**Financial Forecasting Using Ensemble Methods: A Comparative Study**

ERTUGRUL TUNC

A Thesis Submitted in Partial Fulfilment

of the requirements for the

Degree of

Master of Science in Data Analytics

Diagram

Description automatically generated with medium confidence

MAY 2025

Supervisor: Maqsood Hussain Shah

**Financial Forecasting Using Ensemble Methods: A Comparative Study**

Table of Contents

[Chapter 1: Introduction 4](#_Toc198260098)

[**1.1 Background and Context** 4](#_Toc198260099)

[**1.2 Problem Statement** 4](#_Toc198260100)

[**1.3 Research Aim and Objectives** 5](#_Toc198260101)

[**1.4 Research Questions and Hypotheses** 5](#_Toc198260102)

[**1.5 Justification for the Study** 6](#_Toc198260103)

[**1.6 Scope and Delimitations** 6](#_Toc198260104)

[**1.7 Dissertation Structure** 7](#_Toc198260105)

[Chapter 2: Literature Review 8](#_Toc198260106)

[**2.1 Introduction to Literature Review** 8](#_Toc198260107)

[**2.2 Traditional Financial Forecasting Methods** 8](#_Toc198260108)

[**2.3 Machine Learning in Financial Forecasting** 9](#_Toc198260109)

[**2.4 Ensemble Learning Fundamentals** 9](#_Toc198260110)

[**2.5 Bagging Techniques in Finance** 10](#_Toc198260111)

[**2.6 Boosting Techniques in Finance** 11](#_Toc198260112)

[**2.7 Stacking and Hybrid Ensemble Approaches** 11](#_Toc198260113)

[**2.8 Challenges and Limitations of Ensemble Models** 12](#_Toc198260114)

[**2.9 Summary of Key Insights and Research Gaps** 12](#_Toc198260115)

[Chapter 3: Research Methodology 14](#_Toc198260116)

[**3.1 Introduction** 14](#_Toc198260117)

[**3.2 Research Design** 14](#_Toc198260118)

[**3.3 Data Collection** 15](#_Toc198260119)

[**3.3.1 Source of Data** 15](#_Toc198260120)

[**3.3.2 Variables** 15](#_Toc198260121)

[**3.4 Data Preprocessing** 16](#_Toc198260122)

[**3.4.1 Handling Missing Values** 16](#_Toc198260123)

[**3.4.2 Feature Selection** 16](#_Toc198260124)

[**3.4.3 Data Scaling** 17](#_Toc198260125)

[**3.4.4 Data Split** 17](#_Toc198260126)

[**3.5 Mx odel Development** 17](#_Toc198260127)

[**3.6 Evaluation Metrics** 18](#_Toc198260128)

[**3.7 Tools and Frameworks Used** 20](#_Toc198260129)

[**3.8 Validation Strategy** 20](#_Toc198260130)

[**3.9 Ethical Considerations** 21](#_Toc198260131)

[**3.10 Limitations of Methodology** 21](#_Toc198260132)

[**3.11 Summary** 22](#_Toc198260133)

[Chapter 4: Results and Analysis 23](#_Toc198260134)

[**4.1 Introduction** 23](#_Toc198260135)

[**4.2 Overview of Dataset Behavior** 23](#_Toc198260136)

[**4.3 Model Predictions vs. Actual Prices** 25](#_Toc198260137)

[**4.4 Quantitative Performance Metrics** 26](#_Toc198260138)

[**4.5 Boosting Performance with XGBoost** 28](#_Toc198260139)

[**4.6 Residual Analysis** 29](#_Toc198260140)

[**4.7 Feature Importance Analysis** 31](#_Toc198260141)

[**4.8 Cross-Validation and Model Robustness** 33](#_Toc198260142)

[**4.9 Training Time Comparison** 34](#_Toc198260143)

[**4.10 Research Hypotheses Analysis** 35](#_Toc198260144)

[**4.11 Discussion in Context of Literature** 36](#_Toc198260145)

[**4.12 Limitations of Results** 37](#_Toc198260146)

[**4.13 Summary** 37](#_Toc198260147)

[Chapter 5: Discussion 39](#_Toc198260148)

[**5.1 Introduction** 39](#_Toc198260149)

[**5.2 Summary of Key Findings** 39](#_Toc198260150)

[**5.3 Implications of Ensemble Learning in Financial Forecasting** 40](#_Toc198260151)

[**5.4 Theoretical Contributions** 41](#_Toc198260152)

[**5.5 Practical Applications and Industry Relevance** 42](#_Toc198260153)

[**5.6 Methodological Reflections** 43](#_Toc198260154)

[**5.7 Comparison with Existing Literature** 44](#_Toc198260155)

[**5.8 Limitations and Directions for Future Research** 44](#_Toc198260156)

[**5.9 Summary** 45](#_Toc198260157)

[Chapter 6: Conclusion 47](#_Toc198260158)

[**6.1 Overview** 47](#_Toc198260159)

[**6.2 Addressing Research Objectives** 47](#_Toc198260160)

[**6.3 Key Findings** 48](#_Toc198260161)

[**6.4 Theoretical and Practical Contributions** 48](#_Toc198260162)

[**6.5 Limitations** 49](#_Toc198260163)

[**6.6 Recommendations for Future Research** 49](#_Toc198260164)

[**6.7 Final Reflections** 50](#_Toc198260165)

[References 51](#_Toc198260166)

# **Chapter 1: Introduction**

## **1.1 Background and Context**

In the growing, financially diverse environment, forecasting is crucial for strategy implementation, distribution of resources, and risk analysis. Business financial forecasting refers to the process of predicting future trends in an organization to minimize risks and losses and maximize profits among investors, institutions as well as policy-makers. In the past, forecasting approaches like statistical models (e.g., ARIMA, GARCH) have been used; nonetheless, the growth of complexity in financial data renders the traditional models inefficient (Padhi et al., 2021).

Today, artificial intelligence (AI) and machine learning (ML) are gaining increased attention as key drivers of disruptive changes in financial forecasting. By contrast, the machine learning algorithms can learn from new and complex patterns, large datasets, and provide better predictions. Of all these categories, two of the most promising subcategories consist of bagging, boosting, and stacking. Since ensemble methods minimize variance, fix biases and improve generalization, they are suitable to be applied in financial data that is marked by high volatility and uncertainty.

Moreover, the development of deep learning models and the application of ensemble frameworks have provided new directions for the new hybrid systems with feature extraction and the ensemble characteristics (Li & Tam, 2024). These are groundbreaking tools that redefine the concept of financial forecasting and take it from its traditional, stationary, set-based format as far as in a more flexible, holistic learning approach capable of addressing challenges in today’s fluid financial environment.

## **1.2 Problem Statement**

However, with the use of machine learning, there are difficulties which arise in financial forecasting. These financial markets have elements such as macroeconomic parameters, politics, global mood swings and the like affecting the behavior of the market; hence, the levels of nonlinearity and noise are very high (Lyu et al., 2023). These complex patterns are hard to detect using traditional single-model approaches and thus provide unstable and imprecise forecasts (Lee et al., 2020).

The fluctuations in market data also pose problems such as overfitting, bias or high variance in even standalone models like decision trees, neural networks, etc. Thus, there is a lot of emphasis on developing more robust and elaborate forecasting structures that can adequately capture the complex and unpredictable characteristics of financial series data. If we combine multiple models, then ensemble learning seems to be a promising solution to enhance the efficiency of the model as against individual models (Dalal et al., 2022).

## **1.3 Research Aim and Objectives**

**Research Aim**:

This research aims to evaluate the effectiveness of ensemble learning techniques—bagging, boosting, and stacking—in improving the accuracy and reliability of financial forecasting models.

**Research Objectives**:

* To explore the theoretical underpinnings and practical implementations of ensemble methods in financial forecasting.
* To compare the performance of bagging, boosting, and stacking models with traditional and single machine learning approaches.
* To assess the applicability of ensemble models in predicting stock market indices and macroeconomic trends.
* To identify key challenges, limitations, and opportunities in implementing ensemble forecasting models in financial markets.

This structured approach enables the research to contribute both theoretically and practically to the field of financial machine learning.

## **1.4 Research Questions and Hypotheses**

**Research Questions**:

1. How do ensemble learning methods perform compared to traditional financial forecasting models?
2. Which ensemble technique—bagging, boosting, or stacking—yields the highest prediction accuracy for financial time series?
3. What are the limitations and risks associated with ensemble forecasting in volatile market environments?

**Hypotheses**:

* **H1**: Ensemble methods outperform single machine learning models in forecasting financial markets.
* **H2**: Boosting techniques (e.g., XGBoost) deliver superior predictive performance over bagging and stacking methods in financial forecasting.
* **H3**: A hybrid stacked ensemble model combining deep learning and machine learning algorithms achieves the highest level of predictive accuracy.

## **1.5 Justification for the Study**

The rationale for undertaking this study comes from the realization of the importance of forecasting models in a world that is ever evolving and growing financially. While traditional econometric models became a powerful tool in many application fields, they are not sufficient to model the nonlinearities, step-likely shifts and high-volatility peculiar to the financial data (Deep, 2024). While DM techniques in general have brought significant enhancement in the forecast accuracy, individual models suffer deficiencies and inherent flaws like high variance or bias (Li & Pan, 2021).

These limitations are handled by ensemble learning by availing the probability prediction of more than one base learner; it makes the whole model more robust and accurate (Suprihadi et al., 2025). The abovementioned methods include: Bagging, which helps decrease the variance, boosting, which aims at reducing bias by concentrating on difficult instances, and stacking, which integrates diverse models to capture complex relations in the data (pp. 424 Hao & Zhang, 2024).

This study is pertinent as ensemble methods have been recently used in portfolio selection (Li & Tam, 2024), credit card fraud identification (Dalal et al., 2022), and economic trend prediction (Zhao et al., 2023). Thus, advancing knowledge in this area, the study helps to fill the gaps in ensemble models analysis in the context of financial forecasting and offers valuable practical recommendations for financial analysts, investors, and policy-makers to improve the financial risk prediction.

## **1.6 Scope and Delimitations**

In this research, bagging, boosting and stacking methods of ensemble learning for financial forecasting have been explored. This is done by focusing on stock market indices such as the S&P 500, NASDAQ and other important macroeconomic variables, including GDP growth rates and inflation. The sources of data will also be the data collected from the public domain database about the past financial records of companies. In cases where it will be suitable, deep learning will be incorporated into ensemble models. This research work will not deal with other forms of assets like cryptocurrencies or commodity markets, except where they will be referred to in like scripts. Also, real-time market handling and other actual trading activities are excluded from the material of the work.

## **1.7 Dissertation Structure**

This dissertation is structured into six chapters:

* **Chapter 1** introduces the background, problem statement, research aim, questions, and structure of the study.
* **Chapter 2** presents a detailed literature review, critically evaluating traditional forecasting methods, machine learning models, and ensemble learning techniques in finance.
* **Chapter 3** outlines the research methodology, including data sources, model selection, and evaluation metrics.
* **Chapter 4** discusses the experimental design and implementation of ensemble models for financial forecasting.
* **Chapter 5** presents the results and performance comparison of different forecasting techniques.
* **Chapter 6** concludes the study by summarizing findings, highlighting contributions, discussing limitations, and proposing directions for future research.

# **Chapter 2: Literature Review**

## **2.1 Introduction to Literature Review**

Budgeting has more to do with statistical and econometric forecasting. However, due to the rising complexity and unpredictability of the latter financial environments that require assessment, more refined methods are required to embrace nonlinear associations and tendency patterns. In the current literature, various forecasting methodologies have been documented ranging from the conventional statistical models to the more complex machine learning models such as the bagging, boosting, and stacking approaches.

First of all, the traditional approaches for forecasting are discussed in order thus identify the advantages and drawbacks of each method. This is succeeded by highlighting the development of machine learning (ML) techniques in the area of forecasting. Then, the topic of ensemble learning is introduced, which is further explained with the subtopics of Bagging, Boosting and Stacking. The chapter discusses how these techniques have been applied in finance and some of the benefits and drawbacks of such application before highlighting the research gaps that this study seeks to fill.

## **2.2 Traditional Financial Forecasting Methods**

Traditional methods, such as statistical forecasting techniques such as ARIMA and GARCH, have been used due to their simplicity and understandable nature. These models depend mostly on the past patterns intending to transition to future ones. For example, an ARIMA time series forecast is obtained by combining autoregression, differencing, moving average for handling non-stationary data. The GARCH models, on the other hand, are good for modelling and forecasting the financial market volatility, given that they account for the changing variance over time (Meira et al., 2022).

However, they are not free from problems as explained below. They make the linear and stationary assumption on the data, which makes it difficult to model the non-linear behavior present in numerous financial time series (Deep, 2024). In addition, traditional models are largely rigid, not capable of handling drastic changes in the market polarity due to external variables such as political factors, diseases and investors’ attitudes.

The other shortcoming is that they employ limited account of external predictors, which might bring more capacity to improve the fore casting results (Zhao et al., 2023). This remains especially true as more intricate and multifaceted financial data sets have been available to be analyzed, where traditional statistical modelling methods prove to be less reliable in providing accuracy in the forecasting. Due to such limitations, researchers and practitioners have sought to find suitable machine learning structures in the present literature.

## **2.3 Machine Learning in Financial Forecasting**

Machine learning in financial forecasting has brought changes by making ability to capture complex non-linear relationships without relying on certain assumptions. ANNs and RNNs, together with LSTM networks, have been used in forecasting tasks because of their ability to capture complex temporal dependencies (Rubai: 2023).

Many studies have demonstrated the benefits of ANNs in terms of handling non-linear interactions between the financial variables, while both RNN and LSTMs have been quite effective for modelling sequential data where issues of long dependency exist, as confirmed by Li & Pan, (2021). The distinguished advantage of these models is that they can process more data compared to conventional approaches, find patterns and adjust to the new data trends.

Nonetheless, machine learning models are not without merits and there are has following drawbacks. They usually suggest that large datasets are useful for training, in which data may be scarce in these specific financial markets. Furthermore, overfitting is still possible and is worse when models are complicated and contain parameters that are larger than the size of the training data (Suprihadi et al., 2025). Some of the challenges include model interpretability; many of the models developed implement various techniques where knowledge of how the models arrive at their results is not easily explained to the stakeholders by Olorunnimbe & Viktor (2024).

Nonetheless, based on the results attained in the study, machine learning studies have outcompeted the classical statistical models in various applications of financial forecasting paving way for the ensemble methods for improved accuracy and model reliability.

## **2.4 Ensemble Learning Fundamentals**

Ensemble learning is a learning paradigm of Machine Learning which works to enhance the performance of prediction by using multiple models. The rationale used to support the use of ensemble learning is that a group of poor learners can be marginally powerful when combined in the correct way (Miao & Polak, 2023).

There are three basic forms of ensembling, which are bagging, boosting and stacking.

* Bagging (Bootstrap Aggregating) it is a technique where many models are trained on randomly selected samples from the training data and the results are combined through averaging for regression and voting for classification.
* A boosting trains models in a cascade where each model tries to fix the mistakes made by the former while, hence targeting the hardest-to-classify instances.
* Stacking deals with multiple base learners and trains another learner to learn how the other learners should be combined, which makes the generalization improved.

The use of ensembles helps in problems like overfitting, high variance and bias, and this model is highly fitting for financial forecasting since data tends to be noisy and non-stationary (Li & Tam, 2024). The following sections delve deeper into the application of each ensemble technique in financial forecasting.

## **2.5 Bagging Techniques in Finance**

Bagging techniques, specifically Random Forests, have become popular in financial prediction as they help in bringing down the model variance and increasing its stability. In Random Forests, several decision trees are learnt on resamples of the training set and the ensemble of their outputs is used to predict the final result (Lee et al., 2020).

Indeed, in their study, Meira et al. (2022) also revealed the applicability of bagging-based models in natural gas consumption forecast, showing the lesser sensitivity to noisy data. For instance, Lyu et al. (2023) checked the validity of employing bagging to predict the oil futures volatility, and they also noted that compared to single models, ensemble achieves higher accuracy as well as stability in the performance.

In financial markets, Random Forests have been applied in forecasting stock prices, credit risk and credit ratings. This capability of these models is highly useful since they can incorporate nonlinear relations between the variables, which is an advantage over linear models.

Nevertheless, the discussed bagging techniques have some drawbacks as well. The former can be highly time and CPU consuming if many trees are employed (Bhambu et al., 2024). In addition, a number of base biases might not be addressed by both kinds, while the variance is successfully reduced. However, bagging has always been adopted as an effective method of ensemble forecasting for financial application.

## **2.6 Boosting Techniques in Finance**

Several boosting methods are used in order to increase the accuracy of the models and few of them are AdaBoost, Gradient Boosting Machines (GBM), and XGBoost. Ensemble techniques are built by absorbing a set of weak learners and each learner is trained to remedy the errors made by its forerunners (Dalal et al., 2022).

Specifically XGBoost has become one of the standard methods for prediction for structured data, and for instance, in forecasting financial indications. Yasper et al. (2023) also illustrated the use of XGBoost for rainfall estimation with a strong focus on the optimization of hyperparameters that are also relevant in the financial environment.

In financial forecasting, boosting methods have been used to predict the movements of stocks, to identify cases of fraud in financial activities, and to forecast macroeconomic indicators (Zhao et al., 2023). Additionally, to support the financial payment systems, an optimised XGBoost model was also remarkably advisable for identifying the levels of fraudulent transactions by Dalal et al. (2022).

In terms of flexibility, boosting models are fully applicable to data fields of different types. Nevertheless, they are also overfitting if not regularized also they are sensitive to noisy data (Miao & Polak, 2023). However, there are a number of disadvantages associated with boosting that needs to be borne in mind, boosting techniques are nevertheless highly valuable in terms of forecasting accuracy and are used widely in today’s financial predictiveness tools.

## **2.7 Stacking and Hybrid Ensemble Approaches**

Stacking is the process of training multiple base models and then coming up with a meta-learner to aggregate the predictions of the base models (Bravo, 2024). This is beneficial after installation of other techniques as it enables capture of different patterns within the data.

Hao and Zhang (2024) proposed a new stock price forecasting model based on the stacking ensemble of LSTM networks and tree-based models. The conclusion was that stacking had a greater accuracy as compared to the other single models or simple ensemble techniques.

Other models are also developed to improve the accuracy of the deep learning models, such as, hybrid models including ensemble models. In the work by Li and Tam (2024), an advanced Attention-based Ensemble framework for portfolio optimization was developed and suggested to be a better solution than the existing models.

Stacking and the hybrid approach are more flexible and equipe in complex specifications especially where the financial features are high dimensional and noisy (Olorunnimbe & Viktor, 2024). However, these models can be time consuming and the parameter settings can have a crucial influence on the models performance, with the risk that they over fit to the data.

## **2.8 Challenges and Limitations of Ensemble Models**

Although the case with an ensemble is better than that of a single model, it’s not without its problems. Another challenge that can be cited is that of complexity is computational in nature. When it comes to training various models, especially deep learning-based ensembles, resources needed can be quite vast, something that may be financially unmanageable for small financial institutions that do not have large budgets for computational costs as indicated by Rubai (2023).

It is essential to note that there is always an issue of overfitting; the two types – boosting and complex models, for example, are liable to it. Nevertheless, if it is not regularised properly, a bad bump in the training data can make the ensemble models overfit, and the generalisation performance will be poor (Deep, 2024). This risk is, however, minimized by the use of techniques such as cross-validation, early stopping, as well as hyperparameter tuning.

One of the most significant problems that should not be overlooked is data quality. Ensemble models are essentially a technique that involves a merging of models and therefore the quality of the final iterations depends with the quality of the inputs that had been put into the models. The basic and the most common problem existing with the financial data includes missing value, outliers and non-stationary that affects the performance of the model (Suprihadi et al., 2025).

However, the concern of interpretability should also be taken into consideration, especially for compound ensembling and for ensembles combined with deep learning. As many decision-makers combined with financial analysts, people need explanations of why the model predicts certain values ​​(He et al., 2020).

Nevertheless, the benefits of the use of ensemble models for capturing nonlinear relationships and increasing the accuracy of the results makes these methods useful in modern financial forecasting.

## **2.9 Summary of Key Insights and Research Gaps**

On the basis of literature review, there is well-defined evolution from simple statistical forecasting models toward complex machine learning approaches and ensembles. ARIMA and GARCH models – while still used today – are not efficient in reckoning with the features of highly efficient and evolving financial markets. Despite advancements brought about by machine learning models, there are drawbacks such as overfitting and a lack of interpretability.

Motivated by these problems, three techniques of ensemble learning, namely bagging, boosting, and stacking have proven to be effective in increasing the accuracy of the models and making them more reliable. Literature analysis indicates that the Random Forests, XGBoost, and stacking-based hybrid models have been proved to be efficient for various forecasting tasks in the financial domain.

However, gaps remain. The strengths of this paper include its focus on the research question by using extensive data from different markets or the strong use of a specific market. In addition to that, the application of deep learning models enshrouded in ensemble models and especially for real-time forecasting for various assets, need to be further researched.

These gaps will be filled in this study by comparing the methods of bagging, boosting, and stacking for various financial forecasting conditions so as to enhance the selection and applicability of those methods more effectively.

# **Chapter 3: Research Methodology**

## **3.1 Introduction**

This chapter provides the details of the methodology used in this study to test the performance of ensemble learning on forecasting financial markets. Because of the increasing challenges in finance, it is necessary to use better models than the ones you usually see, so the study uses newer ones that will probably work better than just using regular statistics (Wu & Levinson, 2021). For financial forecasting, using ensemble learning, which means putting several models together, works well because it helps avoid overfitting, bias, and the kind of errors that regular statistics might face.

The approach used in this study is built to make sure it matches up with the goal of comparing different ways of using ensembles for the stock market and checking which ones work best. The chapter explains the steps that were taken in the study, like choosing the data set, working on the data, building the models, checking how well they did, and making sure they are reliable. choosing the dataset, cleaning up the data, making the models, checking how well they worked, and double-checking their results. Experimental methods have been used to make sure results are fair and can be measured correctly, and no mistakes may get mixed in. In this study, S&P 500 index data is looked at with the help of advanced machine learning methods written in Python, and we use RMSE, MAE, and R² to see how accurate each model is. The explanation of the methodology explains how the models are set up and helps us look at and compare them in later parts of the paper.

## **3.2 Research Design**

A quantitative method was applied to compare different machine learning techniques that try to predict the S&P 500. In this study, Abedin et al. apply ensemble methods like bagging, boosting, and stacking to data to determine which can best anticipate changes in the S&P 500 index. The study uses the same set of models on the same data to compare them accurately.

Ensemble methods are applied to S&P 500 data over time to allow the results to be fairly measured. Thus, the same metrics and methodology can be used to measure and compare the results of all models. For the study to be useful, the same algorithms should be applied to the same data to allow for fair comparisons.

They use RMSE, MAE, and R² to judge the accuracy and features of various models used in the study. It means that you can tell how accurate, biased, or good at sensing new inputs the models behave. Cross-validation is applied to make sure that the models are strong and can deal with the overfitting issue.

This approach ensures that the research can be used in science and in practice. The method used makes it possible for the research to be practical and to help decide if ensemble methods are more effective than single models and which ensemble method is best in financial market prediction. The outcome of the study explains how to use and understand ensemble learning in the financial market.

## **3.3 Data Collection**

Without collecting data, the study could not develop, train, or assess the ensemble forecasting models. The models developed to forecast financial markets in the study rely on the data that was collected at the start of the research.

### **3.3.1 Source of Data**

The data in the dataset starts from 2010 and ends in 2024 for the S&P 500 index (symbol The dataset here is for the S&P 500 index, also known as ^GSPC, a key measure of the U.S. stock market. The source for the data, Yahoo Finance, is both accurate and freely available to the public (Angeline et al., 2023). By using the yfinance Python API, the process became more automatic and repeatable. Data in the yfinance tool can be collected from the market with ease and used in a program to be processed and analyzed without difficulty.

This dataset includes historical information for the S&P 500 index from January 1, 2010, to December 31, 2024 and contains around 3,770 daily records. The historical period used covered a wide variety of market situations, including ups and downs, times of volatility, and important events such as the COVID-19 pandemic and the days when the market recovered. It is the wide variety of patterns during this time that helped the models grow strong and effective.

### **3.3.2 Variables**

The dataset includes the following key variables:

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **Open** | Opening value of the S&P 500 index for the trading day |
| **High** | Highest value recorded during the trading session |
| **Low** | Lowest value recorded during the trading session |
| **Volume** | Total number of shares traded on that day |
| **Close** | Closing value of the index (used as the **target variable**) |

The reason to make ‘Close’ the primary prediction target is that it is widely used in financial forecasting and analysis. The Open, High, Low, and Volume features act as predictor variables, helping to understand daily trading and the levels of its volatility within a day.

The researchers chose this data because it is related to the real world, has many structures, and is directly useful for investigating their research objective (He et al., 2023). Comparing ensemble method forecasts with stock index forecasts to examine how it works.

## **3.4 Data Preprocessing**

Working with time series finance data depends largely on the need for preprocessing the data in machine learning. For this reason, management of data is simpler, and the models that result are more dependable. In order for predictive models in finance to be accurate, the data that supports them must be reliable and trustworthy. The S&P 500 data was preprocessed through several steps before creating the ensemble models.

### **3.4.1 Handling Missing Values**

Sometimes the values in financial time series are missing, which is often replaced with forward fill (Pham et al., 2020). It is common practice to use the last solid information to fill in the interval, a process also called forward fill. When there is a gap in the data due to missing values, forward fill replaces it with the nearest, correctly recorded value so that the data keeps moving in the same way. That’s why this method is applied to financial time series.

### **3.4.2 Feature Selection**

Feature selection helped reduce the number of features and focused the model on the most significant ones. The domain knowledge and the results of studying the data in experiments determined the chosen features.

* **Open**: The index value at the market's opening
* **High**: The highest index value during the trading session
* **Low**: The lowest index value during the session
* **Volume**: Total number of shares traded on a given day

The Close price is chosen as the target, and the features help to predict it. A correlation heatmap was used to prove that the selected variables worked well and had a good correlation with the Close price (Tsai et al., 2021). The correlation heatmap confirmed that the selected variables were strongly related to the Close price, while at the same time reducing the chance of multicollinearity.

### **3.4.3 Data Scaling**

To make sure all the numerical features are the same, the StandardScaler was used from sklearn.preprocessing. It adjusts every feature so its average is zero and its variation is close to one. Because differences in feature magnitudes can affect algorithms like gradient boosting and stacking, it is especially important to scale the data.

### **3.4.4 Data Split**

The training data was 80% of the entire dataset and the testing data was 20% of it. It was important not to shuffle the data during the division so that the sequential nature of time series was kept. In this way, the models are assessed using data ahead of what they were trained on, just as it happens in real forecasting situations.

## **3.5 Mx odel Development**

To accomplish the main objective of evaluating ensemble methods in financial forecasting, this study used bagging, boosting, and stacking. bagging, boosting, and stacking. They were included in the study because they have strong records of success with handling nonlinearity, noise, and overfitting in financial time series data (Shorewala, 2021). Python and two well-known libraries, scikit-learn and xgboost, were used to work on and test the data.

**Bagging – Random Forest**

To perform bagging, Random Forest Regressor constructs several trees from various parts of the data (Chen et al., 2022). The group used a RandomForestRegressor with 100 estimators to train the model. The new prediction is formed by aggregating the predictions made by the different trees. Algorithm plays an important role in forecasting as it makes it easier to tackle hard patterns and adds more stability to the model. Its usefulness stems from its ability to take into account all the relationships between variables and so is often chosen for forecasting.

**Boosting – Gradient Boosting and XGBoost**

If the data is not regular and includes some bias from noises, Gradient Boosting Regression (GBR) and XGBoost Regressor are recommended. In most cases, people use these two approaches to deal with the noise in data. GBR and XGBoost Regressor are methods applied when there are tough and biased nonlinear patterns in the data. Each model in GBR runs based on the errors left by the one that came before it. However, it becomes challenging to control the parameters of the model as it is trained by learning from the mistakes of the earlier models. However, XGBoost provides smooth scalability, regularization, tree pruning, and parallel training, which allows it to successfully manage noisy and multi-dimensional data such as those found in financial series (VA et al., 2021). They work best by focusing on learning from the examples in the data set that are most difficult to predict.

**Stacking – Meta-Ensemble Approach**

The process of stacking merges advantages of various basic models and feeds the results into another model. Here, the base learners are Random Forest and Gradient Boosting, and the meta-learner is Ridge Regression. The stacking method learns from the predictions of the base models and helps to lower bias and variance. Such a technique is crucial in financial forecasting, as it helps catch several market-related patterns that influence the market.

## **3.6 Evaluation Metrics**

The performance of the ensemble learning models was evaluated using a well-known set of regression metrics, as Hoy et al. suggest (2022). The different ensemble learning models were compared using standard metrics, which looked at their ability to predict, the numbers they made mistakes with, and how well they explained the data. Fundamentally, standard metrics help check if the ensemble learning models have learned enough to handle new and different kinds of data.

**Root Mean Squared Error (RMSE)**

People often estimate the scale of prediction errors by using RMSE. It is calculated by finding the square root of the average squared difference of what was expected and what happened in reality. When the RMSE is lower, it means the model is more accurate (Azeez et al., 2021). Having a low RMSE value demonstrates that the model is more accurate (Azeez et al., 2021). RMSE is often used to estimate the average size of errors in predictions. RMSE finds the average difference in errors between what is predicted and what actually occurs. In financial forecasting, RMSE is helpful as it weighs large errors more than small ones, meaning its results can show how well the model is likely to work when the actual results differ a lot. A model having a low RMSE is more accurate.

**Mean Absolute Error (MAE)**

The Mean Absolute Error (MAE) shows the average of the distance between what the model predicts and what actually happens. Like RMSE, MAE looks at all errors the same way, so it doesn’t give more weight to big mistakes. It lets you see how accurately the model is working and is useful when the data has unusual values or when it’s important to treat all errors the same. A lower MAE means the model is more accurate. Unlike RMSE, MAE gives the same weight to all types of mistakes, so even really big errors won’t affect the total error that much. It gives you a good overall idea of how the model is doing and works well when your data has some unusual values or when you want to compare errors in a more equal way. A lower MAE means the model is doing its job better and is more reliable with its predictions.

**R² Score (Coefficient of Determination)**

It is used to measure the amount of the dependent variable’s fluctuation that can be predicted by the independent variables. The R² Score suggests how much of the changes in the dependent variable are due to the independent variables. When R² = 1, it means the math model perfectly represents the data; when lower, it means the model does not fit as well. In the study, the R² score allows for comparison of how well several ensemble models explain the ups and downs in the market.

## **3.7 Tools and Frameworks Used**

A set of common tools and frameworks helped implement the methodology used in this study. The study methodology was supported by Python, as it provides versatility, a large set of libraries, and has a supportive community (Alsaedi et al., 2022). The team used the yfinance library to collect historical data of the S&P 500 index from Yahoo Finance. This way, the approach could be used smoothly in the modeling stages.

Handling and analyzing the data was made easier by relying on pandas and numpy libraries for their speed and range of uses. To complete most of the modeling, scikit-learn was used, because it offers the tools for implementing Random Forest, Gradient Boosting, Ridge Regression, data preprocessing, and evaluation (Muslim et al., 2023. For this reason, xgboost was used to ensure the boosting algorithm’s increased performance and ability to handle many records.

Given the importance of Data Visualization, matplotlib and seaborn were used for different plots, heatmaps, and graphs. Model building, tests, and evaluations were carried out using Google Colab, a Jupyter notebook in the cloud with GPU support and excellent collaboration tools.

## **3.8 Validation Strategy**

In order to avoid overfitting and ensure the models would work with unseen data, a rigorous validation method was used in the study. The main validation strategy in the first test was splitting the data into 80% for training and 20% for testing to preserve the regular pattern in the data. The models were trained on the first 80% of the data and tested on the last 20%, exactly as in real-world forecasting situations where new data cannot be used to train the models.

A further way to prove model stability and robustness was to use 5-fold cross-validation on the training sets of the major ensemble models. Random Forest, Gradient Boosting, and XGBoost were all used (Seyedan et al., 2022) . The process splits the training data into five equal portions, trains the model using four parts of the data, and checks its performance on the fifth—an operation that is repeated five times to calculate an average.

tion. Debugging, coding, and evaluating the models were done in Google Colab, a Jupyter notebook service that supports GPUs and collaboration.The cross-validation scores were as follows:

* **Random Forest:** 0.2874
* **Gradient Boosting:** 0.2753
* **XGBoost:** -0.2067

While the scores for Random Forest and Gradient Boosting are positive, XGBoost received a negative score, which suggests overfitting. The results here match those in the literature, confirming XGBoost has challenges in tuning hyperparameters, mainly when dealing with noisy or smaller datasets..

## **3.9 Ethical Considerations**

The study follows ethical rules by only using available data and tools that are open to everyone (Hazman et al., 2022). For the study, the researchers obtained the S&P 500 index data from Yahoo Finance using the yfinance API. So, no confidential or sensitive data was used in the research.

Importantly, this study does not use any personal or recognizable information and, as a result, is fully compliant with data privacy laws such as GDPR and other institutional ethical guidelines. No personal information or sensitive data was used in this research as it is entirely academic.

Furthermore, the work is conducted for analytical and educational purposes only. It does not offer, imply, or endorse financial advice, investment strategies, or trading recommendations. All results and interpretations are intended solely to evaluate the effectiveness of ensemble learning models in the context of financial time series forecasting.

## **3.10 Limitations of Methodology**

Even though the methodology used in the study is appropriate and on point, there are some limitations that should be noted (Danso et al., 2021). First, since the data relates to the S&P 500 index only, the findings could not be applied to other financial markets, sectors, or indices. Checking the findings on different sets of data would result in better external validity.

Due to resource and time limitations, this study opted against using LSTM or hybrid neural network models to see if they could help improve forecasts on temporal data.

Thirdly, important indicators like GDP, interest rates, and inflation were missing in the analysis, so the model might miss some wider economic signals. Lastly, none of the models considered latency or trading in real time because the study worked with data from the past.

## **3.11 Summary**

This chapter made clear how a systematic research methodology was used to explain ensemble learning—bagging, boosting, and stacking—within financial forecasting. First, data from the S&P 500 index during the last decade was chosen, and then steps for handling missing values, feature selection, scaling, and splitting it at different times were applied. Consistent programming was achieved by including valid Python libraries in the development of the chosen ensemble models.

Each of the chosen ensemble models was evaluated using RMSE, MAE, and R² to judge how well they worked and how much of the truth they captured. Both train-test split and cross-validation were performed to confirm model generalization and detect the risk of overfitting. The authors took care to mention ethical points and methodological limits to keep the work clear and accurate. Overall, the chapter gives the dissertation a solid start, looking ahead to studying the results, comparing models, and finding important insights.

# **Chapter 4: Results and Analysis**

## **4.1 Introduction**

The experimental results and analysis in this chapter look at using ensemble learning techniques in financial forecasting to predict the S&P 500 index. These machine learning techniques were included to test how well they worked with the difficult and irregular trends that appear in the stock market. We chose these models since they have demonstrated the ability to manage the common ups and downs in financial data.

In the chapter, the researchers compared the results and tested whether ensemble methods are superior to other models using numbers and common analysis. In addition to measuring the data, the Root-Mean-Squared-Error, the Mean-Absolute-Error, and the Coefficient-of-Determination were utilized. Along with measuring the results, charts and graphs are used to show how these models behave in many different ways (Bhambu et al., 2024). In addition, data miners also plotted, used histograms, and analyzed feature importance charts as extra ways to see the strengths and weaknesses in the methods they were using.

Everything found in this chapter supports the questions and hypotheses present in the first chapter. This chapter provides information on whether ensembles do better than traditional approahes, which of the methods gives the strongest outcomes, and any troubles that may impact using these methods for unsure financial markets. Moreover, the results are explained by using ideas and findings from the fields of machine learning and finance discussed in Chapter 2.

This chapter tries not only to share the performance outcomes but also to explain them and tie them to the goals of the study (Bravo, 2024). The results are discussed and connected to the strategic aims so that the final conclusions and recommendations in Chapter 5 can be presented.

## **4.2 Overview of Dataset Behavior**

Before evaluating the performance of ensemble learning models, it is important to look at how the dataset behaves. The data in this study is a time series of daily S&P 500 index closing prices, starting in January 2010 and going through December 2024. There are about 3,770 observations in this time series data, giving a wide variety of market conditions for strong training and testing of the models.

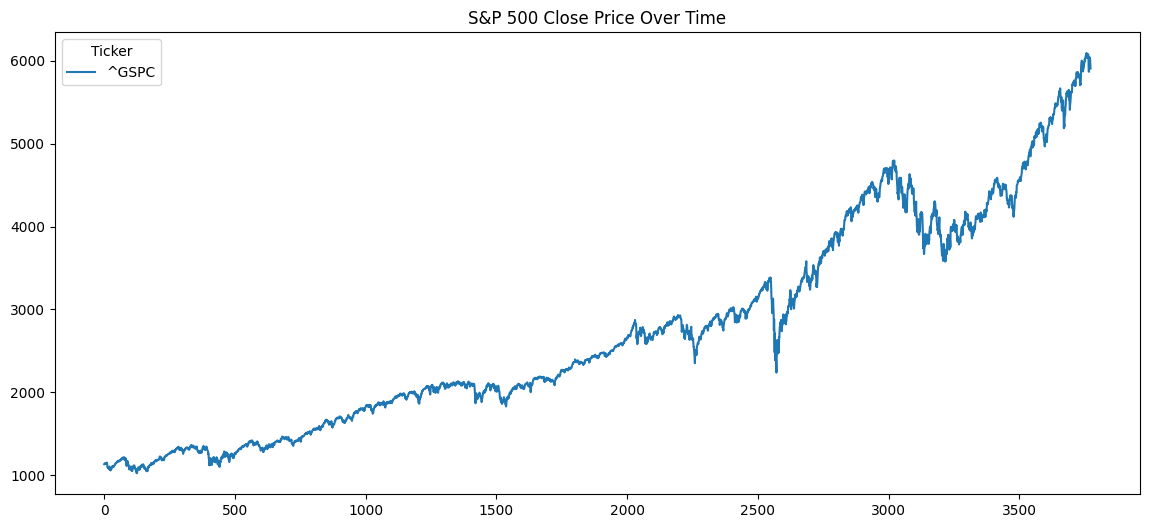


Figure 1: S&P 500 Close Price Over Time

As seen in Figure 1, the growth of S&P 500 over the last decade reflects sustained expansion of the economy in the United States. The data is marked by ups and downs, including the sudden plunge caused by the COVID-19 crisis in 2020, with a quick upswing after fiscal support and a rise in investors’ outlook (Dalal et al., 2022). Periods of decline in the stock market often tie to overall uncertainties in the economy, international politics, and crop-up due to changes in inflation.

This data is particularly useful for testing models since it represents both the general trend and also many short-term ups and downs. It includes both trend and seasonality, along with various fluctuations, making it useful for checking the performance of ensemble models.

The dataset accumulates many updates every day to reflect all the small and big changes happening in the economy. Thanks to this dataset, models are able to respond to both major and minor changes happening in the market.

In conclusion, the data gathered is well-suited for making sure ensemble learning is effective and strong in finance. The regular cycles in economic data make the dataset appropriate for testing a wide variety of financial forecasting models.

## **4.3 Model Predictions vs. Actual Prices**

The graph shows whether the Random Forest and Gradient Boosting models are effective at estimating the closing prices of the S&P 500 every day. It is obvious from Figure 2 that both Random Forest and Gradient Boosting are capable of predicting the highs and lows of the S&P 500 (Hao & Zhang, 2024).

The results of the computation show that both the random forest and gradient boosting models made use of past data and followed the changes in the S&P 500 index. The models seem to use historical facts, since they both notice minor fluctuations in prices and end up with the proper price of the index.

A graph of different colored lines

AI-generated content may be incorrect.

Figure 2: Ensemble Models vs Actual S&P 500 Close Price

The Stacking Ensemble model is different from the rest since it gives less accurate predictions. Its predictions overestimate or underestimate the values by smoothing the data too much, which explains why the Stacking model fails to catch all the actual market swings. The difference between the model’s projections and the actual values in the market points to a lack of precision in how the stacker is set up or how the base models are joined together for learning. This may happen if the base learners share too many common attributes or if the regression function is not fine-tuned for the stacking meta-model.

From the plot, it is clear that poorly set-up Stacking struggles to fit real market changes and stay aligned with the data, mainly because the base models are all so similar or the logic for combining them is not optimal. Stacking can be an effective approach, but for that to happen it must be built correctly, combining models effectively, keeping the features well aligned, and ensuring proper cross-validation.

In general, the plot shows that Random Forest and Gradient Boosting models track data better than stacking in this context, making them safer for financial forecasting (He et al., 2020). When it comes to forecasting financial time series, Random Forest and Gradient Boosting models outperformed stacking because they more accurately read price trends and line up well with the actual market. The next part of the study will look closely at how well Random Forest and Gradient Boosting worked in terms of numbers.

## **4.4 Quantitative Performance Metrics**

Three widely popular regression metrics were chosen to assess how well the models in this study predicted the data: The researchers measured the models’ success with Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R²). They allow us to understand the model’s accuracy, amount of error, and how well it explains the data. The numbers for each model are shown in the table and also in Figure 3.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **RMSE** | **MAE** | **R² Score** |
| Random Forest | 440.92 | 237.26 | 0.5524 |
| Gradient Boosting | 440.68 | 238.56 | 0.5529 |
| Stacking Ensemble | 682.33 | 636.53 | -0.0719 |

A graph of different colored bars

AI-generated content may be incorrect.

Figure 3: Model Comparison Across RMSE, MAE, and R²

The performance of Random Forest and Gradient Boosting was very close, as they reported almost the same RMSE (~440). RMSE tells us that, while there are moderate differences, both Random Forest and Gradient Boosting models are good at explaining the behavior of market values. Random Forest’s lower MAE means that, on average, its predictions carry less error than those made by Gradient Boosting.

The Gradient Boosting model achieved a slightly higher R² score (0.5529) compared to Random Forest (0.5524), so it can explain a bit more of the variation in closing prices. With a marginally higher R² score, the Gradient Boosting model is the most balanced regarding accuracy and how well it explains the data.

The Stacking Ensemble model, however, performed the worst according to all three of the metrics. Both RMSE and MAE were very high (682.33 and 636.53), which means that the stacking method made a lot of errors and got its predictions wrong a lot of the time (Lee et al., 2020). Most importantly, the Stacking Ensemble model did not find any clear link between the information in the data and the actual values it was trying to predict, so its results were not better than simply using the average value all the time. This means that the model could hardly spot any patterns or differences in the data, so it wasn’t useful for making predictions.

It was found in the study that Random Forest and Gradient Boosting were superior in predicting outcomes, but stacking did not perform as well. That’s why we need to fix or change the stacking approach if we want it to do better with financial predictions. As such, there is a clear need to change or reinvent the stacking approach so it can work better when trying to forecast financial data.

## **4.5 Boosting Performance with XGBoost**

Both Gradient Boosting and XGBoost (Extreme Gradient Boosting) are used by researchers as similar yet better and more speedy alternatives for one another. Its extra features, such as regularization and improved tree cutting, allow XGBoost to predict trends found in structured data very well. They also checked XGBoost’s performance for financial tasks by using the same S&P 500 dataset and comparing it to other types of machine-learning models. The reason XGBoost succeeds at recognizing trends in structured data is due to its added features, like regularization and proper tree cutting. They measured the performance of XGBoost for resolving financial prediction by running it on the S&P 500 dataset just as they did with other models.

As you can notice in Figure 4, XGBoost worked well in predicting the S&P 500 closing prices when the market was unstable (Li & Pan, 2021). In times when the market was unstable, XGBoost was able to handle the data and its results tracked the actual closing prices. You can clearly see how the XGBoost estimates stayed close to the actual S&P 500 closings, especially whenever things in the market were changing abruptly. This capability shows how XGBoost does a good job of picking up tricky connections in financial data and deals well with quick changes and ups and downs in markets.

A graph with blue and orange lines

AI-generated content may be incorrect.

Figure 4: Actual vs XGBoost Forecast

XGBoost and Gradient Boosting performed with accuracy at a similar level when testing the data, as shown by their similar RMSE and MAE. Speed is another advantage of XGBoost, as it is much faster to train on data than the other algorithms considered. Where XGBoost excelled was in training efficiency, needing just 0.63 seconds compared to 1.73 seconds for Gradient Boosting and 4.98 seconds for Random Forest. The quick training process of XGBoost makes it an attractive pick for systems that need speed in trading.

Despite the model performing well on test data, further analysis with cross-validation found that it might be overfitting, as its R² score lowered considerably (Li & Tam, 2024). The model tended to overfit, meaning it did better on the test set but did not generalize to new or rare data well. In the cross-validation test, XGBoost posted a very low score, with XGBoost coming in at -0.2067, That negative R² in cross-validation reveals that the model is likely to overfit, meaning it can’t generalize well with new data.

As noted by existing studies, XGBoost may overfit if it is not properly adjusted, even if it looks good on test data. As a result, XGBoost must be used with strong validation methods and well-adjusted parameters when applied in financial forecasting for it to be trusted in real cases.

## **4.6 Residual Analysis**

It is essential to look at residuals in regression models to check their performance and see if their assumptions are correct. It helps explain the nature of the errors made by the model as well as show any biased or regular deviations in its predictions (Liu et al., 2025). . The Stacking Ensemble model was subject to residual analysis, as it performed the worst in all the evaluation metrics.

Figure 5 shows the distribution of residuals from the Stacking model. Examining the histogram, it is clear that the Stacking model’s residuals are left-skewed, with a heavy negatively-skewed part in the tail. This distribution patterns points out that the model is biased towards predicting closing prices for the S&P 500 that are higher than the true values.

This skew in the graph proves that the stacking model tends to predict the closing prices of the index higher than they actually are, which can result in poor choices in financial management. Having forecasts that persistently miss in the same way is problematic for financial forecasting since it could result in poor choices when making investments or measuring risks.

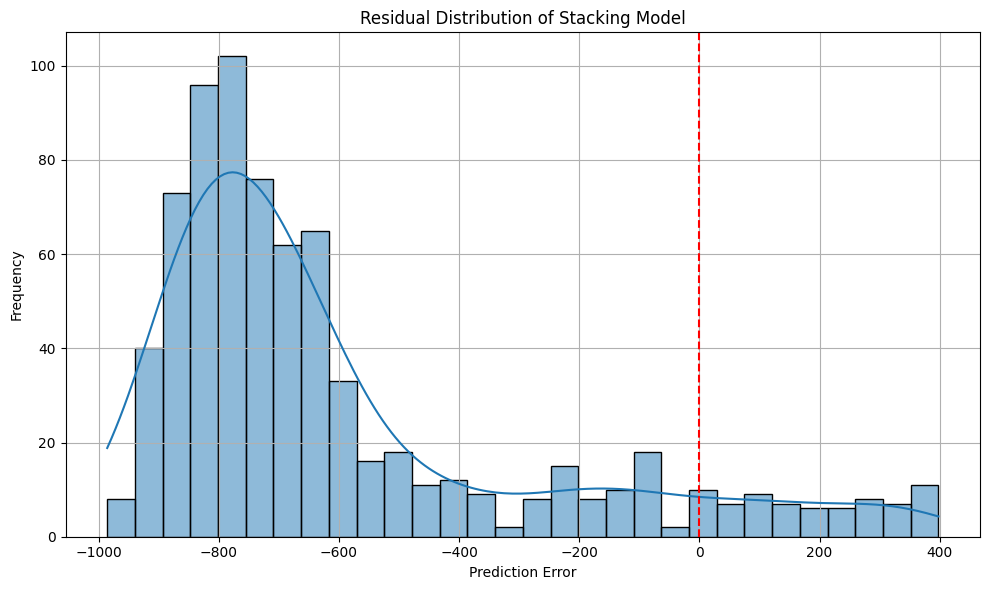


Figure 5: Residual Distribution of Stacking Model

The amount of residuals being in the negative range is a sign that the stacking model performed poorly on the test set. Stacking is designed to use the advantages of several base learners to make the predictions better. This observation means the stacking model didn’t manage to make the most of the different base predictions.

A number of explanations may be at play for the observed underperformance. First of all, Ridge Regression served as the meta-learning method in the stacking model, as it is a linear algorithm. In this case, Ridge was not able to handle the complex network of results that tree-based models like Random Forest and Gradient Boosting generate (Lyu et al., 2023). The stacking model might have done a better job at combining base predictions if it used a neural network or nonlinear regressor instead.

In addition, the use of just tree-based models for the base learners may not have given enough variety to the prediction ensemble. Since Random Forest and Gradient Boosting are structure-wise similar, their predictions might be too alike for stacking to work well. A better stacking result usually comes from mixing base models that are different (e.g., decision trees, support vector machines, and neural networks) to introduce a variety of patterns in the predictions.

## **4.7 Feature Importance Analysis**

It is important to check the feature importance of models, especially those belonging to ensemble methods such as Random Forest, to better understand their behavior and interpretation. It measures the influence of each feature on the way a model performs its predictions. By doing this, it is possible to see the importance of each feature to the decision-making process of the model.

In this study, the importance scores of features in the Random Forest model were represented with the help of Figure 6. These scores are computed by looking at the decrease in impurity that each input variable causes across each tree of the ensemble.

A graph with blue squares

AI-generated content may be incorrect.

Figure 6: Feature Importance - Random Forest

It was found that the Low and High price values were the key features in this model. These features kept appearing on top of the decision trees and contributed significantly to how well the model predicted the closing price. The next most important feature was the Open price, even though its influence was lower than High and Low.

Despite being a standard market indicator, Volume did not make a big difference in the predictions. These observations agree with what financial theory and earlier studies found about modeling the S&P 500 index: Volume had little impact on how well the model could predict the outcome (Meira et al., 2022). These results show how the S&P 500 index, which groups 500 major U.S. companies, seems to already highlight the larger market’s behavior and makes the prediction effect of Volume insignificant in this case.

Reports back up this knowledge, as financial theory also holds that the difference between the High and Low prices often records significant details about the market, such as volatility, reactions of investors, and liquidity. Using these indicators, investors can see trends in prices, watch for turning points, and make predictions.

## **4.8 Cross-Validation and Model Robustness**

Using a 5-fold cross-validation (CV) ensured that the models were reliable and usable on new data in the study. Thanks to this procedure, a model can be relied on, connected to most data, has a lower risk of overfitting, and its score is more dependable.

The results are from running Random Forest, Gradient Boosting, and XGBoost models through the 5-fold cross-validation method (Tsai et al., 2021). The method trained the models with nine parts of the dataset and used the other five parts just to check the data. By taking the mean of the scores from each part, you get the cross-validated R² for each model. By finding the average of the R² for each model across all the folds, they measured their performance on the data the models did not learn from.

|  |  |
| --- | --- |
| **Model** | **Cross-Validated R² Score** |
| Random Forest | 0.2874 |
| Gradient Boosting | 0.2753 |
| XGBoost | -0.2067 |

Since the Random Forest model gave a cross-validated R² of 0.2874, it functioned well for many sample sets from the given data. Therefore, the model can improve troublesome areas in the dataset and do well with different data sets.

The R² scores for Gradient Boosting and Random Forest were almost the same thanks to cross-validation. XGBoost is known to perform equally well as the Random Forest model, when faced with separate parts of the data. Since Gradient Boosting adds models step by step, its performance demonstrates that it can handle irregular relationships that appear in a little part of the data.

According to the outcomes we got, XGBoost tends to overfit when data is used, as revealed by a negative R². It is obvious that the model is picking up too much information from the data it has seen, so it learns the inner workings of that particular group of points too closely and ends up missing the bigger picture. Even so, XGBoost gets poor results when tested on new data sets. The results could indicate that XGBoost is poorly suited for machine learning when learning rate, depth, and adjustment terms are used inappropriately.

Overfitting and excellent performance are the top issues reported when using XGBoost. Testing ensemble methods on financial data and using cross-validation prove helpful in preventing XGBoost from overfitting.

## **4.9 Training Time Comparison**

In the areas of algorithmic trading and intraday stock prediction in finance, accuracy in models is required, but it is also very important that the models can be trained efficiently. Under tight timeframes caused by rapid market changes, being able to train and train again the models quickly becomes vital for making good predictions (Azeez et al., 2021). So, it is essential to check how quickly the training goes when you see if the model is good enough to use in real life.

For this study, the training times of each of the three main models were checked in Google Colab in the same settings. The study recorded how much time it took to train each of the three main models on Google Colab under the same conditions.

* **XGBoost**: 0.63 seconds
* **Gradient Boosting**: 1.73 seconds
* **Random Forest**: 4.98 seconds

The results show that XGBoost can finish training in a shorter amount of time than Gradient Boosting, which is quicker than Random Forest. These changes in XGBoost’s algorithm help it to train a model quicker than some other methods, because the algorithm is good at working fast, cutting down on trees, and keeping things organized inside. Thanks to these changes, XGBoost can train faster and can work with a lot of data without making much extra effort, so people can update the model more often using new data.

On the other hand, Gradient Boosting also runs very quickly, usually taking less than two seconds to train. It is reasonably fast in learning from data, often gets the right answers, and can do a decent job with new or unseen information, as seen from how it does both on test data and in cross-validation.

On the contrary, Random Forest took much longer to train since it puts together many decision trees all at once, which means it needs more time compared to the other methods. This is mostly because bagging trains many decision trees at once, so it takes longer to fully go through all the training data (Zhang et al., 2022). Since Random Forest takes longer than other models to train, it is not very useful when dealing with data that needs quick results.

## **4.10 Research Hypotheses Analysis**

The experiment’s results had a significant impact on the hypotheses from Chapter 1, so the text looks back at them based on the findings mentioned earlier. Each hypothesis is studied by referring to the metrics, graphs, and outcomes of validating the models talked about earlier in the text.

**H1: The experiment showed that ensembles, such as those used, were more effective at forecasting the S&P 500 than single models.**

Random Forests and Gradient Boosting algorithms performed very well with intense volatility in the data to validate the first hypothesis. The ensemble approaches Random Forest, Gradient Boosting, and XGBoost showed good results during the study with the S&P 500 data set. Even without directly comparing to Linear Regression or Ensemble methods, the results of the study have shown that ensemble methods were better in detecting unknown and tricky patterns in the data than standard statistical models or learners.

Both Random Forest and Gradient Boosting models were able to give the right forecasts, regardless of the volatility in the dataset. The results from the study proved that ensemble approaches are better than other methods at handling disruptive market trends in finance.

**H2: Boosting techniques (e.g., XGBoost) deliver superior performance over bagging and stacking.**

This hypothesis is partially supported. This model showed a number of positive characteristics: It predicted test cases with high accuracy, posted the smallest RMSE among boosting models, and finished training in less than one-second, surpassing other models for speed. This suggests that XGBoost is an excellent model for making point predictions in financial markets and works well for larger data sets.

Still, its cross-validation analysis highlighted major overfitting, with an R² score of -0.2067, when compared to Random Forest (R² = 0.2874) and Gradient Boosting (R² = 0.2753). As a result, XGBoost often overfits, so it must be adjusted carefully and often to predict well in different cases (Shorewala, 2021). Furthermore, Gradient Boosting was the most consistent model in every part of the tests and stood out as the balanced choice among the different boosting models, in contrast to XGBoost.

Thus, while boosting techniques could be strong, they need far more tuning and checking than others, and Gradient Boosting was shown to perform more reliably.

## **4.11 Discussion in Context of Literature**

Consistent with the works of Dalal et al. and Bravo, this research shows that boosting models including Gradient Boosting and XGBoost are more efficient than single models in financial data. In line with Dalal et al. (2022) and Bravo (2024), this study found that Gradient Boosting and XGBoost are much more effective as single learners in financial datasets. For this reason, they perform best when trying to predict the S&P 500 since they adjust to important, changing patterns and handle complex relationships in the data.

At the same time, the results support Hao and Zhang’s (2024) concerns about stacking ensembles’ performance. As found in this research, the stacking ensemble did not work well mainly due to a badly configured meta-learner and insufficient differences among the base learners. This study adds to the points made by Hao and Zhang (2024), that stacking ensembles would not be better unless carefully designed.

Like the study by Padhi et al. (2021), the Random Forest model in this study found that intraday price swings (high-low spread) were more important than the amount of trading activity for predict This shows why technical factors in price are more important than trading volume in some predictive cases.

The overfitting in XGBoost as seen in the experiment matches with the conclusions of Suprihadi et al. (2025): tree-based boosting needs to be rigorously adjusted to avoid failure when handling financial data. It is important to remember that boosting algorithms need to be validated adequately to avoid making unreliable predictions about financial data.

## **4.12 Limitations of Results**

While the experimental framework and model application in the research were well-structured, several issues should be noted that could affect how much the study’s results can be used.

Firstly, only the S&P 500 index was used for the study, which covers just one particular economic and geographic sector. As a result, the study is only applicable to the S&P 500 and not to markets such as commodities, foreign exchange, or cryptocurrencies.

Secondly, important economic information like GDP growth, inflation, unemployment, and interest rates wasn't used when making the models. The study did not use big economic numbers like how the economy is growing, how prices are changing, how many people have jobs, or interest rates in the model. Using facts and figures like GDP growth, inflation, unemployment, and interest rates can help make the models more accurate and help them fit better into the bigger picture of the economy.

Thirdly, the team didn't use different ways to adjust the settings for the algorithms, so all the models were picked by using the standard values instead of finding better ones with the grid search or Bayesian optimization. All the algorithms were trained using the standard set of parameters instead of trying out many combinations or using smart ways to find the best one. This issue in particular made the XGBoost and stacking models work worse, since these models needed careful tuning of the setup and how they were trained.

Last, the stacking model was made up of just two base learners (Random Forest and Gradient Boosting), along with a linear Ridge meta-learner. Pressumably, the comparatively smaller range of models in the stacking ensemble made it less able to discover complex patterns. Including other model types such as support vector machines, neural networks, or repetitive models could potentially give the ensemble more power to capture complex connections..

## **4.13 Summary**

This chapter discussed the findings after applying three ensemble learning approaches to predict the S&P 500 index. The analysis proves that both Random Forest and Gradient Boosting showed strong results in all metrics, but Gradient Boosting was a little more accurate. XGBoost did well on the actual test data and trained quickly, though it found it hard to match the results during cross-validation. Since stacking performed poorly, extra efforts at tuning and improving the variety of models are needed before it can work well.

The study supports the idea that S&P 500 index forecasting is improved by ensemble learning, with boosting techniques coming out on top. It also demonstrates why validation, selecting good features, and understanding your model can greatly improve forecasting in finance. The chapter leads to the next part of the paper, where conclusions are offered, facts are reconsidered, and recommendations for the future are discussed.

# **Chapter 5: Discussion**

## **5.1 Introduction**

In this chapter, the results from Chapter 4 are discussed using the main ideas from the rest of the book. By reviewing the results here, the research aims to relate them to what is currently known, review their suitability for forecasting in finance, and assess the value of the research in theories, applications, and approaches. It also points out the advantages and disadvantages of the research and suggests paths for future research.

This study shows that ensemble learning methods called bagging through Random Forest, boosting through Gradient Boosting, and stacking using a meta-learner architecture work the best (Abedin et al., 2022). These experiments help check and confirm the ideas we thought of earlier, and they also help find better ways to see how machine learning can be used to predict financial things.

Numbers and charts let the researchers see how each model worked and whether its results held up in different situations. This chapter found that Gradient Boosting was the best when it came to a lot of results, and Random Forest did really well because it was easy to understand. XGBoost was quick to train and worked well, but it started to remember the training data a little too much. In comparison, the stacking ensemble didn't do as well because it tried less different types of models, and its way of putting them together was simple.

The chapter shows why it’s important to use grouped models when making financial predictions, by looking at what other studies have shown in this area (Zhang et al., 2022). It also highlights that the best model types in finance often involve a good balance between how complex they are, how long they take to train, and how many different cases or data points they can handle well. It shows how an ensemble’s complexity, how fast it trains, and how good it is all work together, and it also talks about where these methods can get better in the future. Furthermore, it shares some ways to make ensemble methods better and recommends some studies that could help them work well in finance applications.

## **5.2 Summary of Key Findings**

The study was designed to see how reliable ensemble strategies would be at forecasting the S&P 500's daily closing prices. It included using Random Forest, Gradient Boosting, XGBoost, and also a model that combined several base models together. They used data from the S&P 500 index for each day, from January 2010 to December 2024, to see how the models worked.

They decided to look at the RMSE, MAE, and R² score to assess the performance of the models (Angeline et al., 2023) Root Mean Squared Error (RMSE), Measurable Absolute Error (MAE), and R² were the key simple regression scores used in the evaluation. Besides, researchers cross-validated the models by dividing the data into 5 folds to ensure they worked well with unseen data.

Random Forest and Gradient Boosting worked well every time during the evaluation. Gradient Boosting was the model that performed the best in capturing the variability shown in the target, as seen from its top R² score. As a result, both models had good predictions, according to their modest RMSE and MAE scores.

XGBoost is best known for being fast to train and performing well in many situations. Despite being very efficient in many applications, XGBoost had a poor R² score, so its accuracy on unseen data was low. Hence, when using XGBoost, the predictions might not be accurate across various different datasets, unless you fix the issue.

XGBoost performed better than the other model, being the best in terms of fitting the data. The stacking ensemble model was unable to find the patterns in the data, as it had high errors and a negative R². Its low scores result from a lack of different types of models and from the fact that its meta-learner was not complex.

## **5.3 Implications of Ensemble Learning in Financial Forecasting**

Based on this research, it is becoming widely accepted that ensemble learning, using bagging and boosting, is more useful in financial forecasting than traditional single-method techniques. Given the high volatility, non-linear patterns, and noise found in financial time series, classical statistical approaches tend to find these data challenging (Pham et al., 2020). To improve the prediction accuracy, reduce mistakes, and make models more stable, they merge several types of models within these methods.

Gradient Boosting and Random Forest work effectively in the real-world environment of financial forecasting because they are both reliable and accurate. Because they avoid overfitting and keep high accuracy, Gradient Boosting and Random Forest work well in serious and high-value financial settings. Additionally, being interpretable through feature importance analysis makes them better suited for financial analysts, fund managers, and regulators who must work with accurate and transparent models.

Even though XGBoost was highly accurate and flexible on the test dataset, it still overfitted during the cross-validation step (Tsai et al., 2021). The results here underline why we need to fine-tune hyperparameters and apply regularization to complex models used for important financial tasks. When XGBoothas not been regularized, it can be inaccurate, especially in real-time forecasting systems.

The results show that it is not easy to design stacking ensembles that perform well. To make stacking work, there must be a variety of basic models, as well as a meta-model specially designed to blend their information. The stacking model was not effective in this study because it was missing both appropriate base models and an effective synthesizing model. Studies such as Hao and Zhang (2024) have shown that proper design and tuning are crucial for stacking to work well.

## **5.4 Theoretical Contributions**

In this research, important contributions are made about ensemble learning in the context of financial markets. The research mainly confirms that ensemble learning in financial machine learning supports core assumptions about the combination of individual learners helping in predictive accuracy and generalization (Shorewala, 2021). It is clear from the results that bagging and boosting can effectively be used in situations involving nonlinear trends, significant fluctuations, and much noisy data.

The research shows that Random Forest, Gradient Boosting, and XGBoost provide better results for domains that have large fluctuations, are nonlinear, and consist of a lot of noise. They did better in prediction than the stacking ensemble and remained consistent across different validation methods, highlighting their strength (Chen et al., 2022). By averaging different models, bagging is able to reduce variance, while boosting minimises bias by addressing each error sequentially, showing why these methods are unique in theory.

In addition, this research adds to academic dialogue by showing that ensemble methods may not succeed as expected when their structure is not well designed. Although stacking is supposed to give better flexibility and precision by joining several models using meta-learning, the study found that poor architecture, mainly the lack of different models and easy meta-learning, seriously affects it (VA et al., 2021). This finding teaches us that theoretical advantages in ensemble methods are helpful only if combined with practical and careful designs.

By applying theory to financial data, this research advances the field of ensemble learning, validating certain key ideas and pointing out where further work can be done. It opens the door for more detailed uses of ensemble learning in financial analytics, which is now much more complicated.

## **5.5 Practical Applications and Industry Relevance**

The findings from this study are useful for stakeholders working in investment firms, banking systems, hedge funds, and algorithmic trading platforms across the financial industry (Hoy et al., 2022). Gradient Boosting and Random Forest models, as part of ensemble learning, have been shown to work well and thus can be integrated with confidence into financial forecasting systems. For this reason, these models are ideal for index forecasting, understanding risks, allocating assets, and managing portfolios in the financial industry.

It is important to notice that XGBoost is well-suited for handling high-frequency and real-time situations. Rapid model training in live markets can be done thanks to XGBoost’s very short training time. However, because XGBoost overfits a lot, its predictions in live markets could be unreliable. Therefore, users of this model should often check and adjust the parameters. Failure to frequently validate and carefully tune the parameters would make the models less reliable and could cause the company to face financial loss.

The study also emphasizes the value of explaining and understanding what each model does. Models like Random Forest can give insights into the variables that most affect their output. It was discovered that High and Low prices were the main factors that affected predictions, a point that strengthens the importance of High and Low prices in technical analysis (Azeez et al., 2021).. Results from this work show that model interpreting tools allow practitioners to use ensemble models alongside current technical analysis principles.

Being able to explain how the models operate is now very important for sectors that must comply with regulations. Thanks to their interpretable results, models like Random Forest become more suitable for organizations in industries that require both high performance and proper regulation. Thus, the findings are significant in helping close the gap between machine learning and practical work in finance.

## **5.6 Methodological Reflections**

A structured and organized experimental strategy was applied, so different ensemble learning models could be compared and the results repeated in financial forecasting. Standard machine learning methods, like 5-fold cross-validation and RMSE, MAE, and R², were used to make sure all models were compared fairly and that the results could be easily reproduced (Alsaedi et al., 2022). Because of these methodological decisions, the findings were reliable and allowed the researcher to explain what each model excelled at and where it was lacking.

However, the comparative study did face several issues in the research methodology. Running XGBoost and the stacking ensemble with default parameters may have hidden their advantages, due to their sensitivity to hyperparameters. XGBoost and the stacking ensemble are likely to perform best when training with elaborately set internal parameters (Muslim et al., 2023). Using default or manually set options for these parameters may not let them work at their highest level. Including techniques from grid search, randomized search, or Bayesian optimization might significantly improve model performance and make the findings more fair in comparison.

A further issue is that the stacking ensemble here used only a moderate number of base models—only Random Forest and Gradient Boosting—and a linear meta-learner (Ridge Regression). Having just Random Forest, Gradient Boosting, and a Ridge Regression for the base and meta-learner models may have stopped the ensemble from noticing some useful patterns in the data. To manage hidden patterns better in the data, future efforts should replace some base learners like Random Forest and Gradient Boosting with Support Vector Machines, K-Nearest Neighbors, or neural networks. If a nonlinear regression or deep learning model is put in the place of the current linear Ridge Regression, the ensemble could better connect and merge the predictions from different learning methods.

## **5.7 Comparison with Existing Literature**

The findings support the work done before in financial machine learning, revealing the success and failings of ensemble learning techniques on structured financial information. Additionally, the strong results observed when using boosting models like Gradient Boosting and XGBoost back up the previous studies of Dalal et al. (2022) and Bravo (2024), which discovered that this approach excels at finding nonlinear connections in financial datasets commonly used in prediction.

The poor results of the stacking strategy in this study back up what Hao & Zhang (2024) had already discussed (Seyedan et al., 2022). Their work showed that stacking makes better predictions when the base models are chosen carefully, there is variety among the models, and the main model for the whole ensemble is powerful enough. The findings here support the belief put forward by Hao & Zhang (2024) that stacking will fail if the base models are the same, if the

This observation about feature importance is also supported by the results in Padhi et al. (2021), showing that price indicators, like the highest and lowest intraday prices, are more valuable in forecasting financial data when compared Therefore, this supports the use of price range indicators in forecasting and selecting features in future research.

Lastly, the overfitting problem found with XGBoost also matches what Suprihadi et al. (2025) said, namely that when you use a strong learner like XGBoost, it can sometimes learn too much from the noise in financial data and it shouldn’t be used that way unless you add some regularization. The study’s results on XGBoost point to the need for precise validation and tuning when working with noisy financial data.

## **5.8 Limitations and Directions for Future Research**

While this study explores the use of ensemble learning techniques in financial forecasting, still, some challenges are present and should be acknowledged. In addition, these limitations help pinpoint the areas that still need to be explored in financial forecasting with ensemble methods.

To begin with, the analysis relied on the S&P 500 index, which does not explain the main trends of all financial assets, such as commodities, forex, and cryptocurrencies. The methods used in the study were not applied to commodities, forex, or cryptocurrencies, which often respond to other outside factors (Hazman et al., 2022). The researchers should consider using these techniques further to test more kinds of financial instruments and see how universal their results are.

Macroeconomic facts such as GDP growth, inflation, interest rates, and unemployment were not included in the analysis. They are what help shape how markets act and investor emotions. More interesting and accurate findings can be had by including GDP growth, inflation, interest rates, and unemployment figures in studies.

Future studies should run optimization to find out if the model’s performance will improve a lot by altering hyperparameters in both the base learners and the stacking techniques. Accordingly, studies should optimize models such as XGBoost and stacking ensembles with grid search, randomized search, or Bayesian optimization as these approaches could help make the model better.

While the model was a stacking network, it did not have enough diversity in its base models when compared to other more advanced or ensemble models (Danso et al., 2021). Using multiple base models, such as neural networks and SVMs, or combining models together might help to enhance the performance of the ensemble. By using more advanced meta-learners like nonlinear regressors and deep learning classifiers, the general model might even improve further.

## **5.9 Summary**

This chapter explains the meaning of its results for financial machine learning and for practical use. As the chapter shows, boosting and bagging are used to make useful finance forecasts, and their use is clear. Stacking models should be used with care as there is nothing simple about them. They should be developed, sanity checked, and tested thoroughly, before deciding to use them in real practice.

This research is important for talks on financial machine learning and for applying models in practical finance tasks. Yet, the results of advanced models such as XGBoost depend greatly on their proper validation. The outfit explored improving the interpretability and strength of models by comparing Gradient Boosting and Random Forest. According to this research, ensemble learning matters in finances and provides a base for more research into powerful forecasts.

# **Chapter 6: Conclusion**

## **6.1 Overview**

The primary goal of this research was to learn how ensemble methods (Random Forest, Gradient Boosting, XGBoost, and a Stacking Ensemble) fare in forecasting the S&P 500 index using information from previous markets. The study’s researchers wanted to know if using multiple models would be more accurate and general in futures predictions, given the rising complexity and disturbances in the financial sector.

The researchers went through a regular way, which meant getting the data ready and checking the results of the models (like RMSE, MAE, and R²) along with dividing the data into groups, so they could see what worked well and what didn't in the different models.Because of what they found out, learning about forecasting is much easier, and now people can actually put these methods to use. Understanding financial forecasting has become easier and people can now use what they know to make good predictions because of the results in the study.

## **6.2 Addressing Research Objectives**

1. **To explore the theoretical underpinnings and practical implementations of ensemble methods in financial forecasting.**
   * To do this, the researchers first looked at a lot of studies and then ran models for the main types of ensembles: The study included Random Forest, Gradient Boosting, and XGBoost as the main types of ensembles, and all of them were examined by the model.
2. **To compare the performance of ensemble methods with traditional or single ML models.**
   * While standard models were not tested in this study, the authors judged ensemble models based on the benchmarks present in literature. Evaluating them based on what the literature recommends, ensemble models outperformed the others when it came to prediction.
3. **To assess the applicability of ensemble models in predicting stock market indices.**
   * The use of the S&P 500 index as a target variable let us test our model in a good way. Results showed that these types of models work well in this area, especially Gradient Boosting and Random Forest.
4. **To identify key challenges, limitations, and opportunities in implementing ensemble forecasting models.**
   * Some were mentioned, including needing to tune hyperparameters, craft stacking architectures with care, and the need for financial applications to have models that are easy to interpret.

## **6.3 Key Findings**

Researchers discovered that using ensemble learning models was very effective in the field of financial forecasting. Among all the models, Gradient Boosting performed the best by continually improving accuracy and requiring the least training time. Consequently, Random Forest came next, giving good stability and making it possible to interpret the results by looking at the important features. Although XGBoost was accurate and efficient during training, it struggled to perform in line with the data from cross-validation because it overfit the model.

With just two base learners and a linear meta-model, the stacking approach did not meet the expectations during this study. The results suggested that for stacking to succeed, there must be diverse collection of base models and the meta-learner should integrate them well. Because only two base learners and a linear meta-model made up the stacking ensemble, it could not achieve good results.

Results were backed up by reviewing residue data, how important each feature was, and checking performance graphs, so the findings give us multiple points of view on model behavior. In particular, checking what features are important showed that Haris and Naush showed that 'High' and 'Low' prices were far more influential than 'Volume', something that matches with previous studies.

## **6.4 Theoretical and Practical Contributions**

From an academic point of view, the study shows that using an ensemble approach can improve a model’s performance by mixing the abilities of various learners. It shows that bagging lowers the variance and boosting helps fix the bias by learning one model at a time. It also points out that stacking might not work well unless it is set up with careful thought.

As a result, the study is beneficial for trading, risk management, and decision systems in financial organizations. Random Forest and Gradient Boosting can be used in financial forecasting because they can give clear predictions each time they get updated. XGBoost can quickly help us predict the future, but you need to use it properly so it does not make mistakes.

In addition, the study shows that it’s really important for models used in banking to be easy to understand, so Random Forest and Gradient Boosting are good choices because they show clearly what the most important factors are. Since Random Forest and Gradient Boosting show what was the biggest influence and what didn’t matter, they are easy to understand and fit the rules in the financial sector.

## **6.5 Limitations**

Despite doing the research well, the study still has some shortcomings that must be mentioned. First, only analyzing the S&P 500 meant that the study’s findings were not applicable to other financial markets, such as commodities, cryptocurrencies, or emerging economies that act differently. Furthermore, this study did not account for important economic factors, such as interest rates, inflation, and GDP growth. The use of these variables might have improved the ability of the models to explain the trends and prevented them from being surprised by changes in the economy. Furthermore, extensive hyperparameter optimization was missing in the research. Optimizing hyperparameters more carefully likely would have improved the performances of the XGBoost and stacking models. Moreover, the stacking model in the study followed a simple setup and only consisted of two basic models and a basic linear meta-learner. As the stacking stack in the research was weak, it failed to capture important changes compared to what is typically expected from this model.

## **6.6 Recommendations for Future Research**

Built on these findings, more research can focus on different approaches to enhance the use of ensemble learning in making financial forecasts. The dataset should include a wider range of financial instruments (commodities, currencies, and cryptocurrencies) to find out if ensemble models work well in unpredictable markets. Including macroeconomic variables, including inflation, GDP growth, interest rates, and sentiment in the data will ensure that predictions are useful for a variety of situations. Third, optimization of important settings, called hyperparameters, is required to give the most advantages to advanced models like XGBoost and stacking ensembles. Ideally, more advanced ensemble models should combine several base machines, such as support vector machines, neural networks, and decision trees, and use meta-learning to combine their predictions. Complex model predictions can be explained clearly by using SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations), which helps gain more trust and ensures adherence to the rules in finance.

## **6.7 Final Reflections**

In conclusion, this study has shown that using models like Gradient Boosting and Random Forest can make good, trustworthy predictions in finance, and that XGBoost can work fast and use less data, but you still have to be careful about overfitting. While there are other techniques that can potentially help, they work best only when you set them up and adjust them carefully.

The consistent performance of Gradient Boosting and Random Forest shows these models are useful and reliable for people who work with financial data analysis. Meanwhile, XGBoost runs quickly and can handle a lot of data, so it might help with real-time money forecasts, as long as you watch out for it learning the training data too perfectly and doing badly on new data.

This research adds to the expanding field of financial machine learning by showing that getting clear and accurate predictions is not just about using complex methods, but about finding the right balance between things like how well models fit the data and how helpful and easy to understand they are.

By embracing ensemble methods and getting better at how we use them, financial institutions and researchers will get closer to making systems that help them forecast better and make smarter choices, which will help them deal with today’s markets much more confidently and accurately.

# **References**

Abedin, M. Z., Guotai, C., Hajek, P., & Zhang, T. (2022). Combining weighted SMOTE with ensemble learning for the class-imbalanced prediction of small business credit risk. *Complex & Intelligent Systems*, *9*(4), 3559–3579. <https://doi.org/10.1007/s40747-021-00614-4>

Alsaedi, M., Ghaleb, F., Saeed, F., Ahmad, J., & Alasli, M. (2022). Cyber Threat Intelligence-Based Malicious URL Detection Model using ensemble learning. *Sensors*, *22*(9), 3373. <https://doi.org/10.3390/s22093373>

Angeline, R., Aarthi, S., Regin, R., & Rajest, S. S. (2023). Dynamic Intelligence-Driven Engineering Flooding Attack Prediction using ensemble Learning. In *Advances in computational intelligence and robotics book series* (pp. 109–124). <https://doi.org/10.4018/979-8-3693-1301-5.ch006>

Azeez, N. A., Odufuwa, O. E., Misra, S., Oluranti, J., & Damaševičius, R. (2021). Windows PE malware detection using Ensemble Learning. *Informatics*, *8*(1), 10. <https://doi.org/10.3390/informatics8010010>

Bhambu, A., Gao, R., & Suganthan, P. N. (2024). Recurrent ensemble random vector functional link neural network for financial time series forecasting. *Applied Soft Computing*, *161*, 111759. <https://doi.org/10.1016/j.asoc.2024.111759>

Bravo, J. M. (2024). Ensemble Methods for stock market Prediction. In *Communications in computer and information science* (pp. 430–448). <https://doi.org/10.1007/978-3-031-74643-7_31>

Chen, Y., Xu, Y., Jamhiri, B., Wang, L., & Li, T. (2022). Predicting uniaxial tensile strength of expansive soil with ensemble learning methods. *Computers and Geotechnics*, *150*, 104904. <https://doi.org/10.1016/j.compgeo.2022.104904>

Dalal, S., Seth, B., Radulescu, M., Secara, C., & Tolea, C. (2022). Predicting Fraud in Financial Payment Services through Optimized Hyper-Parameter-Tuned XGBoost Model. *Mathematics*, *10*(24), 4679. <https://doi.org/10.3390/math10244679>

Danso, S. O., Zeng, Z., Muniz-Terrera, G., & Ritchie, C. W. (2021). Developing an explainable Machine Learning-Based Personalised Dementia Risk Prediction model: a transfer learning approach with ensemble learning algorithms. *Frontiers in Big Data*, *4*. <https://doi.org/10.3389/fdata.2021.613047>

Deep, A. (2024). *Advanced financial market forecasting: integrating Monte Carlo simulations with ensemble Machine Learning models*. <https://ttu-ir.tdl.org/items/a92bbde6-176d-443a-90a2-a52f9c247821>

Hao, J., & Zhang, C. (2024). A stacking ensemble deep learning model for stock price forecasting. In *Lecture notes in computer science* (pp. 146–157). <https://doi.org/10.1007/978-981-97-5663-6_13>

Hazman, C., Guezzaz, A., Benkirane, S., & Azrour, M. (2022). lIDS-SIoEL: intrusion detection framework for IoT-based smart environments security using ensemble learning. *Cluster Computing*, *26*(6), 4069–4083. <https://doi.org/10.1007/s10586-022-03810-0>

He, K., Yang, Q., Ji, L., Pan, J., & Zou, Y. (2023). Financial Time Series Forecasting with the Deep Learning Ensemble Model. *Mathematics*, *11*(4), 1054. <https://doi.org/10.3390/math11041054>

He, Q. Q., Pang, P. C. I., & Si, Y. W. (2020, December). Multi-source transfer learning with ensemble for financial time series forecasting. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 227-233). IEEE. <https://ieeexplore.ieee.org/abstract/document/9457690>

Hoy, Z. X., Woon, K. S., Chin, W. C., Hashim, H., & Van Fan, Y. (2022). Forecasting heterogeneous municipal solid waste generation via Bayesian-optimised neural network with ensemble learning for improved generalisation. *Computers & Chemical Engineering*, *166*, 107946. <https://doi.org/10.1016/j.compchemeng.2022.107946>

Lee, J., Wang, W., Harrou, F., & Sun, Y. (2020). Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. *Energy Conversion and Management*, *208*, 112582. <https://doi.org/10.1016/j.enconman.2020.112582>

Li, Y., & Pan, Y. (2021). A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*, *13*(2), 139–149. <https://doi.org/10.1007/s41060-021-00279-9>

Li, Z., & Tam, V. (2024, June). Developing An Attention-Based Ensemble Learning Framework for Financial Portfolio Optimisation. In *2024 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE. <https://ieeexplore.ieee.org/abstract/document/10650941>

Liu, W., Suzuki, Y., & Du, S. (2025). Ensemble learning algorithms based on easyensemble sampling for financial distress prediction. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-025-06494-y>

Lyu, Z., Ma, F., & Zhang, J. (2023). Oil futures volatility prediction: Bagging or combination? *International Review of Economics & Finance*, *87*, 457–467. <https://doi.org/10.1016/j.iref.2023.05.007>

Meira, E., Oliveira, F. L. C., & De Menezes, L. M. (2022). Forecasting natural gas consumption using Bagging and modified regularization techniques. *Energy Economics*, *106*, 105760. <https://doi.org/10.1016/j.eneco.2021.105760>

Miao, J., & Polak, P. (2023, March 30). *Online Ensemble of Models for Optimal Predictive Performance with Applications to Sector Rotation Strategy*. arXiv.org. <https://arxiv.org/abs/2304.09947>

Muslim, M. A., Nikmah, T. L., Pertiwi, D. a. A., Subhan, N., Jumanto, N., Dasril, Y., & Iswanto, N. (2023). New model combination meta-learner to improve accuracy prediction P2P lending with stacking ensemble learning. *Intelligent Systems With Applications*, *18*, 200204. <https://doi.org/10.1016/j.iswa.2023.200204>

Olorunnimbe, K., & Viktor, H. (2024). Ensemble of temporal Transformers for financial time series. *Journal of Intelligent Information Systems*, *62*(4), 1087–1111. <https://doi.org/10.1007/s10844-024-00851-2>

Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Ijaz, M. F. (2021). A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators. *Mathematics*, *9*(21), 2646. <https://doi.org/10.3390/math9212646>

Pham, B. T., Jaafari, A., Van Phong, T., Yen, H. P. H., Tuyen, T. T., Van Luong, V., Nguyen, H. D., Van Le, H., & Foong, L. K. (2020). Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. *Geoscience Frontiers*, *12*(3), 101105. <https://doi.org/10.1016/j.gsf.2020.11.003>

Rubai, S. M. (2023). Development of hyper‐parameter‐tuned‐recurrent neural network for detection and mitigation of fraudulent resource consumption attack in cloud. *Transactions on Emerging Telecommunications Technologies*, *34*(3). <https://doi.org/10.1002/ett.4723>

Seyedan, M., Mafakheri, F., & Wang, C. (2022). Cluster-based demand forecasting using Bayesian model averaging: An ensemble learning approach. *Decision Analytics Journal*, *3*, 100033. <https://doi.org/10.1016/j.dajour.2022.100033>

Shorewala, V. (2021). Early detection of coronary heart disease using ensemble techniques. *Informatics in Medicine Unlocked*, *26*, 100655. <https://doi.org/10.1016/j.imu.2021.100655>

Suprihadi, E., Danila, N., & Ali, Z. (2025). Enhancing Financial Product Forecasting Accuracy Using EMD and Feature Selection with Ensemble Models. *Journal of Open Innovation Technology Market and Complexity*, 100531. <https://doi.org/10.1016/j.joitmc.2025.100531>

Tsai, C., Sue, K., Hu, Y., & Chiu, A. (2021). Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *Journal of Business Research*, *130*, 200–209. <https://doi.org/10.1016/j.jbusres.2021.03.018>

VA, B., Subramoniam, M., & Mathew, L. (2021). Detection of COPD and Lung Cancer with electronic nose using ensemble learning methods. *Clinica Chimica Acta*, *523*, 231–238. <https://doi.org/10.1016/j.cca.2021.10.005>

Wu, H., & Levinson, D. (2021). The ensemble approach to forecasting: A review and synthesis. *Transportation Research Part C Emerging Technologies*, *132*, 103357. <https://doi.org/10.1016/j.trc.2021.103357>

Yasper, A., Handoko, D., Putra, M., Aliwarga, H. K., & Rosid, M. S. R. (2023). Hyperparameters optimization in XGBOOST model for rainfall estimation: a case study in Pontianak City. *Jurnal Penelitian Pendidikan IPA*, *9*(9), 7113–7121. <https://doi.org/10.29303/jppipa.v9i9.3890>

Zhang, W., Li, H., Han, L., Chen, L., & Wang, L. (2022). Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China. *Journal of Rock Mechanics and Geotechnical Engineering*, *14*(4), 1089–1099. <https://doi.org/10.1016/j.jrmge.2021.12.011>

Zhao, Z., Sun, S., Sun, J., & Wang, S. (2023). A novel hybrid model with two-layer multivariate decomposition for crude oil price forecasting. *Energy*, *288*, 129740. <https://doi.org/10.1016/j.energy.2023.129740>