

Ensemble Machine Learning Methods for Stock Price Forecasting: A Comparative Out-of-Sample Analysis

Samir K. Safi

Department of Statistics and Business Analytics, United Arab Emirates University, Al Ain, United Arab Emirates

Abstract:

This study compares the predictive performance of ensemble methods, specifically, decision trees, random forests, and gradient boosting, focusing on their tuning and performance in predicting stock prices using a 500-company dataset from Yahoo Finance. The analysis focuses on understanding stock price fluctuations, trading volume variability, and the distribution of dividends and stock splits. The ensemble methods were evaluated using root mean square error, mean absolute percentage error, and the coefficient of determination to comprehensively assess prediction accuracy and error reduction. The results revealed that decision trees outperformed random forests and gradient boosting in accuracy and error rate after hyperparameter tuning. Random forests demonstrated high performance when tuned, effectively capturing intricate and nonlinear interactions within the dataset. However, gradient boosting exhibited mixed results. This study provides insights into the strengths of each model and highlights the advantages of ensemble methods in enhancing stock price prediction accuracy. The findings provide a foundation for further enhancements in financial forecasting models.

Keywords:

Decision Trees, Random Forests, Gradient Boosting, Machine learning, Financial Forecasting Models.