Advanced Time Series Forecasting with Prophet and Bayesian Optimization

Student: Pon Enitha

1. Dataset Description

This project uses real-world financial market data from the S&P; 500 Index (^GSPC), obtained using the yfinance API. Ten years of daily data were downloaded, and the most recent eight years were retained after cleaning. Missing dates were corrected to daily frequency, and missing prices were interpolated. Stationarity was checked using the Augmented Dickey–Fuller test, which showed non■stationarity (ADF Statistic ≈ 0.3983, p-value ≈ 0.9814), which is expected for stock prices.

2. Preprocessing & Forecasting Setup

The dataset was split into training data and a 60■day hold■out test set. Performance was evaluated on short-term (7■day) and medium-term (30■day) horizons using RMSE, MAE, and MAPE. These metrics were computed for both baseline and optimized Prophet models.

3. Baseline Prophet Model

A default Prophet model was fitted. Performance:

7■Day Horizon – RMSE=339.60, MAE=339.54, MAPE=8.66%

30■Day Horizon – RMSE=345.78, MAE=345.53, MAPE=8.81%

4. Bayesian Optimization Using Optuna

Bayesian Optimization (Optuna) tuned key Prophet hyperparameters:

• seasonality_prior_scale

• changepoint_prior_scale

• n_changepoints

• seasonality_mode

The training data was internally split (85%/15%) for validation. Optuna minimized validation RMSE and found the best configuration (CV RMSE ≈ 293.38) that significantly outperformed the baseline.

5. Final Optimized Prophet Model

Retrained using the best hyperparameters, the optimized model achieved:

7■Day Horizon – RMSE=282.74, MAE=282.55, MAPE=7.21%

30■Day Horizon – RMSE=284.61, MAE=284.14, MAPE=7.24%

This represents roughly a 20% improvement in accuracy.

6. Residual Diagnostics

Residual Time Plot:

Residuals oscillate around zero, indicating no bias or drift and confirming the model's stability.

ACF Insights:

Strong autocorrelation at lags 1–5 (typical for financial markets), then gradual decay with no remaining seasonal structure.

PACF Insights:

Only lag■1 is significant; all later lags fall within confidence limits. This indicates no major unmodeled structure.

7. Comparative Analysis

Across both 7■day and 30■day horizons, the optimized model consistently outperforms the baseline Prophet model in RMSE, MAE, and MAPE. This demonstrates the benefit of Bayesian Optimization in improving Prophet's predictive accuracy.

8. Final Model Justification

The optimized Prophet model is selected because:

• It provides significantly lower forecasting error.

• It models trend and seasonality more effectively.

• Residuals show no systematic patterns.

• Bayesian Optimization ensures reproducible and principled tuning.

• The model remains interpretable and stable.

9. Conclusion

This project covers data acquisition, cleaning, baseline forecasting, Bayesian Optimization, evaluation across horizons, residual analysis, and final model justification. All required deliverables—dataset, predictions, metrics, plots, and analysis—are successfully produced. The optimized Prophet model is the final recommended solution due to its superior accuracy and well■behaved residuals.