.