Vrije Universiteit Amsterdam PricewaterhouseCoopers

Master Thesis

Predictive Modeling in Stock Market Price Forecasting: Challenging the Efficient Market Hypothesis

Author: Sander Genot (2669350)

1st supervisor: René Bekker
daily supervisor: Jerom de Valk (company)
2nd reader: Kevin Luck

August 28, 2024

Abstract

Predicting stock market prices is challenging due to market unpredictability, with traditional theories like the efficient market hypothesis (EMH) suggesting that prices already reflect all available information. This study aims to enhance prediction accuracy for the top 25 S&P 500 stocks by developing multiple models that integrate historical data, technical indicators, fundamental data, industry trends, macroeconomic factors, and social media sentiment. Machine learning models including LSTM, SVM, Random Forests, and Linear Regression were used to analyze these diverse data sources. In addition, multiple different NLP transformer models, different versions of BERT, were used to perform sentiment analysis on sentiment data. Results showed that simpler models performed well with basic data, while adding sentiment analysis improved some predictions but also introduced noise. Moreover, the transformer models pre-trained on financial corpus outperformed those pre-trained on standard text. Frequent rebalancing strategies outperformed sentiment-based approaches. The study concludes that integrating diverse data can improve predictions, but model simplicity and careful data selection are crucial for success.

Contents

1

Introduction

Predicting stock market prices remains a formidable challenge. This is exacerbated by the unpredictable nature of financial markets. In other words, stock markets are dynamic, non-linear, non-stationary, non-parametric, noisy, and chaotic. This makes analyzing price behavior and movements in the market quite challenging. Theories like the Efficient Market Hypothesis (EMH) and the Random Walk Theory (RWT) further emphasize this intricacy. In the RWT, it is assumed that stock prices move in a manner similar to a random walk. According to the EMH, which was presented by (1) and (2), stock prices take into account all relevant information and only the change in reaction to new information. This implies that utilizing knowledge that is readily available to the public, it is theoretically impossible to consistently outperform the market.

Criticism of the EMH has led to an increasing number of studies questioning its validity and introducing new and successful approaches that combine technical analysis indicators and chart patterns, patterns within a chart when prices are graphed, with methods from econometrics, statistics, data mining, and artificial intelligence (3). Therefore, despite the previously mentioned drawbacks, a new wave of research aimed at improving the precision of stock market forecasts has been driven by the development of sophisticated computer tools and a wide range of data sources. In order to develop a more thorough understanding of market dynamics, there has recently been a growing trend to integrate several data sets, such as mood analysis, technical indicators, historical pricing, and industry trends.

In particular, sentiment analysis has become increasingly important in highlighting the emotional and psychological aspects of the market. Rich information from social media sites, like Twitter, is now easier to acquire because of the digital era, and these datasets offer unique insights into how the public feels about certain stocks or the market as a whole.

1. INTRODUCTION

The big data paradigm, which seeks to increase the precision of forecasting models, is consistent with this integration of new data streams with conventional financial indicators.

Sentiment can influence short-term market volatility, leading to discrepancies between market value and the book value of a firm. As Nobel Prize winner Robert Shiller demonstrated, however, fundamental reasons eventually push the share price to represent the underlying value of the business (4). On the one hand, the predictive value of sentiment research has been questioned even though it has made it simpler to include the emotions of the market in forecasts regarding stock movement and price. According to research such as (5) and (6), sentiment on social media may not be indicative. On the other hand, research by (7) and (8) has shown that including market sentiment can improve the accuracy of forecasting models, highlighting the significant influence of market sentiment on stock movements.

Moreover, the potential of combining technical indicators and market sentiment with fundamental data and sector-specific trends offers a promising avenue of research. When combined, these components offer a comprehensive method for comprehending and projecting market behavior that goes beyond the constraints usually connected with stock market forecasting. Using these many information sources could lead to new ways of interpreting the intricacies of the stock market, giving analysts and investors alike a more sophisticated and useful toolkit as computational finance develops.

The primary objective is to improve the forecasting accuracy of the stock market by performing a thorough, data-driven analysis of the top 25 large-cap stocks in the S&P 500. In other words, the ultimate goal of this research is to develop a model that can accurately forecast changes in stock prices. The method incorporates information from multiple sources including historical price data, technical indicators, fundamental data, industry trends, macroeconomic variables, and sentiment data from social media sites like Twitter and financial news.

The secondary objective is to test the validity of traditional financial theories, such as the EMH and RW, by incorporating these various data sources. These objectives converged in the research question: "How does incorporating multiple data sources, different Machine-/Deep Learning techniques, and sentiment analysis with Natural Language Processing enhance the accuracy of stock price predictions?"

This study uses a variety of benchmark models in addition to modern methods including time series analysis with LSTM networks and NLP tasks with bidirectional encoder representations from transformers. Other machine learning models, such as SVR, RF, and LR, are also investigated for their robustness in regression tasks relevant to stock price prediction. These models present a variety of viewpoints on the data, each with a unique ability to capture various facets of market activity.

Advanced NLP techniques are employed to measure stock sentiment. Specifically, optimized versions of BERT (Bidirectional Encoder Representations from Transformers) are leveraged. This makes it possible to analyze sentiment in great detail that is present in large amounts of textual data from social media and financial news, giving important insights into public opinion and how it may affect stock prices.

More specifically, to determine which BERT model is best suited for financial sentiment analysis, the sentiment analysis will start by comparing it to carefully selected datasets like the Financial Phrase Bank. The next stage involves using zero-shot learning to classify the sentiment of tweets connected to stocks by choosing the best-performing BERT model. This makes it possible to handle tweets that are extracted from the hugging-face dataset "twitter-financial-news-sentiment".

The paper is structured as follows. Following the introduction, Section 2 delves into prior studies that form the basis of this research. In Section 3, the methods for gathering data and preparing it for training and testing, including feature engineering, are described. The methodology in section 4 describes the problem in more detail and delves into the stock price and movement prediction models and sentiment analysis models employed together with their evaluation metrics. The findings are shown in Section 5, which also includes a comparison of the baseline (naive) model. In Section 6, the conclusion provides a concise summary of the research by revisiting the main objectives and methods used. It highlights the key findings and offers an interpretation of the results in the context of the research questions. The paper concludes in Section 7 with a discussion of the findings, implications for future research, and finally a reference list.

2

Background & Literature

The first section, *Traditional Finance*, delves into cornerstone theories such as the Random Walk Theory and the Efficient Market Hypothesis, which have traditionally dictated the understanding of market behaviors. This section also explores how these theories reconcile with empirical market behaviors, setting the stage for discussing their limitations and the circumstances under which they may not hold.

In the second section, *Challenging the Efficient Market Hypothesis*, alternative approaches are explored including fundamental and technical analysis. This part of the chapter highlights the limitations of traditional theories in capturing the complexities of real-world markets and discusses how advancements in financial technologies and methodologies have led to improved predictive capabilities. Specific attention is given to innovations in market forecasting that leverage computational and quantitative models, reflecting the shift towards more data-driven approaches in finance.

The third section, *Sentiment Analysis in Financial Markets*, assesses the role of investor sentiment and its quantification through advanced NLP techniques. It covers the integration of sentiment analysis into financial prediction models, detailing the impact of emerging technologies such as BERT and its financial derivatives on stock price prediction. This section not only highlights the evolution of sentiment analysis, but also discusses its practical implications and effectiveness in enhancing market forecasts.

Through a detailed exploration of these areas, the chapter aims to provide a comprehensive background, preparing the reader for a deeper investigation of modern financial market analytics in the following chapters.

2.1 Traditional Finance

The exploration of financial markets has long been dominated by theories aiming to understand and predict stock prices and movements. Among the cornerstone theories in this domain are the RWT (9) and the EMH (1, 2), each offering unique perspectives on the nature of stock price changes and the efficiency of markets.

2.1.1 Random Walk Theory

According to the RWT, which was covered in-depth by (9), price fluctuations in stocks follow a pattern akin to a random walk and are therefore both unpredictable and random. Essentially this theory challenges the viability of consistently obtaining returns higher than the market average by market timing or stock selection strategies. It claims that attempting to predict future stock values using historical price movements is futile. This theory is predicated on the notion that stock prices are meaningless forecasts since they are just the product of a multitude of random occurrences and information coming together. In other words, the random walk theory, as outlined by (10), suggests that the market price of a particular stock should remain independent of its previous price. Early evidence for this idea came from empirical research by (10, 11, 12), which demonstrated the random walk nature of stock price fluctuations and hence put conventional stock forecasting techniques to the test.

A common mathematical representation of a random walk for stock prices is given by the following equation:

$$S_{t+1} = S_t + \epsilon_t \tag{2.1}$$

where:

- S_t is the stock price at time t.
- S_{t+1} is the stock price at time $t+1$.
- ϵ_t is a random variable representing the change in stock price, often modeled as a normal distribution with mean zero and some variance σ^2.

2.1.2 Efficient Market Hypothesis

Conversely, the EMH, outlined by (1) and (2), states that a stock's price mirrors all existing information, granting equal information access to all market participants. According to the

EMH, this informational efficiency makes it impossible for investors to achieve consistently higher returns through either technical or fundamental analysis, as stock prices adjust so rapidly to new information that any attempt to trade on it is likely futile.

The EMH differentiates into three forms: **1)** The weak form, which negates the utility of technical analysis by asserting that past prices are already reflected in current prices. **2)** The semi-strong form, which asserts that neither technical nor fundamental analysis can offer an investor edge since all public information is accounted for in stock prices. **3)** The strong form, which claims that all information, public or otherwise, is factored into current stock prices, leaving no room for informational advantages in the market.

According to the EMH, markets are efficient because investors make logical decisions and promptly adjust prices when they see an opportunity to benefit. The empirical challenges to the EMH have been met with serious scrutiny despite its broad acceptance. Evidence of deviations from ideal market efficiency can be found in documented market oddities like the weekend effect (13), which notes lower returns on Mondays, and the January effect (14), which notes that stocks typically perform better in January.

Moreover, studies on behavioral finance have revealed patterns of investor irrationality. An expansion of conventional finance, behavioral finance, examines the psychological factors and biases influencing investor choices and market performance. This area of study recognizes that investors are not always logical and are frequently swayed by their own prejudices and emotions, which results in predictable but frequently poor financial judgments. Behavioral finance provides answers for a range of market anomalies, that are not fully explained by conventional financial theories such as the EMH. These anomalies include tendencies toward overreaction (15) and the disposition effect (16, 17, 18), which contradicts the rational investor model assumed by the EMH. These anomalies further challenge the notion of market efficiency and rationality posited by the EMH.

Because behavioral finance incorporates psychological aspects into the examination of investor behavior and market dynamics, it has made a substantial contribution to our understanding of financial markets. It offers a more sophisticated comprehension of financial decision-making by acknowledging the substantial influence that emotions and cognitive biases have on investor behavior and, in turn, on the results of markets. This viewpoint adds a great deal to the field of finance research and offers regulators, investors, and portfolio managers wise counsel for navigating the intricacies of the financial system.

2.1.3 Reconciling Theory with Practice

Financial regulations and investment strategies have been significantly impacted by the application of the RWT and the EMH in the actual world. This conventional wisdom, however, needs to be reassessed given the persistence of market anomalies and the expanding corpus of data about investor behavior. In (19), the authors introduced the Adaptive Market Hypothesis as a framework for combining behavioral finance and the EMH. The authors make the argument that shifting investor behavior and market conditions have an impact on market efficiency, which is a dynamic process.

Furthermore, the strict interpretations of the RWT and EMH have been challenged by new directions in market analysis demonstrated by developments in data analytics and financial technology. The emergence of advanced algorithms and machine learning models suggests possible avenues for detecting subtle patterns and connections in financial markets, suggesting a more intricate comprehension of market dynamics that goes beyond the dividing line of predictability and randomness.

2.2 Challenging the Efficient Market Hypothesis

The EMH is contested in the investment sector by a variety of analytical approaches and sophisticated forecasting methods that go against the hypothesis's claim of market efficiency.

2.2.1 Fundamental and Technical Analysis

The two primary analysis approaches used in the investment industry are technical analysis and fundamental analysis.

Fundamental Analysis Fundamental analysis, which considers a company's intrinsic value to uncover potential investment opportunities, is a crucial part of investing strategy. With this method, analysts look at a range of financial data, including cash flow, balance, and income statements, to determine the stability and health of a company's finances. In order to understand how the macroeconomic environment affects the performance of the company, other economic data are also considered, such as rates of inflation and bond prices.

The fundamental analysis aims to find and utilize these qualitative elements, such as market trends and industry position. Fundamental analysis is criticized for being time-consuming and vulnerable to analysts' prejudices when analyzing economic and financial data. Moreover, in highly efficient markets, it is believed that all known information is already reflected in stock prices, potentially diminishing the effectiveness of this analysis.

Technical Analysis Technical analysis, on the other hand, is based on the notion that changes in the price of the stock market and trade volume might indicate future price trends. This approach makes extensive use of charts and patterns to pinpoint possible buying or selling opportunities, including oscillators, trend lines, moving averages, and candlestick patterns. Technical analysts operate under the assumption that behavioral finance plays a substantial role in decision making. In other words, market psychology influences trading in a way that allows the prediction of when a stock will rise or fall based on past patterns.

The idea that history repeats itself and that patterns in stock price movements can be examined and forecasted is a fundamental principle of technical analysis. Traders who prefer to make short-term investments over long-term ones tend to favor this strategy the most. Technical analysis, however, is viewed with suspicion due to its dependence on subjective and interpretable chart patterns and indications. In addition, critics point out that relying solely on historical data without taking into account the fundamentals of a company may ignore larger changes in the market or the economy that could have an impact on stock prices.

Integration of Fundamental and Technical Analysis Although technical and fundamental analysis is often seen as mutually exclusive approaches, some analysts and investors support a more integrated strategy to make use of the advantages of both approaches. Investors may be able to improve their investment strategy by combining the long-term outlook of the fundamental analysis with the accuracy of the technical analysis timing (20). This would enable them to make informed decisions about the optimal timing for buying or selling by thoroughly understanding market conditions and the value of the company.

This integrated approach recognizes that market prices can be impacted by a complex interaction of basic reasons, investor sentiment, and previous trading patterns, allowing for a comprehensive perspective of investing opportunities (21). Combining these approaches could provide a more flexible and sophisticated approach to managing the risks associated with stock investing as the financial markets continue to change (22).

2.2.2 Advancements in Market Prediction Techniques

As described in (23), advances in stock market analysis and prediction techniques can be categorized into four primary approaches: statistical methods, pattern recognition, machine learning (ML), and sentiment analysis. These approaches fall predominantly under the broader umbrella of technical analysis, with certain machine learning techniques also bridging the gap to include fundamental analysis for a more comprehensive market prediction strategy.

Before the incorporation of machine learning into financial analysis, stock price prediction relied heavily on statistical techniques. These conventional methods established the foundation for a methodical approach to comprehending market dynamics by frequently supposing linearity, stationarity, and normality. One of the best examples of these statistical techniques is time series analysis, which allows analysts to monitor and forecast changes in stock prices over time by arranging sales data and stock prices in a chronological order.

The Auto-Regressive Moving Average (ARMA) and its variant, the Auto-Regressive Integrated Moving Average (ARIMA), are pivotal in this realm, offering models that capture autocorrelations within time series data (24). Similarly, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model addresses volatility clustering, a common feature in financial time series, by modeling the variance of the current error term as a function of the variances of the error terms of previous time periods (25), (26).

Pattern recognition focuses on finding recurrent themes in stock data, especially in Open-High-Low-Close (OHLC) candlestick charts. By offering visual clues for possible market moves, this practice not only enhances technical analysis, but also presents a data-driven methodology to identify trends, reversals, and continuation patterns.

Research in this domain has used various computational techniques to establish pattern detection techniques, ranging from Bayesian networks, which offer a probabilistic approach to modeling uncertainties in market trends (27), to more complex models such as neural networks and machine learning algorithms that adapt and learn from the data (28).

Studies investigating the predictive capacity of combining multiple statistical methods have further enhanced the field of financial analysis. For example, combining GARCH and ARIMA models has been shown to be successful in forecasting heteroskedasticity, capturing the dynamics of mean and volatility in financial time series (29). Furthermore, new pathways to improve prediction accuracy and investment strategies have been made possible by applying machine learning techniques, such as SVM and RF, to the patterns found through statistical analysis (30).

Despite these advances, the challenges of modeling financial markets remain. Critiques regarding the small sample sizes in some studies (8, 31) highlight the importance of robust data sets and comprehensive analysis. In order to overcome these challenges, the current study examines 25 stocks for a period of 10 years. The goal is to offer more comprehensive and dependable insights into stock price fluctuations and to advance the rapidly developing field of financial analytics.

Innovations in Financial Market Forecasting Machine learning has become a cornerstone in the realm of financial market predictions, with both supervised and unsupervised learning methods offering innovative strategies to forecast market movements. In recent years, different ML techniques, such as supervised, unsupervised, and reinforcement learning, have been used effectively to solve different multidisciplinary problems in real life (32).

Supervised learning facilitates the prediction of future stock prices by leveraging historical data and algorithms that are trained on datasets with labeled input-output pairs. This method operates on the principle that past stock price patterns can be used to forecast future trends. Numerous studies, such as (33), that offer a thorough overview of pattern recognition and ML techniques applied in different domains, including finance, have shown the effectiveness of supervised learning in financial forecasting.

Conversely, unsupervised learning investigates unlabeled data in order to find latent structures or patterns without the need for pre-established labels or categories. This method works especially well for finding unusual or new patterns in the financial markets that are not always obvious. Unsupervised learning relies heavily on methods like clustering and dimensionality reduction to reveal intricate links in financial datasets. The authors of (34) discuss in more detail the possibilities of unsupervised learning in financial analysis, including the creation of deep learning architectures that may extract valuable data from large, unlabeled datasets.

Because technical analysis is widely accepted by financial advisors and technical data is so readily available, machine learning is becoming the preferred method of stock market forecasting. This preference is supported by the fact that technical indicators and stock prices are updated daily, providing a rich dataset for analysis. The significance of this transition is highlighted in (35), which examines the use of technical analysis in financial markets, providing a foundational understanding of how historical price and volume data can inform future market movements.

Attempts to predict financial trends have also explored SVM for daily stock price direction (36) and weekly movements on indices like NIKKEI 225 (37). The authors of (38) introduced a fusion model that combines the hidden Markov model (HMM), the artificial neural network (ANN) and the genetic algorithm (GA) to predict market behavior, demonstrating the potential of hybrid approaches.

The use of Artificial Neural Networks, Support vector machines, and random forest have turned out to be pivotal tools, leveraging historical prices and technical indicators as fundamental inputs, has been explored extensively (28, 39, 40, 41, 42). The success achieved through these studies highlights the versatility and potential of machine learning in navigating the complexities of finance.

Deep learning (DL) leverages multi-layered networks and has achieved notable advancements in managing and interpreting massive amounts of data. Two DL techniques that are particularly good at identifying patterns in data are convolutional neural networks (CNNs) and long-short-term memory (LSTM) networks (43, 44). These algorithms can also detect complex and nonlinear patterns in financial time series. (45) and (46) provide a thorough examination of DL techniques and their applications in a variety of fields, including finance, and address how deep learning affects financial forecasting.

Furthermore, studies have embraced ensemble approaches and Recurrent Neural Network (RNN) methodologies in addition to conventional machine learning techniques. Research such as (47) has shown how these hybrid models, especially those that use RNN, work well enough to outperform traditional forecasting models in terms of accuracy. The creation of a two-stage fusion model by (30) provides additional evidence of the noteworthy progress made in stock price prediction and supports the idea that multi-stage techniques are preferable to single-stage models.

LSTM networks were used by (48) to forecast stock returns in the Chinese stock market. According to the study, LSTM models are better at capturing the long-term dependencies of stock price sequences, resulting in more accurate forecasts. The research contributes to the increasing body of data demonstrating LSTM networks' efficacy.

The problem of using noisy and chaotic news data for stock prediction is discussed in (49). The authors significantly improve stock price prediction by creating a DL framework that gathers and filters pertinent data from financial news. Their study serves as an example of how important model architecture and data preprocessing are when using text data for financial forecasting.

Furthermore, (50) presented a new DL-based data augmentation technique to improve the robustness of the model and avoid overfitting. By using LSTM layers, this method enhances model performance and highlights the significance of incorporating cutting-edge computational approaches into stock market analysis. It is intended for financial time series forecasting.

A DL model that rates stocks according to their potential returns was proposed by (51). Using a temporal relational ranking model, the study provides a more reliable prediction mechanism than previous methods by capturing relational interdependence between stocks throughout time. This approach emphasizes how relational and temporal models might improve stock market predictions.

2.3 Sentiment Analysis in Financial Markets

Sentiment analysis is the task of extracting sentiments or opinions of people from the written language (52). Sentiment analysis has emerged as a key tool in various applications, from evaluating product and restaurant reviews (53, 54) to analyzing the nuanced language of financial markets. In the realm of general sentiment analysis, the objective is straightforward: to discern consumer emotions and opinions about products, services, or experiences. For example, a product review stating, "I absolutely love this phone; its battery life is incredible," is clearly positive. Such direct expressions allow sentiment analysis tools to easily categorize feedback as positive, leveraging common indicators of satisfaction.

2.3.1 Application to Financial Markets

However, when sentiment research is used in the financial industry, the circumstances are different. This time, the emphasis shifts to analyzing the sentiment found in financial news, analyst reports, earnings calls, and financial statements. The intricacies of articulating financial outcomes, expectations, and market patterns are navigated by financial sentiment analysis. A statement like "The company's operating margin is under considerable pressure due to increased raw material costs," though not overtly negative, signals a potential concern for investors, indicative of a negative financial outlook.

The complexity of financial language lies in its specialized terminology, which often carries different connotations than in everyday speech. For instance, the term "exposure," typically neutral, assumes a risk-related meaning in a financial context, such as in "exposure to foreign markets."

Impact on Stock Price Prediction The utility of sentiment analysis in predicting stock prices, especially through the analysis of social media and financial news sentiment, has increasingly been recognized. Incorporating sentiment data alongside historical price information has been shown to significantly enhance prediction accuracy, posing a challenge to traditional financial theories like the random walk theory.

Efforts to adapt sentiment analysis to financial contexts have explored various textual representations, including bag of words, noun phrases, and named entities, integrating these with predictive models like linear regression and SVM (7). Nonetheless, these methods often fall short in capturing the mood information crucial for financial analysis.

Alternatively, (55) approached this by quantifying collective emotions such as hope and fear, examining their correlation with stock market indicators through mood-tagged tweets. This shift towards a more nuanced understanding of sentiment in financial markets underscores the evolving nature of sentiment analysis, bridging the gap between generic sentiment evaluation and its application in financial forecasting.

2.3.2 Evolution of Sentiment Analysis Techniques

Recent efforts in sentiment analysis can be broadly categorized into two approaches: traditional machine learning methods that rely on text features extracted through techniques like word counting, and DL methods that represent text through sequences of embeddings. While traditional methods struggle to capture the semantic nuances conveyed by specific word sequences, DL approaches offer a more nuanced analysis but require a substantial amount of data to learn effectively (56, 57, 58).

The authors of (57) stand out for applying machine learning to the study of financial language by evaluating the sentiment of financial documents using lexicon-based techniques and a "bag-of-words" approach. Similarly, (58) used supervised machine learning techniques to identify sentiments about financial institutions by analyzing n-grams from tweets that contained financial information.

In (59), sentiment analysis was leveraged on Yahoo Message Board comments for stock price prediction, integrating various NLP techniques to derive sentiment, which, along with historical price data, served as input to an SVM for trend forecasting. This study also explored additional classification models, such as LDA topic, JST-based and Aspect-based models.

The authors of (60) introduced a hybrid approach, merging LSTM with investor sentiment analysis for the Chinese stock market prediction. Essential to this process is text preprocessing, notably in Chinese, involving text segmentation, stop word removal, and conversion of text to vector representations via Word2vec, a tool that employs continuous bag-of-words (CBOW) and Skip-gram models.

One of the pioneering studies to apply DL for financial sentiment analysis was conducted by (56). They demonstrated that LSTM neural networks, applied to company announcements, could predict stock market movements more accurately than traditional machine learning models. Their findings also highlighted the benefits of pretraining models in larger corpora to enhance accuracy.

Further investigations have underscored the efficacy of LSTM-based models when combined with sentiment analysis. In (61) investor sentiment was extracted from forum posts, integrating it with historical market data within a network to forecast CSI300 and sentiment, noting that LSTMs surpassed SVM benchmarks, with sentiment features notably enhancing next day open price predictions.

Similarly, (62) utilized textual data from newspapers and numerical time-series data within LSTMs to predict the open prices for ten companies, achieving significantly higher profits compared to models relying solely on numerical data. Further studies have explored various neural network architectures for financial sentiment analysis. The authors of (63) found CNNs to be the most effective for analyzing sentiment in the StockTwits dataset. The authors of (64, 65) employed doc2vec and LSTM networks, respectively, achieving state-of-the-art results in classifying financial news sentiment.

Advancements in NLP: The Emergence of BERT and Its Financial Derivatives

In 2018, Google introduced BERT, which revolutionized the field of NLP. Rather than processing words one at a time in order, this ground-breaking model makes use of the Transformer architecture, a DL model that processes words in connection with all other words in a sentence using self-attention mechanisms. The novel aspect of BERT is its capacity to comprehend a word's context by taking into account both the word's left and right surroundings.

The model is pre-trained on a vast corpus of unlabeled text, including the entire Wikipedia and the BooksCorpus. It employs two novel strategies: masked language modeling (MLM) and next-sentence prediction. As described in (66), MLM is the process within BERT that randomly masks words in the input and then attempts to predict them based on the context provided by the non-masked words in the sequence. This approach allows BERT to learn a rich understanding of language, including word relationships and sentence structure.

Transfer Learning and BERT BERT's usage of transfer learning is one of its main characteristics. BERT can be adjusted with extra output layers after it has been pre-trained, which allows it to be flexible enough to handle a variety of NLP jobs without requiring significant changes to the model architecture. With comparatively little additional training, researchers and practitioners can use this capability to make use of BERT's profound comprehension of linguistic nuances for certain applications, such as sentiment analysis, question answering, and language inference.

After (66) introduced BERT, the NLP community set out to investigate, optimize, and expand BERT's capabilities. This trip resulted in the development of numerous noteworthy variants and enhancements that are suitable for a variety of applications.

RoBERTa: A Robustly Optimized BERT Approach One of the first notable developments was the announcement of RoBERTa (Robustly optimized BERT technique) by (67). By training the model on larger mini-batches, over longer periods of time, with more data, and without the next sentence prediction aim, RoBERTa modifies BERT's pre-training procedure. Additionally, it modifies the training data's masking pattern dynamically. These adjustments allowed RoBERTa to outperform multiple benchmarks, setting new standards for model effectiveness and efficiency in NLP tasks.

DistilBERT: A Distilled Version of BERT Recognizing the need for more computationally efficient models without compromising performance, (68) introduced DistilBERT in 2019. DistilBERT applies knowledge distillation techniques during the pre-training phase, effectively reducing the size of the BERT model while retaining 97% of its language understanding capabilities and speeding up its performance. This distilled version opened the door for deploying state-of-the-art NLP models in environments with limited computational resources.

FinBERT: Tailoring BERT for Finance Recognizing BERT's potential, (69) developed FinBERT, a variant of BERT specifically trained on financial texts. FinBERT was pre-trained on a large financial corpus, including corporate reports, Earning Call Transcripts, and financial news articles, to grasp the unique jargon and expressions used in the financial domain. This specialized training enables FinBERT to outperform its generic counterpart in financial sentiment analysis and other finance-related NLP tasks, offering more accurate predictions and insights.

FinancialBERT: Advancing Financial NLP Further building on BERT's foundation, (70) introduced FinancialBERT, another domain-specific adaptation pre-trained on an even wider array of financial documents. The training of this model included diverse sources, such as Bloomberg News and SEC filings, to capture the breadth of language used in the financial sector. FinancialBERT demonstrated significant performance improvements over both the original BERT and FinBERT, showcasing its ability to accurately interpret complex financial narratives and sentiments.

Impact on Financial Sentiment Analysis The development of BERT and its financial derivatives represents a significant leap forward in the application of NLP to the financial sector. By understanding the context and subtleties of financial language, these models offer unprecedented precision in sentiment analysis, allowing analysts to gain deeper insights from financial texts. Their success underscores the potential of advanced NLP technologies to transform financial analysis, risk assessment, and decision-making processes by providing more nuanced and sophisticated tools for interpreting market sentiments.

3

Data

This section provides an outline of the data used in this research, its sources, and the methods used for its collection, processing, and analysis. The types of data sets involved will be discussed in addition to their relevance to the study and any limitations they present. Figure 3.1 shows an overview of all the different data sets involved and highlights some features that are included in these datasets such as volume, moving averages, and ESG scores.

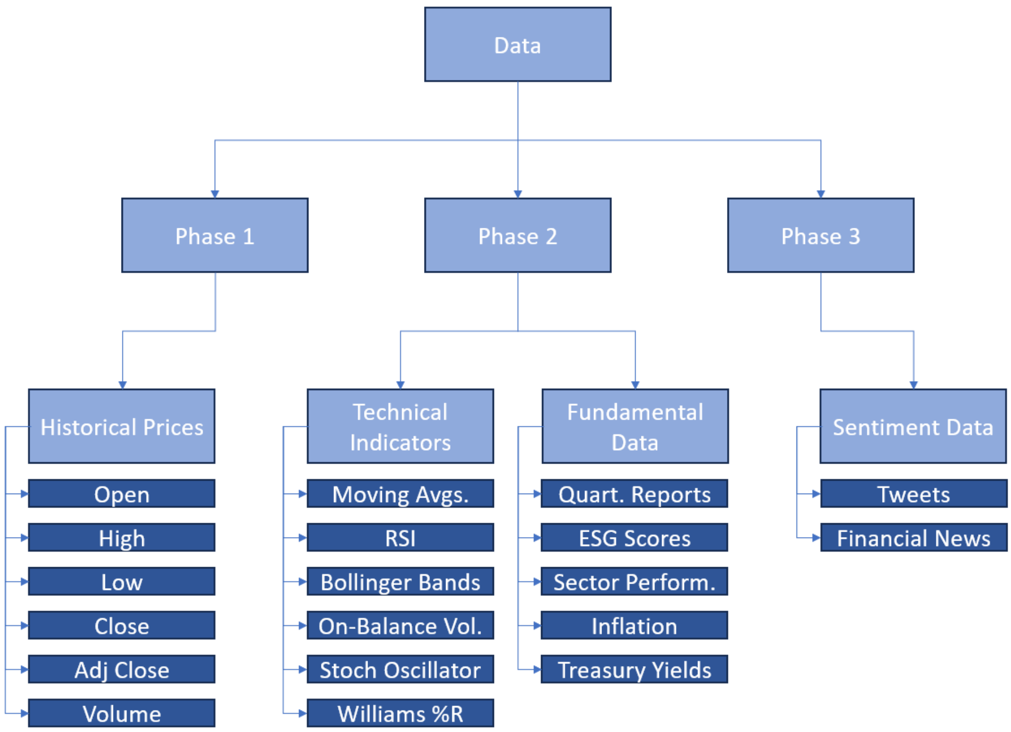

Figure 3.1: Data Source Overview

3.1 Data Collection

For the prediction of stock price movements, this methodology will employ an integrated framework comprising three comprehensive datasets. The primary dataset encompasses historical stock price data, providing a foundation for analysis of past market performance. This dataset includes detailed records of stock prices over time, offering insights into historical trends and patterns.

The second dataset enhances the historical price data by incorporating date/time features, technical indicators, fundamental data, and sector-specific data. This includes a variety of technical analysis tools, such as moving averages and relative strength indices, alongside key fundamental data such as liquidity ratios and earnings per share, and other metrics, such as inflation and bond prices, that may influence stock performance. By integrating technical indicators with fundamental data, macroeconomic features, and industry-related trends, this data set aims to provide a more nuanced understanding of the factors driving stock prices.

The third dataset focuses on financial sentiment, drawing data from multiple sources to gauge the market's emotional and psychological state. This includes analysis of sentiment expressed in financial news articles and social media platforms such as Twitter. By examining the sentiment surrounding specific stocks this dataset seeks to capture the intangible factors that can significantly impact stock movements.

3.1.1 Financial Data

Historical Prices From Yahoo Finance detailed records on the historical prices of the top 25 large-cap stocks are compiled over a period of 10 years from Feb 26, 2014 untill Feb 26, 2024. The top 25 large-cap stocks collectively account for a combined weight of 44.22% in the S&P 500 index, see table 1. These stocks are carefully chosen from the SPDR SP 500 Trust ETF (SPY), which is the oldest ETF tracking the SP 500 index according to investopedia. As of September 20, 2023, SPY manages assets totaling $406.6 billion, with its portfolio weightings it offers a reliable approximation for investment strategies targeting the S&P 500 index. Even though SPY and the S&P 500 index may not align perfectly, the ETF's weightings as of September 20, 2023, serve as a substantial indicator of the index's largest constituents.

Data has been collected over a comprehensive period of 10 years starting from Feb 26, 2014 and ending on Feb 26, 2024. For each stock the data consists of 3775 rows, which includes open, high, low, close, adjusted close prices, and volume for each stock, as depicted in Table 3.1. This will provide a robust foundation for analyzing historical trends and patterns. The adjusted close prices, specifically modified for dividends and stock splits, are emphasized to accurately track price movement over time. Although other researches (59) have alternatively focused on closing prices. This choice of daily data frequency over a decade allows for a detailed examination of stock performance, capturing both short-term fluctuations and long-term trends, enhancing the depth and relevance of the analysis.

Table 3.1: Stock Price Data for AAPL

Date	Open	High	Low	Close	Adj Close	Volume
Feb 26, 2014	18.70	18.75	18.41	18.48	16.27	257,765,200
Feb 27, 2014	18.469	18.89	18.43	18.85	16.60	301,882,000
(...)	(...)	(...)	(...)	(...)	(...)	(...)
Feb 26, 2024	182.24	182.76	180.65	181.16	180.91	40,867,400
Feb 27, 2024	181.10	183.92	179.56	182.63	182.38	54,318,900

Technical Indicators For the study of stock price prediction, an array of technical indicators is utilized. The technical indicators are derived from historical price data to shed light on market trends and potential future price movements. The research incorporates the widely recognized technical indicators listed in table 3.2

Fundamental data Fundamental data provides a crucial insight into a company's financial health, reflecting its earnings, assets, liabilities, equity, and other financial metrics that investors use to assess its intrinsic value and long-term viability. They detail the organization's revenues, expenses, profitability, assets, liabilities, and cash flow operations, thereby offering a comprehensive insight into its fiscal stability and operational efficiency.

For each stock the data spans a decade, covering the period from September 27, 2014, to December 30, 2023, and comprises 38 quarterly financial summaries for a specific stock. The dataset contains a broad spectrum of metrics as can be seen in table 3.3

ESG Scores The data provided, see tables 3.4 and 3.5, represents Apple Inc.'s (AAPL) Environmental, Social, and Governance (ESG) scores over eight years, from 2014 to 2021. ESG scores are increasingly used by investors to evaluate companies based on their sustainability practices, social responsibility, and governance quality. These scores can impact investment decisions, as they reflect how well a company manages risks and opportunities associated with environmental stewardship, social impact, and leadership ethics.

Technical Indicator	Explanation	No. Days
Moving Average (MA)	Calculates the average of price data over a defined period. Assists in smoothing out price variations and emphasizing underlying market trends.	50, 100 , 200
Exponential Moving Average (EMA)	Similar to the simple MA but instead the EMA places greater emphasis on more recent price data. This technical indicator thereby reacts more sensitively to price changes. (35)	50, 100, 200
Relative Strength Index (RSI)	Momentum oscillator that assesses the extent of recent price movements to determine whether a stock is overbought or oversold.	14
Moving Average Convergence Divergence (MACD)	The MACD is a momentum indicator that follows trends by illustrating the relationship between two moving averages of a stock's price.	12, 26
Bollinger Bands	These are lines plotted at two standard deviations above and below a simple moving average (MA) of a stock's price. (71)	None
On-Balance Volume (OBV)	On-Balance Volume stands as a significant momentum indicator that leverages volume flow to forecast changes in stock price.(72)	None
Stochastic Oscillator	A momentum indicator that measures a stock's closing price relative to its price range over a designated period, oscillating between 0 and 100.(73)	14
Average Directional Index (ADX)	A indicator used to quantify the strength of a trend without considering its direction. (74)	14
Williams %R	Is a momentum indicator that identifies overbought or oversold conditions in a stock's price.	14, 28

Table 3.2: List of Technical Indicators with Explanations and Sizes Used

Industry Sector Understanding the sector weightings within the S&P 500 index is critical for analyzing the broader market trends and their potential impact on the index's overall value. As of August 31, 2023, the distribution of sector weightings in the S&P 500, according to S&P Dow Jones Indices, highlights the dominance of certain sectors and the relative insignificance of others in terms of their contribution to the index's performance.

Fundamental Data	**Explanation**
Income Statement	Revenue, gross profit, operating profits, EBITDA, and income before unusual items.
Balance Sheet	Cash, short-term investments, total assets, total debt, and common equity.
Cash Flow	Net cash flow, depreciation/amortization, capital expenditures, net change in cash, and free cash flow net of dividends.
Per Share Data	Dividend yield, diluted EPS, shares used to calculate EPS.
Profitability and Return	Margins (gross, EBITDA, operating, net) and returns (free cash flow yield, ROE, ROA, ROIC).
Growth	Growth in revenue, operating profit, EBITDA, income, and EPS.
Financial Strength	Ratios: debt/asset, debt/capital, debt/equity, interest coverage, dividend coverage.
Enterprise Value	Market cap, total debt, cash, and short-term investments.
Earning Power	Asset turnover, income before tax margin, pretax ROA, pretax ROE, tax complement.
Liquidity	Ratios (current, quick), operational metrics (receivables, payables, inventory turnover, net trade cycle).

Table 3.3: List of Fundamental Data with Explanations

The S&P 500's composition by sector and their respective index weightings are presented in Table 3.4.

The sector weightings, obtained from Investopedia, within the S&P 500 index, as shown in 3.6, play a pivotal role in determining the overall valuation of the index and its responsiveness to shifts within specific market sectors. For example, the Information Technology sector, which boasts a significant weighting of 28.2%, has a pronounced impact on the index's performance. In contrast, sectors such as Energy, Materials, Real Estate, and Utilities, each with weightings below 5%, exert a comparatively minor influence on the index's overall value.

In addition to sector weighting, the performance of each sector can serve as a useful and informative tool for understanding broader market dynamics. Utilizing data from Novel Investor the analysis can be enriched further. The Data from Novel Investor provides

ESG Factor	Explanation
ESG Combined Score	Overall rating combining environmental, social, and governance scores. Grades range from "A" (highest) to lower grades. Apple improved from "C-" in 2014 to "C+" in 2021.
ESG Score	Focuses on ESG factors excluding controversies. Apple improved from "B-" in 2014 to "A-" in 2021.
Environmental Pillar Score	Reflects Apple's impact on the environment. Improved from "B-" to "B".
Social Pillar Score	Assesses relationships with employees, suppliers, customers, and communities. Improved from "C+" in 2014 to "A-" in 2021.
Governance Pillar Score	Evaluates management quality, board, and ethics. Apple consistently scores "A" or better.
ESG Controversies Score	Measures management of ESG controversies. Apple's score remained "D-" from 2018 to 2021.
Resource Use Score	Focuses on resource efficiency. Apple consistently scores "A".

Table 3.4: Example ESG Scores and Factors

Table 3.5: AAPL ESG Scores (2014-2021)

Year	ESG Combined	ESG Score	Environ.	Social	Governance	Controversies	Resource Use
2021	C+	A-	B	A-	A+	D-	A+
2020	C	A-	B	A-	A	D-	A+
2019	C	B+	B	B	A-	D-	A+
2018	C	B+	B-	B+	A	D-	A+
2017	C	B+	B-	B	A	D	A+
2016	C	B	B-	B-	A	C+	A
2015	C-	B-	B-	C+	A-	C-	A+
2014	C-	B-	B-	C+	A	D+	A-

detailed performance metrics for the S&P 500 sectors from 2009 to 2023. This extended dataset allows for a comprehensive examination of long-term trends and cyclical behaviors within the index. 3.7 shows a snippet of the long-term performance trends across selected sectors of the SP 500.

Table 3.6: S&P 500 Sector Weightings as of August 31, 2023

Sector	**Index Weighting**
Information Technology	28.2%
Healthcare	13.2%
Financials	12.5%
Consumer Discretionary	10.6%
Communication Services	8.8%
Industrials	8.4%
Consumer Staples	6.6%
Energy	4.4%
Materials	2.5%
Real Estate	2.4%
Utilities	2.4%

Table 3.7: Sector Performance from 2009 to 2023 (Selected Years)

Sector	**2009**	**2010**	**2011**	**2015**	**2020**	**2023**
COND	41.3%	27.7%	6.1%	10.1%	33.3%	42.4%
CONS	14.9%	14.1%	14.0%	6.6%	10.8%	0.5%
ENRS	13.8%	20.5%	4.7%	-21.1%	-33.7%	-1.3%
FINL	17.2%	12.1%	-17.1%	-1.5%	-1.7%	12.2%
HLTH	19.7%	2.9%	12.7%	6.9%	13.5%	2.1%

Macro-Economics This research also incorporates macro-economic data, highlighting the importance of broader economic indicators on stock market dynamics. It includes monthly inflation data which provides insights into the general economic environment, see table 3.8. Inflation also reflects changes in consumer purchasing power and potential shifts in central bank policies.

Table 3.8: Monthly and Annual Inflation Rates

Month	Monthly Inflation Rate (%) (seasonally adjusted)	Annual Inflation Rate (%) (not seasonally adjusted)
January 2024	0.3	3.1
December 2023	0.2	3.4
November 2023	0.1	3.1
October 2023	0.0	3.2
September 2023	0.4	3.7
August 2023	0.6	3.7
July 2023	0.2	3.2
June 2023	0.2	3.0

In addition to inflation, this study also takes into account Treasury Bond prices. Table 3.9 shows the yields on U.S. Treasury securities at different maturities, recorded at specific dates, typically towards the year's end. The yields are expressed as annual percentages and provide insight into the interest rate environment, investor expectations about future inflation, and the overall economic outlook. U.S. Treasury securities are considered risk-free assets, making these yields benchmarks for other interest rates.

Table 3.9: U.S. Treasury Yields at Year-End

Date	1 Mo	3 Mo	6 Mo	1 Yr	2 Yr	3 Yr	5 Yr	7 Yr	10 Yr	20 Yr	30 Yr
12/31/2014	0.03	0.04	0.12	0.25	0.67	1.1	1.65	1.97	2.17	2.47	2.75
12/30/2014	0.03	0.03	0.12	0.23	0.69	1.11	1.68	2.00	2.20	2.49	2.76

3.1.2 Sentiment Data

Model Selection Data For model selection, the approach adopted involves using sentiment data from existing, publicly accessible repositories. Renowned platforms such as

Hugging Face constitute a rich source of such datasets, notably those tailored for financial sentiment analysis. The repositories offer structured compilations of textual data, encompassing tweets, and news headlines, each annotated with sentiment labels. This pre-structured and pre-labeled data is instrumental for the initial phases of model training and validation, where models are exposed to a variety of sentiment expressions to learn nuanced differentiation between positive, negative, and neutral market sentiments.

For example, the "twitter-financial-news-sentiment" dataset on Hugging Face offers a curated selection of financial news articles, each meticulously annotated with sentiment labels. The sentiment labels are categorized into three primary types: positive, indicating optimistic or favorable views; neutral, denoting unbiased or informational content; and negative, reflecting pessimistic or adverse opinions. In the context of the "twitter-financial-news-sentiment" dataset, an entry might include for example, a news headline stating, "Nomura points to bookings weakness at Carnival and Royal Caribbean," might be labeled as "negative" due to its unfavorable implications for the company's financial health. Conversely, a tweet expressing concerns over decreasing debt levels within a particular sector might be tagged as "positive".

Another valuable resource for this research is the Financial Phrasebank, a publicly accessible dataset that offers a comprehensive collection of financial phrases and sentences. The Financial Phrasebank compiles an extensive range of sentences derived from financial news articles, each meticulously annotated to reflect sentiments such as positive, neutral, or negative. This dataset is particularly useful for analyzing the linguistic nuances and sentiment expressions prevalent in financial discourse.

To account for the inherent subjectivity in sentiment analysis, each sentence in the collection received between 5 to 8 annotations, facilitating a robust consensus on the sentiment expressed. To cater to different levels of consensus and provide an objective basis for comparison, four alternative reference datasets were created, classified based on the degree of annotator agreement:

1. Sentences with 100% Agreement (Sentences_AllAgree.txt): This subset contains sentences where there was unanimous agreement among annotators on the sentiment expressed.

2. Sentences with More Than 75% Agreement (Sentences_75Agree.txt): Includes sentences where over 75% of annotators concurred on the sentiment.

3. Sentences with More Than 66% Agreement (Sentences_66Agree.txt): Consists of sentences with more than two-thirds of annotators in agreement.

4. Sentences with More Than 50% Agreement (Sentences_50Agree.txt): Features sentences where a simple majority (over 50%) of annotators agreed on the sentiment.

Each of these datasets is presented in a machine-readable format, with sentences separated from their annotated sentiment by an "@" symbol, for instance, sentence@sentiment. The sentiment labels used are "positive", "neutral", or "negative" allowing for straightforward integration into sentiment analysis models.

Here are two examples, see 3.10 and 3.11 from the 100% annotator agreement and 50% annotaor agreement datasets, showcasing how sentences are annotated with sentiments. The dataset structure and annotation process, as detailed above, ensure a nuanced and accurate representation of sentiment in financial contexts.

Table 3.10: Example Sentences from the Financial Phrasebank - 100% agreement

Sentence	Sentiment
"According to Gran , the company has no plans to move all production to Russia , although that is where the company is growing"	Neutral
"For the last quarter of 2010, Componenta's net sales doubled to EUR131m from EUR76m for the same period a year earlier, while it moved to a zero pre-tax profit from a pre-tax loss of EUR7m"	Positive

Table 3.11: Example Sentences from the Financial Phrasebank - 50% agreement

Sentence	Sentiment
"In Sweden , Gallerix accumulated SEK denominated sales were down 1% and EUR denominated sales were up 11 %."	Neutral
"Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications , the statement said."	Neutral

Testing Data Collection The study places a strong emphasis on gathering up-to-date sentiment data while gathering it for testing purposes. Although new data is preferred, it is very hard to get. Instead, sentiment data considering tweets for Apple, Google, Tesla, Microsoft, and Amazon in the period from january 2015 to december 2019 will be used. This kind of data derived from Twitter, could provide an instantaneous gauge of the general mood of the market, catching the pulse of sentiment as it develops.

The number of tweets for each stock differs both in total and daily. Specifically, Apple has 527033 tweets, Google has 157145 tweets, Microsoft has 68959 tweets, Amazon has 247800 tweets and Tesla has 699704 tweets. These tweets are collected from the hugging face dataset "twitter-financial-news-sentiment", ensuring a comprehensive and reliable source for sentiment analysis.

3.2 Preprocessing & Feature Engineering

3.2.1 Financial Data

Target Variable Given the realistic assumption that there is no access to the stock data the day before, the closing price on day $t + 1$ is predicted using information up to day t. Using the closing prices, the target variable can be defined as the closing price of the next day.

To incorporate this feature into our dataset, each stock's daily closing prices are shifted by 1 row. The outcome is a new column in our dataset, now serving as the target variable for our predictive models. This enables the leveraging of current and historical price data in forecasting the actual future prices, under the premise that future stock information is not accessible at time t.

Date/Time-Features Temporal features are crucial in identifying trends, seasonal patterns, and anomalies within the financial markets. Extracting useful information from any date/time attributes into financial data analysis is crucial and offers a varied understanding of market dynamics and investor behaviour. Table 3.12 an overview of the calculation process and its significance:

Technical Indicators Preprocessing financial data is a pivotal step in preparing the dataset for analysis. This involves calculating various technical indicators from the historical price and volume data, i.e. feature engineering. Each of these indicators and their strategy, mostly binary features indicating buy or sell signals, will be represented as separate features in the dataset, calculated for each stock based on its historical price and volume data. This enriched dataset will then serve as the basis for subsequent analysis,

Date Feature	Explanation	Formula
Extraction of Date Components	Fundamental temporal attributes like `Year`, `Month` and `Day` are directly extracted from the date column to analyze specific time period influences on financial metrics. This is done to capture possible behavioral effects such as the previously mentioned January effect (14) and weekend effect (13).	$\texttt{Year} = \text{date.year}$ $\texttt{Month} = \text{date.month}$ $\texttt{Day} = \text{date.day}$
Calculation of Cyclic Features	Sine and cosine transformations are applied to time components (e.g., `Month_sin`, `Month_cos`) to account for cyclical nature of time, ensuring continuity between cycle ends and starts. This captures seasonal trends and weekly patterns.	$\texttt{Month\_sin} = \sin\left(\frac{2\pi\times\texttt{Month}}{12}\right)$ $\texttt{Month\_cos} = \cos\left(\frac{2\pi\times\texttt{Month}}{12}\right)$
Identification of Special Periods	Binary features like `Is_Monday`, `Is_Month_End`, `Is_Quarter_End`, and `Is_January` highlight specific periods of interest, reflecting unique financial behaviors or anomalies.	$\texttt{Is\_Month\_End} = \begin{cases} 1 & \text{if Day} = \text{last day of month} \\ 0 & \text{otherwise} \end{cases}$ $\texttt{Is\_Quarter\_End} = \begin{cases} 1 & \text{if date} \in \{\text{last day of quarter}\} \\ 0 & \text{otherwise} \end{cases}$

Table 3.12: Date-Related Features

enabling the exploration of relationships between these technical indicators and stock price movements. They are calculated as follows:

- **Simple Moving Average (SMA)** calculates the average closing price over a specified number of days, n.

  $$\text{SMA}_n = \frac{\sum_{i=1}^{n} \text{Close}_i}{n}$$

  Exponential Moving Average (EMA) places greater emphasis on recent prices, using a factor k to adjust the weighting.

  $$\text{EMA}_t = (\text{Close}_t \times k) + \text{EMA}_{t-1} \times (1 - k), \quad k = \frac{2}{n+1}$$

Strategy: MAs help identify trend direction. A short-term MA crossing above a long-term MA suggests a buy signal; the reverse signals a sell. EMA crossovers provide similar signals.

- **Relative Strength Index (RSI)** RSI assesses the magnitude of recent gains to losses to identify overbought or oversold conditions.

$$\text{RSI} = 100 - \frac{100}{1 + \frac{\text{Average Gain}}{\text{Average Loss}}}$$

Strategy: Buy signals are typically given by RSI values crossing above 30 and sell signals below 70.

- **Moving Average Convergence Divergence (MACD)** MACD indicates the difference between two EMAs.

$$\text{MACD} = \text{EMA}_{12} - \text{EMA}_{26}$$

Strategy: A buy is suggested when MACD crosses above its signal line (9-day EMA of MACD), and a sell when below.

- **Bollinger Bands** Consists of a middle SMA and upper/lower bands determined by the standard deviation (SD) over the same period as the SMA.

$$\begin{aligned}\text{Middle Band} &= \text{SMA}_{20},\\ \text{Upper Band} &= \text{Middle Band} + 2 \times \text{SD}_{20},\\ \text{Lower Band} &= \text{Middle Band} - 2 \times \text{SD}_{20}\end{aligned}$$

Strategy: Buy when the price hits the lower band; sell when it reaches the upper band.

- **On-Balance Volume (OBV)** OBV uses volume flow to anticipate price changes.

$$\text{OBV}_t = \text{OBV}_{t-1} + \begin{cases}\text{Volume}_t & \text{if Close}_t > \text{Close}_{t-1}\\ -\text{Volume}_t & \text{if Close}_t < \text{Close}_{t-1}\\ 0 & \text{if Close}_t = \text{Close}_{t-1}\end{cases}$$

Strategy: Increasing OBV indicates potential buy signals; decreasing OBV suggests sells.

- **Stochastic Oscillator** This indicator compares the closing price to the price range over a given period, in this example 14 days, often used to predict reversal points.

$$\%K = \frac{\text{Close} - \text{Low}_{14}}{\text{High}_{14} - \text{Low}_{14}} \times 100$$

Strategy: Buy signals are given when %K rises above 20 (indicating oversold conditions) and sell signals when it falls below 80 (overbought).

- **Williams %R** Similar to the Stochastic Oscillator, Williams %R also identifies overbought or oversold levels.

$$\%R = \frac{\text{High}_n - \text{Close}}{\text{High}_n - \text{Low}_n} \times -100$$

Strategy: Readings below -80 suggest buy (oversold); above -20 indicate sell (overbought).

- **Average Directional Index (ADX)** Measures trend strength. The formula used to calculate the basic ADX is:

$$\text{ADX} = \frac{\text{Smoothed Moving Average of DX}}{n}$$

Strategy: An increasing ADX indicates a strengthening trend, suitable for trading in the direction of the trend. Values above 25 suggest strong trends. For a detailed computation, see in the Appendix table 2.

Fundamental data The financial summary data, in other words the quarterly reports, undergo multiple prepossessing steps. First, any rows containing missing data are dropped to ensure completeness. After dropping rows with any missing values, the dataset is transposed such that each row represents a different time period with the financial metrics as columns. This aligns the data structure with the historical data. After this is done, the first row is set as column headers to accurately describe each feature. Then after converting the date information into date-time format to synchronize with the historical price data, the data of the financial summary and earnings per share are merged with the historical price data set.

ESG Scores The first step is to perform the score mapping by translating the ESG grades into numerical features. This conversion facilitates quantitative analysis, allowing for a more straightforward comparisons and aggregations of ESG performance across different periods and companies. The mapping follows a scale where 'A+' has the highest score (7) and 'D-' is the lowest (-4), see table 3.13.

Table 3.13: ESG Score Mapping

ESG Grade	Numerical Score	ESG Grade	Numerical Score
A+	7	C+	1
A	6	C	0
A-	5	C-	-1
B+	4	D+	-2
B	3	D	-3
B-	2	D-	-4

Industry Sector & Macro-Economics The first step is to map a single sector from the table 3.6 to each stock. The next step includes mapping the sector performance, see figure 3.2, to the corresponding years for each stock. A similar approach is used for the macro-economic factors such as inflation and treasury rates. These factors do not change in value based on different stock, however, the factors have different values for different months and years. Therefore, the data must be mapped to corresponding year and/or month.

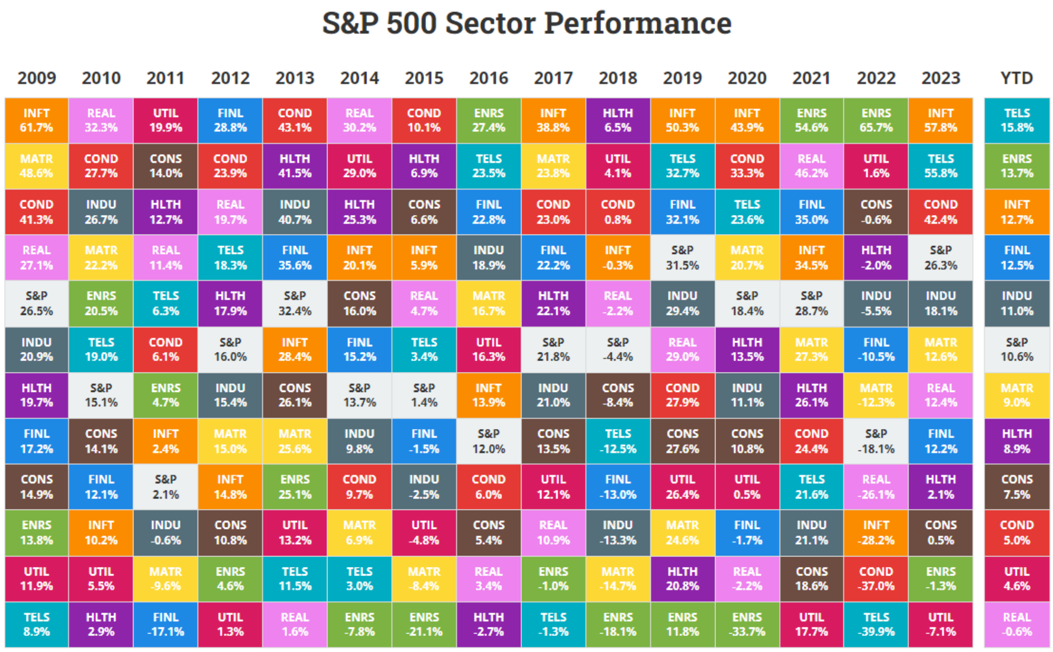

Figure 3.2: Sector Performance

3.2.2 Sentiment Data

Given the quality and structure of both the financial phrase bank (for training) and the hugging face dataset (for testing), extensive preprocessing is not necessary. However, several basic and fundamental cleaning procedures still need to take place to ensure the accuracy and reliability of the sentiment analysis. These procedures include:

- Deduplication: Check for duplicate tweets and remove any duplicate tweets to ensure that each tweet is unique and does not artificially inflate sentiment scores.

- Language Standardization: Check for non-English tweets which should in turn be excluded, ensuring consistency in language and making sentiment analysis more accurate.

- Tokenization: Tweets should be tokenized, converting text into a format suitable for sentiment analysis algorithms. This involves breaking down sentences into individual words or tokens.

As it turns out, the data does not include any duplicate tweets meaning that no tweets have to be removed. Moreover, upon further investigation all sentences are English tweets meaning there is consistency in tweets with respect to language. The final preprocessing step involves tokenizing the tweets using the BERT model. This is done through processing the text into smaller parts called tokens, which are often words or subwords. BERT uses a sophisticated tokenization method that uses WordPiece tokenization to break the words into subwords or characters, allowing the effective handling of rare words and misspellings. An example of this is the word "playing" which might be tokenized into "play" and "##ing" where "##" indicates that "ing" is a suffix.

3.3 Feature Selection

Feature selection is essential in the development of predictive models for financial markets, as it significantly influences both model performance and efficiency. This sub-chapter focuses on identifying the most impactful variables to enhance model accuracy and streamline computation. This is done by addressing data quality issues, missing values, and feature selection.

3.3.1 Missing Values

After combining all relevant data into a single dataset for each stock, it becomes evident that missing values predominantly appear at the beginning or the end of the time series. This pattern is attributed to the calculation methodologies of certain features, such as moving averages, which inherently produce NaN values during their initial periods. Additionally, some data such as Earnings Per Share (EPS) or Environmental, Social, and Governance (ESG) scores may not be immediately available for recent periods such as the start of 2024. Consequently, missing values are not distributed randomly but are concentrated at specific intervals along the time series.

Given this context, selecting an appropriate method to handle missing values is crucial for preserving the dataset's integrity and ensuring accurate analyses and models. The strategy chosen significantly influences model performance and the accuracy of subsequent analyses. Possible solutions are **1)** deletion of rows or columns, **2)** imputation by mean, median, or mode, **3)** imputation by leveraging regression models or techniques that use similiarity measures such as K-Nearest Neighbors (KNN), **4)** the absence of data might carry usefull information an should be retained in the data.

Considering these methods and the dataset's characteristics, the most suitable approach depends on the specific feature and the extent of missing data. The table below summarizes the top 20 features by the number of missing values in the combined dataset:

Table 3.14: Top Features with Most Missing Values

Feature	Missing Values	Feature	Missing Values
4 Mo	2217	RSI-14	14
2 Mo	1466	ADX-14	14
ESG Combined Score	790	%K	13
MA-200	199	High_14	13
MA-100	99	Low_14	13
MA-50	49	%R	13
Lower_Band	19	plus_di-14	13
Upper_Band	19	minus_di-14	13
Middle_Band	19	RSI-Position	1
MA-20	19	MA-Position	1

In addressing the significant number of missing values for features such as "4 Mo" and "2 Mo," these columns were excluded from the dataset due to the potential for introducing bias or inaccuracies through imputation. For moving averages, regression imputation was deemed the most suitable approach as it preserves the trends indicated by these features better than backward or forward filling methods, which might obscure genuine data trends.

The ESG Combined scores, characterized by stability across periods for other stocks, were imputed using the forward fill method. This is done under the assumption that the last observed score remains applicable in the near future.

For features with fewer than 20 missing entries, temporal or sequential imputation techniques including forward and backward filling, were employed. These methods are particularly effective for time-series data where the proximity of observations can provide a reliable basis for imputation. Moreover, these methods also maintain the temporal coherence of the dataset.

3.3.2 Normalization

Normalizing the data is necessary before proceeding to the next stage. This is important for a number of reasons. First, there is a wide variety in the scale of financial data, including technical indicators. For instance, prices may be in the tens, hundreds, or even thousands, yet trade volume may be in the millions. These differences, in the absence of normalization, have the potential to distort the analysis by favoring variables with bigger scales thus leading to less accurate predictions.

This is resolved via normalization, which places all data on a common scale, facilitating more equal processing and analysis of the data by algorithms. This is particularly crucial for machine learning models, as the size of the input features can have significant impact on the convergence of algorithms and the precision of predictions. By ensuring that every feature contributes appropriately to the model's decision-making process, normalization produces outcomes that are easier to interpret and more dependable.

The normalization process is applied as follows: 1) Removal of non-Feature Columns: Columns such as 'Date' and 'Target_variable' are dropped from the feature set, as these are not required for model training. 2) Application of Normalization: The remaining features are normalized using the minmax scaler, i.e, log scaling. 3) Reconstruction of DataFrame: Post-normalization, the numpy arrays are converted back into pandas DataFrames to retain the original structure and facilitate further analysis or model training.

The logarithmic scaling normalization method was selected over other normalization methods because time series models generally assume that the variance of a seris remains constant over time. According to (75), forecasts based on the log transformation can be much better if the log transformation results has a more stable variance. If it turns out that log transforming yield higher variance, direct forecasting without normalization is preferred. Besides variance, log transforming the data can transform potential non-linear trend to linear trends, which allows for easier data analysis.

$$\text{Let } X_t \text{ be the original time series data at time } t.$$

$$\text{The log-transformed data } Y_t \text{ is defined as: } Y_t = \log(X_t).$$

$$\text{Use } Y_t \text{ for modeling if: } \operatorname{Var}(Y_t) < \operatorname{Var}(X_t).$$

3.3.3 Feature Selection

Since the model's performance does not automatically improve with more features added to the dataset, feature selection is the last and final stage within this chapter. The feature selection approach involves multiple methodologies to identify significant features in different stocks, including large, medium, and small stocks in various sectors. This strategy is adopted to tackle several issues. Some issues occur because the dataset contains a large number of features and aggregated data from many sources. Among these difficulties are:

1. **Curse of Dimensionality:** As more features are added, the complexity of the model increases without necessarily increasing the amount of useful information. This causes several problems, such as data sparsity, increased computation time, over fitting, performance degradation, and visualization challenges.

2. **Multicollinearity:** Adding more features increases the risk of linearly dependent predictors, or, in other words, multicollinearity. Having two or more highly correlated features makes it difficult to assess the independent effect of each feature on the target variable. This can, in turn, lead to unstable estimates of coefficients in predictive models, making the interpretation of feature importance more complicated.

3. **Dilution of Feature importance:** With many features, especially in higher-dimensional datasets, the relative importance of a single predictor may seem diluted. While the correlation of individual features with the target remains unchanged, it becomes challenging to identify the most influential predictors among a larger set of features.

The initial step involves a correlation analysis to assess the linear and ordinal relationships between features and target variable (e.g 'Close'). The correlation analysis involves calculating the Pearson, Kendall, and Spearman correlation coefficients for each feature with respect to the target variable. This analysis aims to capture different aspects of these relationships:

1. **Pearson Correlation:** Identifies the degree of linear correlation between two continuous variables.

2. **Kendall and Spearman (Rank-Based) Correlations:** Evaluates the ordinal association and are useful for identifying non-linear relationships that pearson might miss.

For example, see table 3.15, features like "EMA-12" and "Upper Band" exhibited strong Pearson correlations with " Close". This suggests their potential utility in predicting stock prices. However, the higher values of Kendall and Spearman correlation also pointed out complexities beyond linear relationships. This indicates a need for models to be capable of capturing and handling such nuances.

Table 3.15: Features with High Correlation with Target Variable

Feature	Pearson	Kendall	Spearman
Previous Close	0.999077	0.965966	0.998313
EMA-12	0.998569	0.959612	0.997681
Low_14	0.997319	0.946373	0.995812
EMA-26	0.997279	0.944278	0.995657
High_14	0.997149	0.943680	0.995343
Upper_Band	0.997120	0.942049	0.995455

Given the varied nature of correlations, the feature selection strategy included methods to harness both linear and non-linear relationships. These methods include:

1. **Principal Component Analysis (PCA):** Used to reduce the dataset dimensionality while retaining variation present in original variables. This helps to adress the issue of multicollinearity.

2. **Automated Feature Selection Tools:** Techniques like Recursive Feature Elimination (RFE) were employed to systematically remove less important features.

3. **Ridge Regression & Sequential Feature Selection:** Applied to combat overfitting and multicollinearity ensuring robust model performance. Additionally, Ridge Regression was used as an estimator for both forward and backward sequential feature selection.

4. **Mutual Info Regression:** This helped capture non-linear relationships between features and target variable.

Following these methods, a robust set of features for each stock was identified, with the average number of features selected typically ranging from 20 to 30. Notably, several features emerged as commonly selected across various stock categories, highlighting their universal relevance and utility in financial modeling.

Commonly Selected Features Across Stocks

- Fundamental indicators like *Open* and volume indicators (*Volume*, *OBV*) appear consistently across most stocks.
- Trend-capturing features such as moving averages (*MA-20*, *EMA-12*) are widely used.
- Financial health indicators including *Total Assets* and liquidity metrics (*Cash & Cash Equivalents*) are crucial for assessing company performance.

Features Common in Large Stocks

- Large-cap stocks like AAPL (Apple) and MSFT (Microsoft) commonly include fundamental financial data such as *Market Capitalization* and *Operating Profit*.
- Complex technical indicators like *OBV-EMA-20*, *PC1*, *PC2*, and *PC3* are also prevalent.
- Basic price-related features such as *Volume* and *Open* are emphasized due to their significant market impact.

Features in Medium/Small Stocks

- Smaller stocks select features related to volatility or specific technical indicators such as *%R* and *minus_ di-14* that capture short-term movements.
- Time-based trend features such as *Day_ cos* and *Month_ cos* are more frequently selected in medium or small-cap stocks.

4

Methods & Design

This chapter presents the methodologies and specific methods employed to address the research questions identified in Chapter 1. It is essential to choose the correct methodologies to ensure the robustness, validity, and reliability of the research findings. The chapter is organized into three main sections: Section 4.1 explains the research design and overarching strategy; Section 4.2 details the machine learning algorithms used (see Figure 4.1 for an overview of the models); and Section 4.3 discusses the evaluation metrics and validation procedures, with a focus on the three-stage testing process.

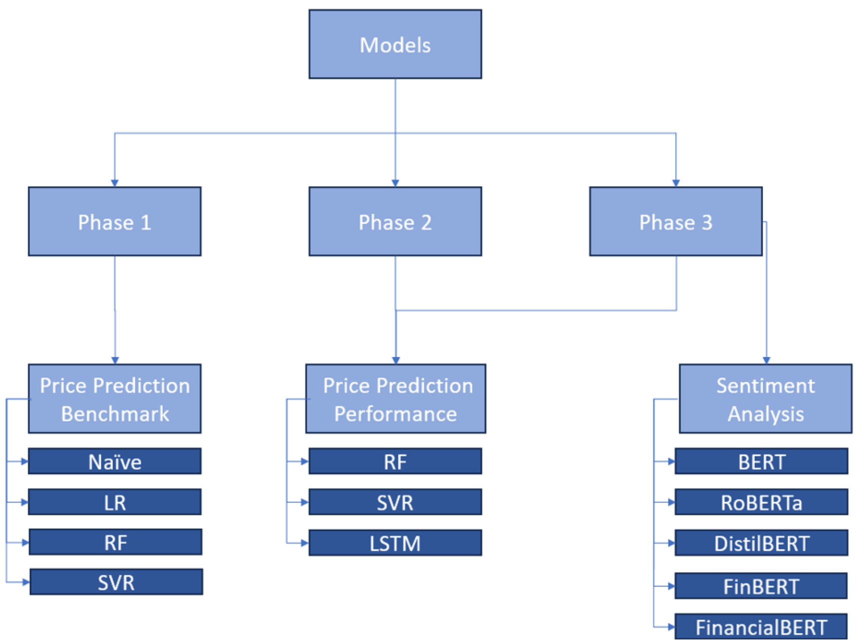

Figure 4.1: Stock Price Prediction Models

4.1 Research Design and Strategy

The research strategy involves a phased testing approach to evaluate multiple models across different phases and compare their predictive accuracy and performance. This phased approach assesses the effectiveness of models when new data is introduced. Each model is applied over a rolling window to prevent look-ahead bias. The rolling window sizes used are 500 days, 250 days, and 125 days, corresponding to approximately two years, one year, and half a year of trading days respectively. These varying window sizes allow the models to capture different temporal dynamics and adjust more effectively to new data, which is crucial for handling the non-stationary nature of financial time series. Moreover, by adjusting the window size, the model's ability to adapt to different temporal dynamics is explored, potentially enhancing its predictive performance across different time periods.

Additionally, the continuous adaptation of models to new data offers a significant advantage in predicting financial markets, where conditions can change rapidly. Due to the high computation time associated with more complex models such as Random Forest, a rolling window is used with a weekly (5-day) adaptation to new data instead of daily. This balance between update frequency and computational efficiency helps to manage the trade-offs between timely model updates and resource constraints.

4.1.1 Phase 1

The first phase, or the benchmark phase, involves simple models that use only open, high, low, and closing prices together with volume. This phase sets a baseline for performance comparison with more complex models in subsequent phases. The benchmark models include:

1. **Naive Model:** Uses the previous day's closing price as a prediction for the next day's closing price.

2. **Linear Regression:** Considers the linear relationship between input features and target predictions.

3. **Random Forest and Support Vector Regression (SVR):** Aim to capture non-linear patterns and relationships in data.

4.1.2 Phase 2

Following the benchmark models in the first phase, phase two introduces a more sophisticated analytical approach by incorporating additional features and utilizing deep learning techniques. Besides the RF and SVR models used in Phase 1, this phase introduces LSTM networks. LSTMs are well-suited for making predictions based on time series data because of their ability to capture long-term dependencies and relationships that simpler models might miss. They are particularly adept at handling the noise and volatility inherent in financial markets (43).

The LSTM models will be trained using a sliding window approach similar to Phase 1 to continuously adapt to new data. This method involves periodically retraining the model on the most recent data, which helps it stay relevant as market conditions change. The performance of the LSTM model will be compared against the benchmark models established in Phase 1 to evaluate the incremental value brought by complex models and additional features in forecasting financial time series.

4.1.3 Phase 3

The final stage, Phase 3, expands the analytical framework to incorporate sentiment analysis, recognizing the significant impact of market sentiment on financial markets. This stage involves training and testing BERT (Bidirectional Encoder Representations from Transformers) models on financial headlines and tweets, followed by the application of the best-performing model to conduct zero-shot learning on financial tweets. The insights derived from sentiment analysis are then integrated into the RF, SVR, and LSTM models from Phase 2 to assess their combined effectiveness in predicting financial market movements.

Training Phase: Evaluating BERT Models for Sentiment Analysis In the training phase, multiple pre-trained BERT models will be evaluated against the financial phrase bank. The fine-tuning process involves adjusting the hyperparameters of the pre-trained models so that they can effectively classify the sentiment expressed in financial texts. Performance metrics such as accuracy, precision, recall, F1-score, and confusion matrices will be used to assess each model's ability to correctly interpret the sentiment of financial headlines and tweets.

The model that demonstrates the highest accuracy and generalizability in sentiment classification will be selected for the subsequent testing phase. This selection ensures that the most capable model is used to analyze real-time data.

Testing Phase: Zero-Shot Learning and Sentiment Integration The chosen BERT model will be employed to perform zero-shot learning on a new dataset of financial-related tweets collected during the period 2015-2019. Zero-shot learning allows the model to classify sentiment on data it has not seen before without additional training, making it highly adaptable to new information.

The model will classify these tweets into predefined sentiment categories (positive, negative, neutral), providing a real-time snapshot of market sentiment. The sentiment data, now structured into a time-series format indicating sentiment trends over time, will be incorporated as an additional input feature into the models developed in Phase 2. This integration aims to evaluate whether sentiment data can enhance the predictive accuracy of financial market forecasts by providing contextual insights that price data alone may not fully capture.

4.2 Machine Learning Algorithms

In this section, various machine learning algorithms are explored, discussing their theoretical foundations as well as practical implementation within the context of the models used.

4.2.1 Phase 1

Naive Model As discussed before, traditional finance supports the assumption that there is no better prediction than yesterday's closing price, given all information up until today. Given this assumption, the Naive Model serves as a fundamental baseline for our predictive analysis. This simplistic approach relies solely on the adjusted closing price of the previous day as the prediction for the current day. Although straightforward, this model provides a reference point against which the performance of more advanced algorithms can be compared. Its simplicity allows quick implementation and serves as a starting point for evaluating the effectiveness of more complex models in capturing the nuances of market behavior.

Linear Regression

Theoretical Framework Linear regression, a fundamental tool in statistical analysis, provides a basic approach for examining the relationship between a dependent variable and one or more independent variables. This model is built on the assumption that there is a linear relationship between the predictors and the target variable, aiming to identify the optimal linear equation that minimizes the residual sum of squares.

Practical Implementation Leveraging prior feature selection efforts, where relevant predictors were identified, the implementation of linear regression is done as follows. Python's sklearn library LinearRegression is used to perform the execution of linear regression analyses. While the parameter tuning process may not be as elaborate as that of more intricate models like random forest, SVR, or LSTM, it remains pivotal for optimizing model performance.

During this phase, the model is subjected to experimentation with varying rolling window sizes. This adaptive approach allows for the capture of temporal patterns within the data, potentially enhancing predictive accuracy across diverse temporal contexts. To illustrate, consider a rolling window of size n, where the regression model is recalculated at each step, including only the most recent n observations:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \epsilon_t, \quad \text{for } t = n \text{ to } T$$

where y_t is the dependent variable, $x_{1,t}, \ldots, x_{k,t}$ are the independent variables at time t, and $\beta_0, \beta_1, \ldots, \beta_k$ are the coefficients estimated using the data from the rolling window.

Fine-tuning of hyperparameters in linear regression involves adjustments to regularization parameters, such as alpha in methods like Ridge or Lasso regression. Regularization aids in curtailing overfitting by penalizing large coefficients, thereby fostering simpler models with superior generalization capabilities.

Random Forest

The available dataset is well-suited for Random Forest, a machine learning technique capable of effectively capturing intricate data patterns in high-dimensional datasets (76).

Theoretical Framework Originally conceptualized by Leo Breiman (77), the random forest machine learning model is designed for both classification and regression tasks. Breiman introduced an ensemble method that aggregates the results of multiple decision trees, using voting for classification and averaging for regression, as illustrated in 4.2.

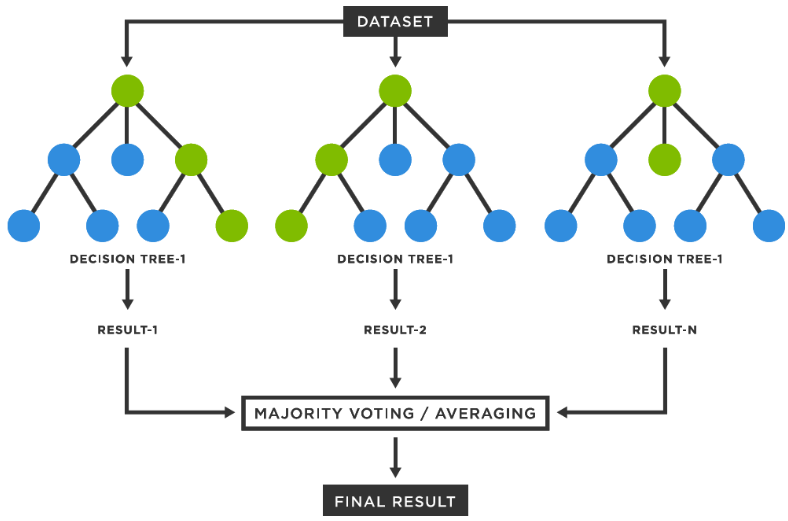

Figure 4.2: Representation of Random Forest, obtained from (78)

Practical Implementation The implementation of the random forest algorithm uses Python's sklearn libraries, specifically the RandomForestRegressor module. Tuning the models is facilitated through sklearn's functions.

In this phase, the model is executed using different rolling window sizes, an approach aimed at capturing various temporal patterns within the data. The subsequent steps involve fine-tuning the model's hyperparameters to optimize its performance. This process entails exhaustive evaluation of various hyperparameter configurations through k-fold cross-validation using the GridSearchCV library. The hyperparameters subjected to optimization include:

1. n_estimators: 50, 100, 200, 300
2. max_depth: None, 10, 20
3. min_samples_split: 2, 5, 10

`n_estimators` defines the number of trees in the ensemble, `max_depth` specifies the maximum depth allowed for each tree, `min_samples_split` determines the minimum number of samples required to split an internal node.

Support Vector Regression

Theoretical Framework Support Vector Regression (SVR), proposed by (79), extends the principles of SVM to the regression context, offering a robust approach for modeling nonlinear relationships between input variables and continuous target variables. SVR seeks to identify the hyperplane that best represents the data by optimizing the margin, which is the distance between the hyperplane and the nearest data points, referred to as support vectors. By employing kernel functions, SVR can effectively capture complex nonlinear patterns in the data.

Practical Implementation In practical application, SVR provides a powerful tool for forecasting continuous outcomes, particularly in scenarios where linear regression proves inadequate due to the presence of nonlinear relationships. Following prior feature selection endeavors, the implementation of SVR commences using Python's sklearn library, which offers robust modules such as SVR.

During this phase, the model undergoes experimentation with different rolling window sizes, akin to methodologies adopted for other regression techniques. Fine-tuning of hyperparameters in SVR is essential for achieving optimal model performance. Key hyperparameters such as kernel type, regularization parameter (C), and kernel coefficient (gamma) require careful calibration.

4.2.2 Phase 2

The RF and SVR models will be used as described in the previous subsection. Additionally, this phase introduces the LSTM network.

Long-Short Term Memory

Theoretical Framework LSTM is a specific architecture of recurrent neural networks (RNNs) developed to address the vanishing gradient problem that frequently arises in traditional RNNs. LSTM networks are equipped with distinct memory cells and gating mechanisms, which allow them to effectively manage and preserve information over extended sequences by selectively retaining or discarding data as necessary. This enables LSTMs to maintain a robust memory of past inputs, making them particularly useful for tasks involving long-term dependencies. This unique architecture enables LSTMs to capture temporal dependencies in sequential data while mitigating the issues of vanishing gradients and exploding gradients. Originally proposed by Hochreiter and Schmidhuber in 1997 (80), LSTM has become a cornerstone in sequential data modeling, finding applications in various domains such as natural language processing, time series forecasting, and speech recognition.

The LSTM architecture (see Figure 4.3) consists of three main gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information into, out of, and within the LSTM cell. The input gate (denoted by $i(t)$) controls which values from the input will be used to update the memory state. The forget gate (denoted by $f(t)$) determines which information should be discarded from the cell state. The output gate (denoted by $o(t)$) decides what part of the cell state should be output as the hidden state. The cell state ($c(t)$) is updated with the new information, which is regulated by the tanh activation function to add non-linearity.

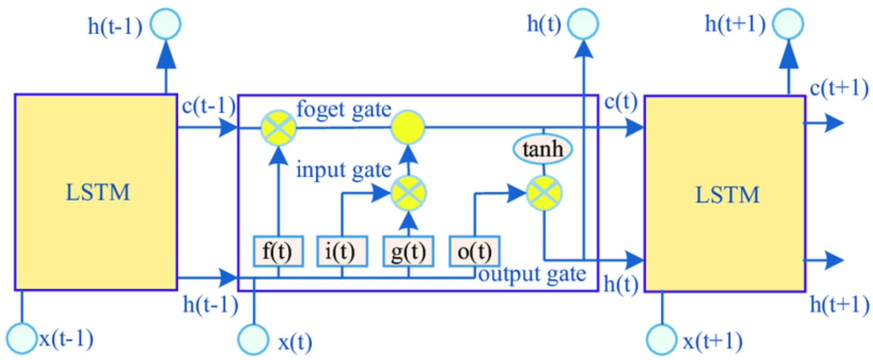

Figure 4.3: LSTM Network from (81)

Practical Implementation LSTM provides a robust framework for modeling sequential data, making it an ideal choice for time series forecasting tasks. Popular DL libraries like TensorFlow or PyTorch offer comprehensive modules for building and training LSTM models.

In practical implementation, the rolling window approach is utilized to ensure that the LSTM model is continuously updated with new data. This involves retraining the model on the most recent subset of the dataset, thereby allowing it to adapt to changing market conditions. The model is trained on a sliding window of past observations and predicts the next value in the sequence. This method effectively captures both short-term and long-term dependencies in the data.

Fine-tuning hyperparameters in LSTM models is pivotal for optimizing performance. Parameters such as the number of LSTM units, the learning rate, the dropout rate, and the batch size are commonly adjusted during the optimization process. Furthermore, exploring architectural variations such as stacked LSTM layers or bidirectional LSTMs can further improve model performance.

For instance, the following hyperparameters are commonly fine-tuned:

1. **Number of LSTM units:** Specifies the dimensionality of the output space.
2. **Learning rate:** Dictates the step size during the gradient descent optimization process.
3. **Dropout rate:** Helps prevent overfitting by randomly setting a fraction of input units to zero during each update in training.
4. **Batch size:** Refers to the number of samples processed in each gradient update.

Additionally, architectural variations such as stacked LSTM layers or bidirectional LSTMs can be explored to enhance the model's ability to capture complex patterns in the data.

4.2.3 Phase 3

Bidirectional Encoder Representations from Transformers

Theoretical Framework BERT, short for Bidirectional Encoder Representations from Transformers, is a state-of-the-art natural language processing (NLP) model introduced by (66). Unlike traditional NLP models that process text sequentially, either from left to right or right to left, BERT uses a bidirectional Transformer architecture, allowing it to consider context from both directions simultaneously. Consider the sentence "The bass was difficult to catch." Traditional NLP models might struggle to determine whether "bass" refers to a type of fish or a musical instrument. BERT, however, examines both the preceding context ("The") and the following context ("was difficult to catch") simultaneously, enabling it to understand that "bass" in this instance refers to the fish.

The objective in training language models is often to predict the next word in a sequence. BERT addresses this challenge using its key innovations. One of the key innovations of BERT is its pretraining strategy, which involves training the model on large amounts of unlabeled text data using two unsupervised tasks: masked language model (MLM) and next sentence prediction (NSP). Whilst MLM helps BERT understand sentence context, NSP helps BERT understand the relationship between pairs of sentences. Through pretraining on vast amounts of text data, BERT learns general language representations that can be fine-tuned for specific downstream tasks, such as text classification, question answering, and named entity recognition.

The architecture of BERT (see Figure 4.4) consists of multiple layers of self-attention mechanisms, which enable it to focus on different parts of the input sequence simultaneously. These self-attention mechanisms allow BERT to capture long-range dependencies and relationships between words in a sentence, leading to more contextually rich representations.

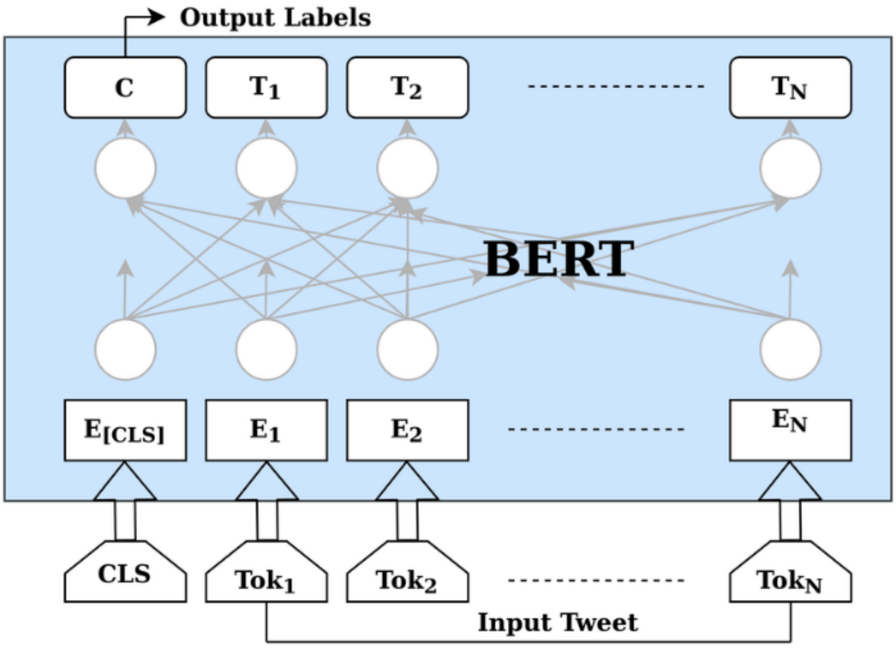

Figure 4.4: BERT architecture for sentiment analysis

As illustrated in Figure 4.4, the architecture includes: **Input Layer**: The input text is tokenized into tokens ($\text{Tok}_1, \text{Tok}_2, \ldots, \text{Tok}_N$) and embedded into vectors ($E_1, E_2, \ldots, E_N$). **Self-Attention Mechanism**: Each token attends to every other token, creating a set of attention scores that help the model weigh the importance of different words in the context. **Output Layer**: The final hidden state corresponding to the [CLS] token is used for classification tasks, predicting sentiment labels such as positive, negative, or neutral.

Practical Implementation In this practical framework, different BERT models will be implemented for sentiment analysis to evaluate their effectiveness and robustness. Since the models used are already pretrained and fine-tuned, the next step is using the best performing model to perform zero-shot learning on a test set. The model will predict the sentiment labels during the period 2015-2019 for multiple different stocks such as Apple.

Zero-shot learning will be employed to classify the sentiment of the text data without requiring additional training on the specific sentiment labels. The pre-trained BERT model, which has been fine-tuned on general sentiment analysis tasks, will be used to predict sentiment labels such as positive, negative, or neutral for each piece of text in the test set.

This approach leverages BERT's ability to understand context and semantic relationships within the text to make accurate sentiment predictions.

4.3 Portfolio Optimization

While traditional evaluation metrics such as MAE, MSE, RMSE, MAPE, and R-squared provide valuable insights into the accuracy and reliability of stock price prediction models, they have limitations when applied to real-world trading. These metrics do not inherently translate into profitable trading strategies, as they do not account for factors such as transaction costs, market conditions, risk management, and the ability to respond to market signals in real-time. Therefore, a comprehensive approach is needed to bridge this gap.

To address this, a rule-based trading simulation will be performed from February 2014 until February 2024 to assess the models in a more realistic and comprehensive manner. This simulation will evaluate the practical application of the models and their ability to generate profits. The focus will be on using predicted stock prices and sentiment predictions, updating portfolio weights daily, weekly, or monthly according to the optimization of the mean-variance portfolio. This will be performed on a subset of stocks (AAPL, AMZN, MSFT, GOOGL, TSLA) that have sentiment data available.

The mean-variance portfolio optimization, originally proposed by Harry Markowitz in 1952, aims to balance the trade-off between risk and return (82). It is calculated by minimizing the portfolio variance while achieving a desired level of expected return. For a portfolio of $N = 5$ stocks, the optimization involves the following steps:

- Expected Returns: Calculate the expected return for each stock based on historical data.

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t}$$

  where μ_i is the expected return of stock i, $R_{i,t}$ is the return of stock i at time t, and T is the number of time periods.

- Covariance Matrix: Compute the covariance matrix of the stock returns, which measures how the stocks move together.

$$\Sigma_{ij} = \frac{1}{T-1}\sum_{t=1}^{T}(R_{i,t} - \mu_i)(R_{j,t} - \mu_j)$$

  where Σ_{ij} is the covariance between stock i and stock j.

- Optimization Problem: Formulate the optimization problem to minimize the portfolio variance, subject to the constraint that the sum of the portfolio weights equals 1 and potentially additional constraints on the expected return.

$$\min_{w} w^T \Sigma w \quad \text{subject to} \quad \sum_{i=1}^{N} w_i = 1 \quad \text{and} \quad w^T \mu = \mu_p$$

  where w is the vector of portfolio weights, μ is the vector of expected returns, Σ is the covariance matrix, and μ_p is the desired portfolio return.

- Solver: Use numerical optimization techniques (e.g., the 'scipy.optimize.minimize' function) to solve the optimization problem and obtain the optimal portfolio weights.

By implementing this framework, the goal is to create a dynamic and responsive portfolio that adapts to changing market conditions and leverages both price and sentiment predictions to enhance performance. This comprehensive approach aims to provide a more realistic assessment of the models' practical utility in financial markets.

4.4 Evaluation Metrics

In this section, the evaluation metrics used to assess the performance of our models are highlighted. Different metrics are employed for stock price prediction models and sentiment classification models to ensure a comprehensive evaluation of their effectiveness and accuracy. It is important to note that in each of the evaluation metrics used, the variable n represents the number of predicted values. Table 4.1 provides an overview of these evaluation metrics.

For sentiment classification, BERT models are used to classify tweets into positive, neutral or negative sentiments. The evaluation metrics used for these models are summarized in table 4.2

Although metrics such as MAE, MSE, RMSE, MAPE, and R-squared are essential to assess the accuracy and reliability of stock price prediction models, they have limitations in the context of real-world trading. These metrics indicate how closely the model's predictions align with actual stock prices but do not inherently translate into profitable trading strategies.

To address this gap, evaluating the effectiveness of the portfolio optimization strategy becomes crucial. The effectiveness will be determined by the profitability of the portfolio based on price predictions, sentiment or a combination of both. This helps to identify the

Metric	**Explanation**	**Formula**
Mean Absolute Error (MAE)	Measures the average magnitude of the errors between the predicted and actual stock prices, without considering their direction. Lower MAE values indicate better performance.	$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert$
Mean Squared Error (MSE)	Measures the average squared difference between the predicted and actual stock prices. More sensitive to larger errors than MAE. Lower MSE values indicate better performance.	$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
Mean Absolute Percentage Error (MAPE)	Measures the average absolute percentage error between the predicted and actual stock prices. Lower MAPE values indicate better performance.	$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left\lvert \frac{y_i - \hat{y}_i}{y_i} \right\rvert \times 100$
R-squared (R^2) Score	Indicates the proportion of the variance in the dependent variable that can be explained by the independent variables. Higher R^2 values indicate better performance. Note that is the average value.	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Table 4.1: Evaluation Metrics for Stock Price Prediction Models

potential strengths and weaknesses of the models, offering a more holistic understanding of their practical utility in financial markets.

Metric	**Explanation**	**Formula**
Confusion Matrix	Provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, helping to understand classification performance in detail.	N/A
Accuracy	Measures the proportion of correctly classified instances out of the total instances. Higher values indicate better performance.	$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
Precision	Measures the proportion of true positive predictions out of all positive predictions. Higher values indicate a lower false positive rate.	$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$
Recall	Measures the proportion of true positive predictions out of all actual positive instances. Higher values indicate a lower false negative rate.	$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$
F1 Score	Harmonic mean of precision and recall, balancing both metrics. Higher values indicate better performance, especially for imbalanced datasets.	$\text{F1 Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

Table 4.2: Evaluation Metrics for Sentiment Classification

5

Evaluation

This chapter presents the obtained results. In addition, the evaluation results are used to justify the design choices and asses the contributions of different aspects in the design toward the overall goals. Detailed evaluation metrics for each phase and model are provided in the Appendix.

5.1 Phase 1 Results

As previously mentioned in Chapter 4, the methods employed in this phase are the Naive, LR, RF, SVR models. To provide a detailed analysis, the focus will be on a subset of five stocks: AAPL, AMZN, XOM, COST, and PEP. These stocks were selected based on their diversity in size, performance, and characteristics. To observe their performance, tables 13 to25 in the appendix will be used.

Benchmark Model First, the performance of the Naive model across all selected stocks improves as the window size decreases. Upon visual inspection of the Naive model's performance in predicting the stock price of AAPL in figure 5.1, it becomes clear that the predictions are a lagged version of the closing price. This is the case for each stock and aligns with the idea that the benchmark should be equal to yesterday's closing price. An example of the results can be seen in table 5.1, where it is visible that the evaluation metrics differ between stocks except for R-squared.

LR and RF These models perform decently well. Generally, the 250-window size provides the best balance between capturing recent trends and maintaining predictive accuracy. For example, AAPL and AMZN show slightly better performance with the 250-window size with regards to MSE, MAE and MAPE compared to shorter windows, indi-

Stock	MSE	MAE	MAPE	R²
AAPL	3.311	1.1107	0.0127	0.9990
AMZN	5.0869	1.4143	0.0144	0.9978
XOM	1.233	0.7738	0.0124	0.9970
COST	23.5173	2.9079	0.0094	0.9990
PEP	2.1062	0.9262	0.0077	0.9982

Table 5.1: Phase 1 Naive Model Performance Metrics (Window Size: 250)

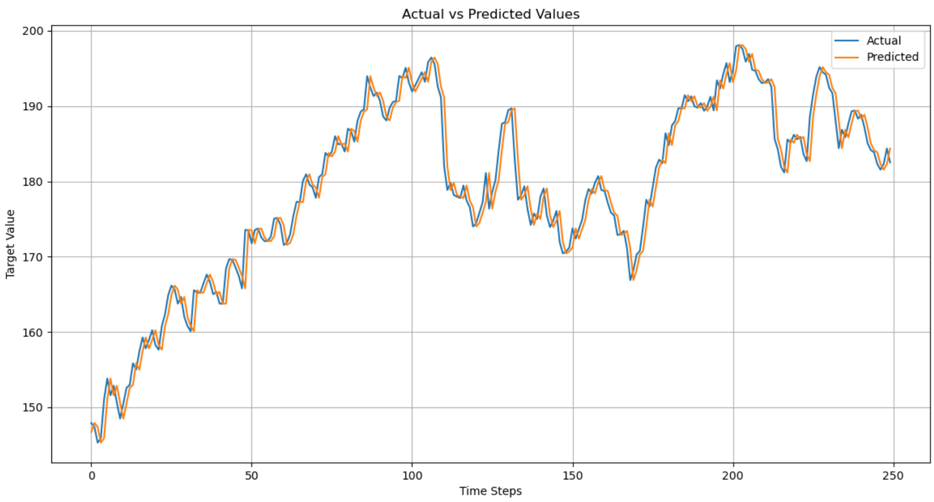

Figure 5.1: Phase 1 Naive Model Prediction AAPL

cating better prediction accuracy. Tables 5.2 and 5.3 show the results, from which can be seen that the LR model yields results very close to the benchmark, while the RF model shows some larger errors. Figures 5.2 and 5.3 show the predictions of the LR and RF models that perform the best for AAPL over the last year.

SVR The SVR model, however, performs significantly worse compared to the LR and RF models. The SVR model shows higher MSE and MAE values, indicating worse predictive accuracy. For instance, AAPL's MSE for the SVR model is 81.7553, see table 5.4, which is significantly higher than the MSE for the LR (3.311) and RF (6.1042) models. Although the SVR model improves with shorter window sizes, its performance remains inferior to the benchmark and the LR and RF models.

Ticker	MSE	MAE	MAPE	R^2
AAPL	3.5162	1.1582	0.0130	0.9989
AMZN	5.2929	1.4454	0.0147	0.9977
XOM	1.7177	0.9523	0.0129	0.9952
COST	29.6008	3.2595	0.0099	0.9988
PEP	2.6011	1.0562	0.0079	0.9968

Table 5.2: Phase 1 LR Performance Metrics (Window Size: 250)

Ticker	MSE	MAE	MAPE	R^2
AAPL	6.1042	1.5646	0.0178	0.9981
AMZN	10.9152	2.0324	0.0212	0.9953
XOM	3.4616	1.2691	0.0176	0.9903
COST	50.6807	4.4446	0.0133	0.9979
PEP	4.0883	1.3722	0.0102	0.9950

Table 5.3: Phase 1 RF Performance Metrics (Window Size: 250)

Performance Analysis There are a few important aspects to analyze before moving on, as they are likely to recur in the coming phases.

First, consider the window sizes. When comparing the results for different window sizes (500, 250, and 125) across various stocks, slight differences in evaluation scores are observed. The Naive Model's results indicate a higher sensitivity to recent stock volatility. For instance, stocks like AAPL and AMZN exhibit larger errors with larger window sizes, meaning that excluding a larger initial interval increases error. This suggests that errors towards the end of the interval (2024) are larger than at the start of the interval (2014).

A similar trend is observed with other models. Generally, smaller window sizes yield better results. However, for machine learning models, smaller windows mean the models train on less data, focusing more on short-term price movements, whereas larger windows capture long-term trends. Short-term price movements generally yield better results. Nonetheless, exceptions exist, such as XOM, which performs better with larger window sizes. This indicates that for certain stocks, capturing long-term trends provides more accurate predictions.

Second, despite showing predictions close to the trend and having a high R-squared score, many predictions are quite inaccurate. This means that having a high R-squared value does not inherently mean accurate predictions. For example, the worst-performing model,

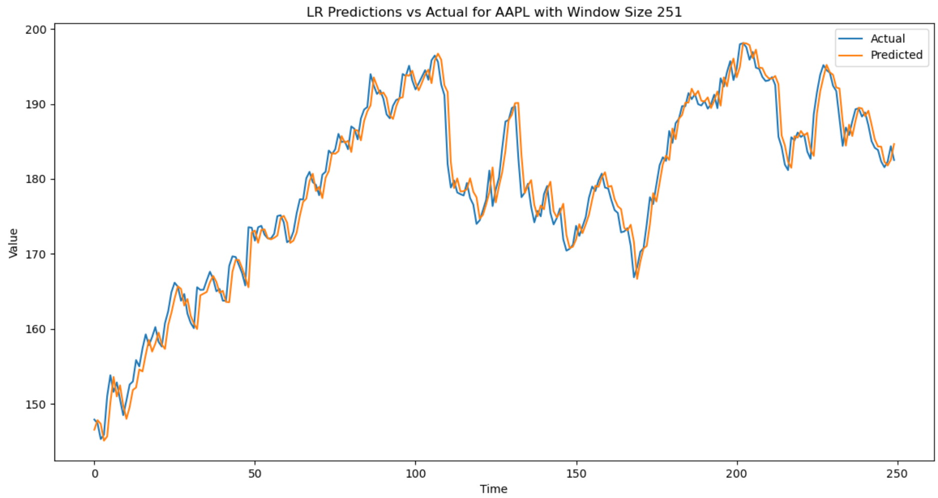

Figure 5.2: Phase 1 LR Model Prediction AAPL (Window Size: 250)

Ticker	MSE	MAE	MAPE	R^2
AAPL	81.7553	7.4247	0.1012	0.9749
AMZN	79.7812	7.4876	0.1039	0.9660
XOM	26.6401	4.3115	0.0597	0.9251
COST	906.3352	23.5585	0.0747	0.9621
PEP	46.4219	5.8901	0.0446	0.9427

Table 5.4: Phase 1 SVR Performance Metrics (Window Size: 250)

SVR, often yields R-squared values above 0.90, which would suggest a strong relationship between the variables and indicate that the model provides a good fit to the data. However, this is not the case, as figure 5.4 shows. Therefore, MSE, MAE, and MAPE are more relevant evaluation metrics to consider.

5.2 Phase 2 Results

In Phase 2 of the analysis, the expanded dataset is used as previously mentioned in chapter 3 and 4. This phase aims to assess the impact of these enriched features on the model's predictive performance. The methods employed in this phase are RF, SVR and LSTM. To provide a detailed analysis, the focus will again be on a subset of five stocks: AAPL, AMZN, XOM, COST, and PEP. To observe their performance, tables 26 until 34 from the

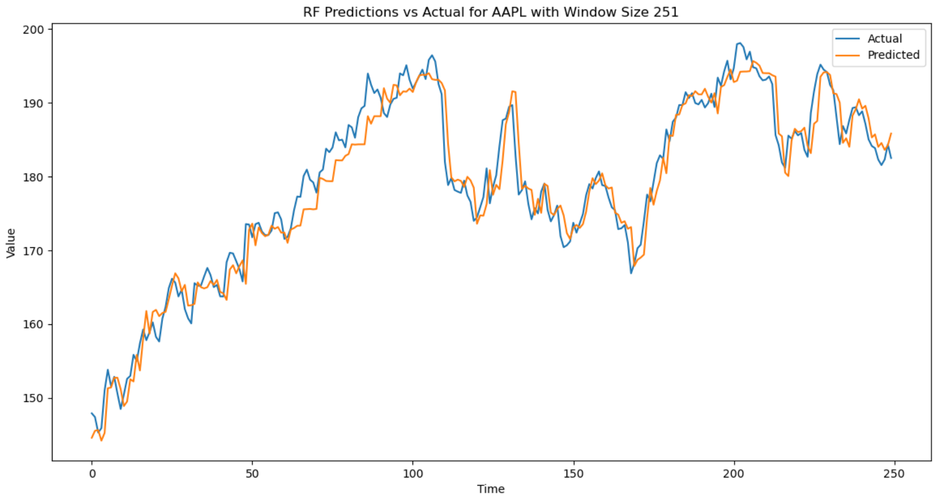

Figure 5.3: Phase 1 RF Model Prediction AAPL (Window Size: 250)

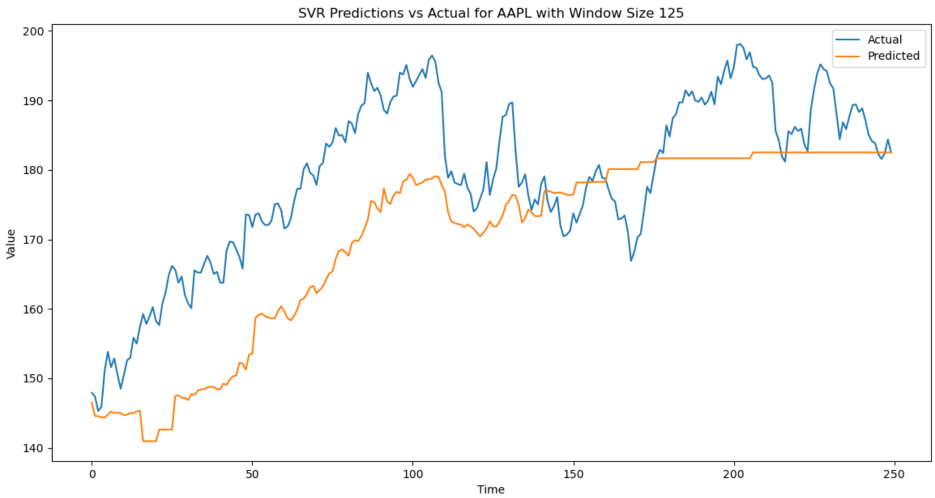

Figure 5.4: Phase 1 SVR Model Prediction AAPL (Window Size: 125)

appendix will be used.

RF The RF model shows decent predictive performance, although it tends to have larger errors compared to results of the RF and LR models in phase 1. The performance metrics in table 5.5 illustrate that for example AAPL has an MSE of 6.5295, MAE of 1.6447, and MAPE of 0.0189. However, RF's larger error margins compared to phase 1 models can be attributed to its increased complexity by adding multiple different features to the dataset. Moreover, figure 5.5 shows the best performing AAPL predictions over the entire interval.

Ticker	MSE	MAE	MAPE	R^2
AAPL	6.5295	1.6447	0.0189	0.9980
AMZN	11.8105	2.1069	0.0221	0.9950
XOM	4.1174	1.3507	0.0189	0.9884
COST	64.4071	5.0205	0.0147	0.9973
PEP	4.1084	1.3719	0.0102	0.9949

Table 5.5: Phase 2 RF Performance Metrics (Window Size: 250)

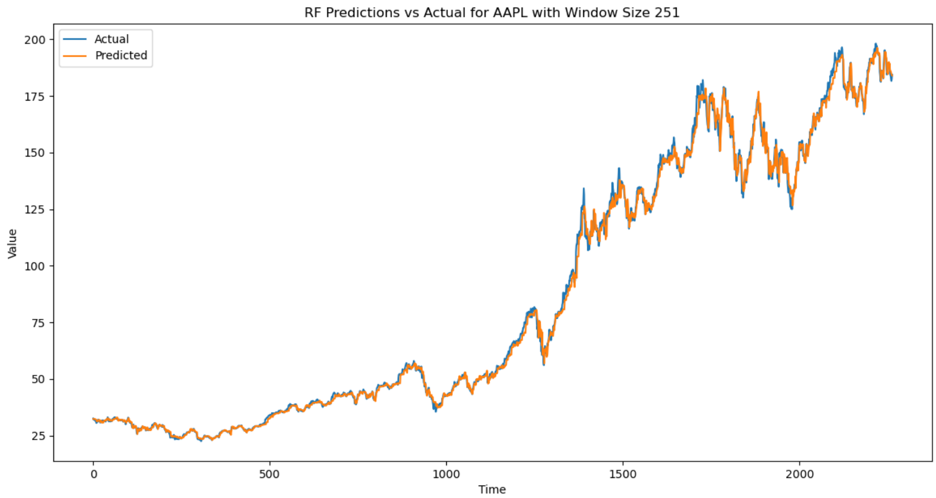

Figure 5.5: Phase 2 RF Prediction AAPL (Window Size: 250)

SVR The SVR model performs, just as in phase 1, significantly worse. As shown in table 5.6, the MSE and MAE values for SVR are substantially higher. For example, AAPL's MSE for SVR is 176.7071, which is significantly higher than both LR (3.5162) and RF (6.5295). Moreover, figure 5.6 shows the best performing AAPL predictions over the entire interval.

LSTM The LSTM model performs, when looking at the evalutation metrics, better than the SVR model. However, its performance is worse than that of the RF model and the

Ticker	MSE	MAE	MAPE	R^2
AAPL	176.7071	10.1106	0.1206	0.9457
AMZN	228.3505	11.8663	0.1379	0.9026
XOM	72.6980	6.8700	0.0995	0.7955
COST	1384.1002	28.1580	0.0834	0.9419
PEP	67.5782	6.8292	0.0507	0.9166

Table 5.6: Phase 2 SVR Performance Metrics (Window Size: 250)

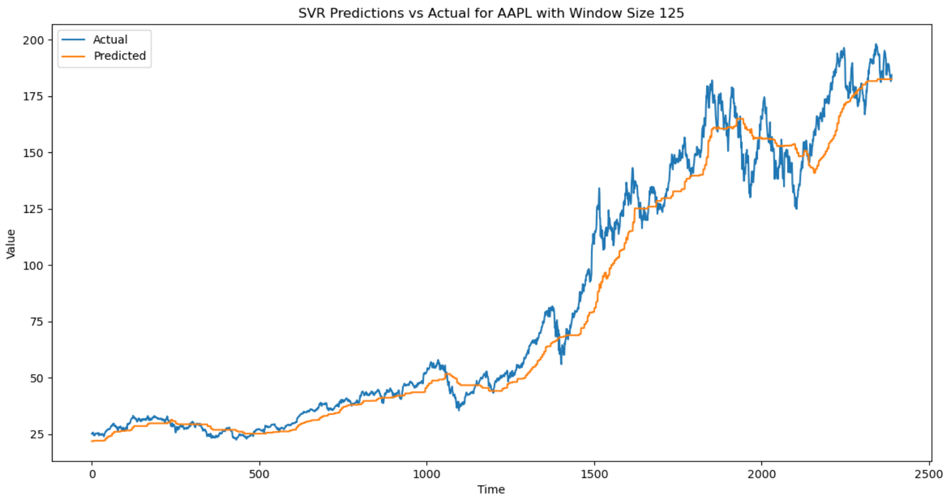

Figure 5.6: Phase 2 SVR Prediction AAPL (Window Size: 125)

benchmark. Note that the performance of the LSTM model increases as the window size decreases, possibly suggesting that capturing short-term price movements is preferable. Table 5.7 shows the evaluation metrics for the selected subset of stocks. Upon examining Figure 5.7, it is evident that, while the LSTM model captures the general trend of AAPL's stock price movements, there are notable fluctuations in the predictions that the model does not capture. These fluctuations are particularly pronounced at the beginning of the time series.

Performance Analysis First, despite its promising performance in phase 1 the RF model in this phase did not outperform its benchmark in phase 1. This could be because of the increased complexity of the RF model in this phase, due to the inclusion of multiple different features in the dataset. Moreover, more data could also have introduces more

Ticker	MSE	MAE	MAPE	R^2
AAPL	50.98	4.71	0.0658	0.9843
AMZN	55.29	5.41	0.0751	0.9764
XOM	13.98	2.79	0.0360	0.9606
COST	415.57	14.14	0.0509	0.9825
PEP	20.31	3.39	0.0261	0.9749

Table 5.7: Phase 2 LSTM Performance Metrics (Window Size: 250)

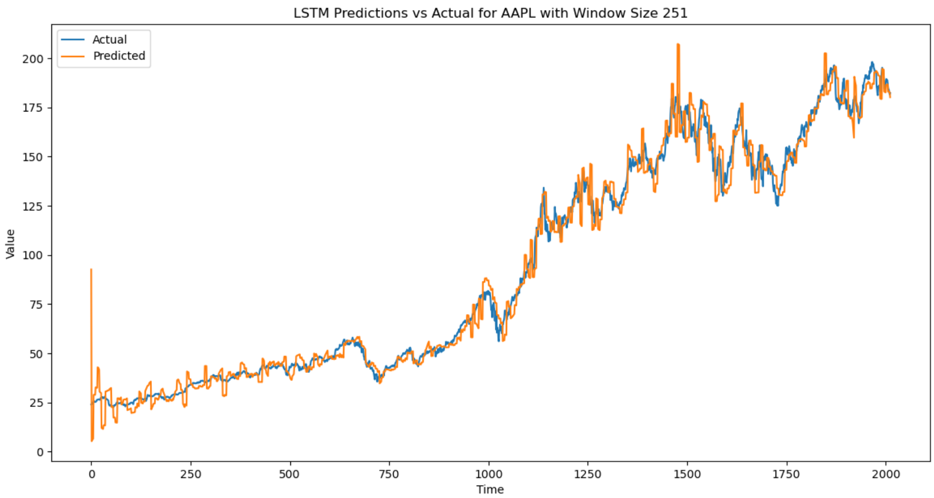

Figure 5.7: Phase 2 LSTM Prediction AAPL (Window Size: 250)

variablility and noise, especially since the data includes periods of high volatility. The model might learn noise in the training data rather than the underlying patterns, which leads to overfitting and less accurate predictions. However, cross validation was used to prevent the overfitting.

Second, as the model was already underperforming in phase 1 the SVR model did not perform better in phase 2. SVR models might struggle with the non-linearity and complexity of stock price movements, suggesting that the added data and its complexity, volatility and noise cannot be captured by SVR. Especially when compared to models like RF that can better capture complex interactions between features through ensemble learning. The inherent assumptions of SVR regarding the linearity of relationships might not hold true for financial time series data, which often exhibit non-linear patterns and volatility clustering. This can lead to significant prediction errors, as reflected in the high MSE and MAE values.

Third, as previously mentioned, the LSTM predictions exhibit significant fluctuations. These fluctuations are particularly pronounced at the beginning of the time series and during periods of high volatility. Several factors could contribute to this behavior:

A **"warm-up" period**, the starting period of a time-series forecast where the model is still adjusting to patterns. But it would not explain fluctuations apart from the start of the interval.

The **window size** of 251 seems to balance capturing long-term trends and short-term fluctuations. Nevertheless, the model's performance improves with smaller window sizes, likely due to the LSTM's ability to capture shorter-term dependencies more effectively. Figure 5.8 illustrates predictions with a smaller window size, where fluctuations decrease but still occur throughout the interval.

Pronounced fluctuations at the start and during high **volatility** periods suggest that the LSTM model is highly sensitive to rapid stock price changes. This sensitivity can lead to overreactive predictions and higher error metrics. Limited data exacerbates this volatility sensitivity, as the model struggles to generalize patterns effectively without sufficient examples of various market conditions, resulting in greater prediction variability.

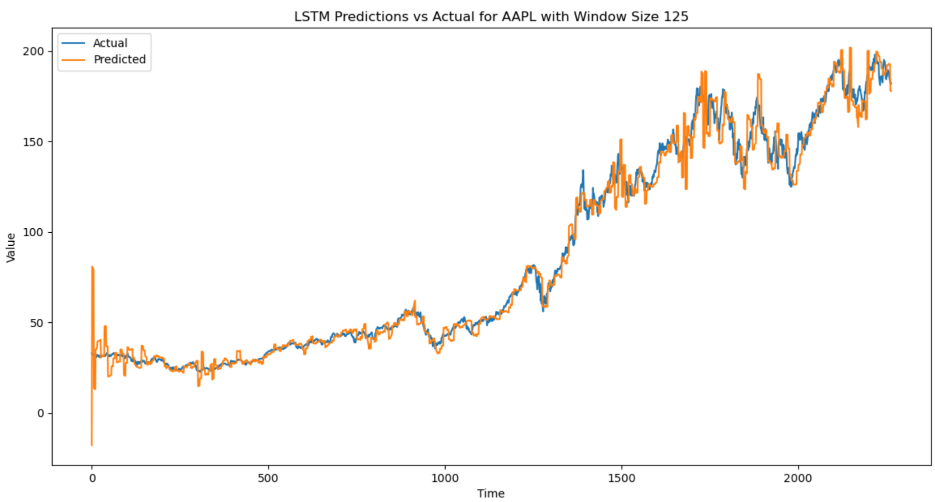

Figure 5.8: Phase 2 LSTM Prediction AAPL (Window Size: 125)

5.3 Phase 3 Results

Sentiment Models Before moving on to the stock price prediction results the performance of the sentiment classification models needs to be adressed. The sentiment classification models, including BERT, RoBERTa, DistilBERT, FinancialBERT and FinBERT, were evaluated to compare their performance on sentiment analysis tasks. To observe their performance, tables 3 to 12 in the appendix will be used.

First, tables 3, 5, 7, 9, 11 show the precision, recall, F1 score, and support of each model predictions. Table 5.8 presents the accuracy and F1 scores for all models. It can be observed that the transformer pre-trained on large financial corpus obtain the highest accuracy and F1 scores, outperforming other transformer models. Among the models that were not pre-trained on financial corpus RoBERTA performs the best while DistilBERT performs the worst.

Taking a closer look at the results obtained from the confusion matrices it becomes clear that BERT mainly predicts neutral and never predicts positive (4). RoBERTa predicts mainly neutral with a promising number of positive predictions and only some negatives and DistilBERT only predicts negative (6, 8). Finbert and FinancialBERT have similar predictions, comparing these transformer models it can be seen that finBERT predicts neutral more often causing some errors while FinancialBERT predicts each class accurately with very small errors (10, 12).

Model	**Accuracy**	**F1 Score**
BERT	61%	47%
RoBERTa	70%	65%
DistilBERT	13%	3%
FinBERT	98%	98%
FinancialBERT	99%	99%

Table 5.8: Summary of Model Performance

In summary, from the confusion matrices and performance metrics, it can be observed that specialized models like FinBERT and FinancialBERT significantly outperform general-purpose models like BERT and its variants. The main reasons for the improved performance of FinBERT and FinancialBERT include their fine-tuning on specific domains, which allows them to capture domain-specific sentiment nuances more effectively.

In contrast, general-purpose models like BERT, RoBERTa, and DistilBERT show varying degrees of success, with RoBERTa performing the best among them. DistilBERT's performance is notably poor, likely due to the compromises made for efficiency. This analysis underscores the importance of domain-specific fine-tuning in achieving high accuracy in sentiment analysis tasks.

In Phase 3 of the analysis, the expanded data set includes sentiment analysis as an additional feature, as previously mentioned in Chapters 3 and 4. The sentiment data ranges from 2015 until 2019, thereby the results will be focussed on a subset of the entire period. Moreover, the sentiment data is only applicable to 5 specific stock AAPL, AMZN, MSFT, GOOGL and TSLA. This phase aims to assess the impact of these enriched features on the model's predictive performance. The methods employed in this phase are the Naive Model, RF, SVR, and LSTM networks. To observe their performance, tables 35 until 38 from the appendix will be used.

Benchmark/Naive Looking at the results of the benchmark model within this subset of 2015 to 2019 it can be noticed that the MSE and MAE are relatively low, indicating that the Naive model is performing well for this simple approach in this period. However, the different MAPE values suggest that there is some variation in prediction accuracy across different stocks. In addition, the same effect is noticed as in phase 1, where the naive model performs better the smaller the window size is.

Ticker	MSE	MAE	MAPE	R^2
AAPL	0.4237	0.4393	0.0106	0.9967
AMZN	1.4317	0.7664	0.0119	0.9974
GOOG	0.5328	0.4913	0.0098	0.9938
MSFT	1.689	0.8744	0.0097	0.9981
TSLA	0.3044	0.3800	0.0204	0.9794

Table 5.9: Phase 3 Naive Model Performance Metrics (Window Size: 250)

RF The Random Forest (RF) model demonstrates decent performance across all tickers with a window size of 250. The MSE, MAE, and MAPE values are low, indicating precise predictions with minimal error. This suggests that the RF model is somewhat effective in capturing the underlying patterns in the stock data, but not as effective as using the benchmark model. Figure 5.9 shows the predictions of the RF model for AAPL over the interval 2015 until 2019.

Ticker	MSE	MAE	MAPE	R^2
AAPL	1.0096	0.7094	0.0168	0.9922
AMZN	2.4462	1.0827	0.0168	0.9956
GOOG	0.9082	0.6842	0.0136	0.9895
MSFT	3.0098	1.2610	0.0138	0.9967
TSLA	0.5542	0.5285	0.0287	0.9627

Table 5.10: Phase 3 RF Performance Metrics (Window Size: 250)

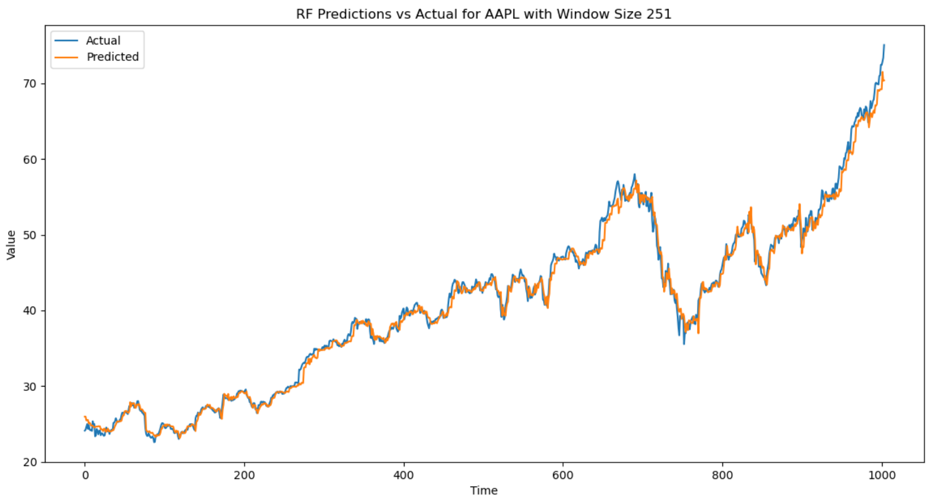

Figure 5.9: Phase 3 RF Prediction AAPL (Window Size: 250)

SVR The SVR model exhibits relatively higher error metrics compared to the Naive and RF models. The MSE and MAE values are notably higher, particularly for AMZN and MSFT, indicating less accurate predictions. The R^2 values, while decent, show that the SVR model is less effective in capturing the variance in the data. Figure 5.10 shows the predictions of the SVR model for AAPL over the interval 2015 until 2019.

LSTM The LSTM model's performance is better than the SVR model but definitely worse than the RF and Naive model. The MSE and MAE values are comparatively higher

Ticker	MSE	MAE	MAPE	R^2
AAPL	19.226	3.873	0.0982	0.8505
AMZN	63.616	6.768	0.1176	0.8862
GOOG	16.030	3.581	0.0744	0.8153
MSFT	120.980	9.874	0.1082	0.8675
TSLA	1.658	1.043	0.0584	0.8884

Table 5.11: Phase 3 SVR Performance Metrics (Window Size: 250)

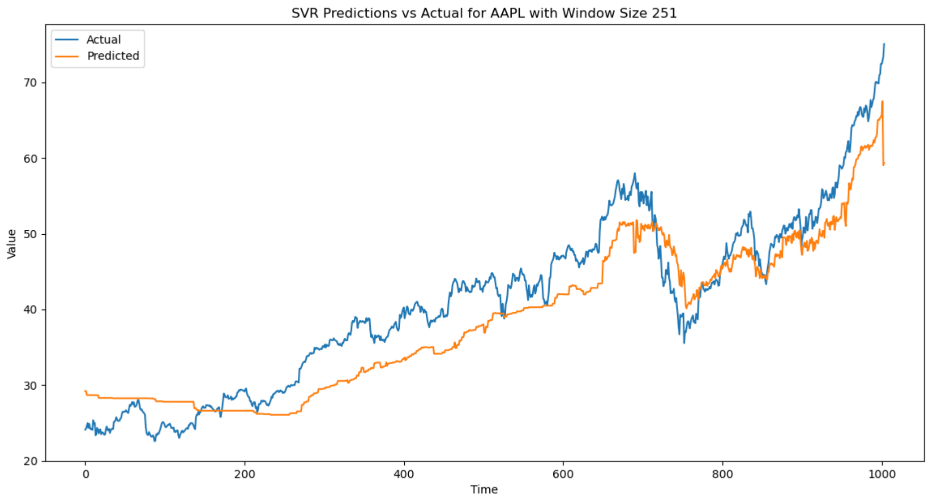

Figure 5.10: Phase 3 SVR Prediction AAPL (Window Size: 250)

than those of the RF model. This indicates that while LSTM can capture temporal dependencies well, it might not be as precise in terms of error metrics for this dataset. Further tuning or alternative configurations might improve its performance. Figure 5.11 shows that the LSTM model captures the general trend but it fails in accuratly predicting the stock price. The same phenomenon appears from phase 2 appears where the LSTM predictions highly fluctuate.

Perfomance Analysis In order to fully analyze the effect of sentiment, it is informative to compare the results with and without the added features on this interval subset. The results of the models' performance on the subset without sentiment can be seen in tables 39 to 41. Looking at the results from these tables, it becomes clear that not adding sentiment yields mixed results. For example, the RF model shows that in the case of

Ticker	MSE	MAE	MAPE	R^2
AAPL	9.668	2.177	0.0492	0.8689
AMZN	26.611	3.807	0.0560	0.9284
GOOG	6.172	1.992	0.0378	0.8665
MSFT	41.335	4.763	0.0502	0.9394
TSLA	2.099	1.150	0.0577	0.7834

Table 5.12: Phase 3 LSTM Performance Metrics (Window Size: 250)

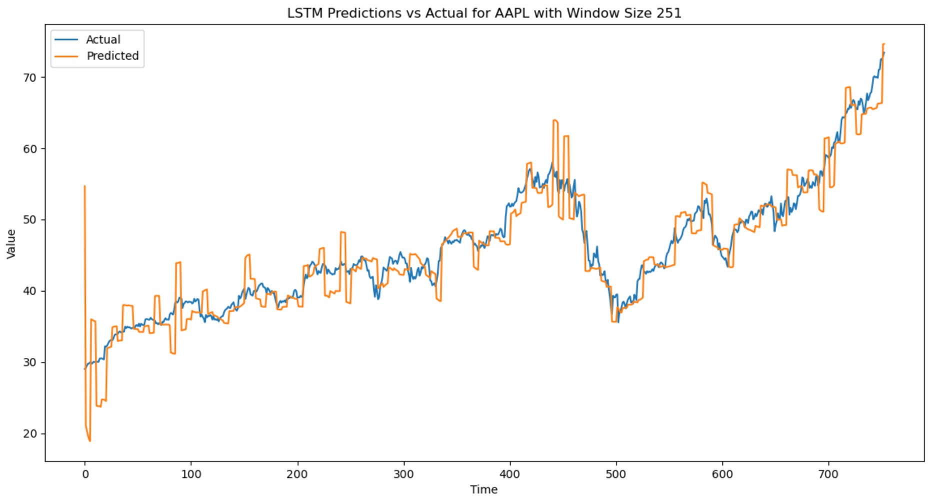

Figure 5.11: Phase 3 LSTM Prediction AAPL (Window Size: 250)

TSLA, the MSE, MAE, and MAPE are lower with the added sentiment, while the other stocks show better performance without the sentiment features. The SVR model performs better overall without the added sentiment data, and the LSTM model performs better with the added sentiment data except for MSFT.

This mixed performance suggests that the impact of sentiment data on model accuracy varies depending on the stock and the model used. For some stocks, sentiment data provides additional context that helps improve the predictive performance of models, particularly for more complex models like LSTM. For other stocks, the inclusion of sentiment data may introduce additional noise or complexity that models like SVR are unable to effectively utilize, leading to reduced performance.

In summary, the inclusion of sentiment data can enhance model performance in certain cases, particularly for models and stocks that can leverage this additional context to better understand market trends and investor sentiment. However, the benefit is not universal across all models and stocks, highlighting the need for careful consideration of feature selection based on the specific application and characteristics of the dataset.

5.4 Portfolio Optimization

As mentioned above in Chapter 4, evaluation metrics such as MAE, MSE, RMSE, MAPE, and R-squared are essential for assessing the performance of stock price prediction models; however, it is crucial to understand their limitations in the context of real-world trading. The obtained predicted prices and sentiment labels will be asses using a rule-based trading strategy. This is done for a subset of stock since the sentiment data is only applicable to five specific stocks AAPL, AMZN, MSFT, GOOGL and, TSLA and the period January 2015 to December 2019.

The process includes the following steps: First, the portfolio weights will be determined using the mean-variance portfolio optimization technique. Second, using predicted prices or sentiment labels, a prediction of whether the stock price is expected to increase or decrease will be made. Based on this decision from either the predicted stock price or sentiment labels the final step uses this information to determine when to buy or sell stock, specifically the stock in the subset. This portfolio optimization makes is such that the usage of predicted prices and sentiment will be assessed in the context of real-world trading.

This section is split into two parts; the first part employs a single, predetermined portfolio weight that will be determined on the first trading day. The second part instead rebalances the portfolio using daily, weekly, or monthly intervals.

5.4.1 Pre-Determined Portfolio Weights

First, the portfolio strategy that leverages price predictions will be assessed based on the graph shown in Figure 5.12. In the initial period (2015-2016), the portfolio value dropped below the benchmark. Moreover, in 2016 both the portfolio and, benchmark showed some volatility with the benchmark performing relatively better. During 2016-2018 the benchmark maintained a higher value compared to the strategy portfolio although both are showing fluctuations. Toward the end (2018-2020), a notable increase in the strategy portfolio can be observed, surpassing the benchmark by a substantial margin. This upward trend continued and showed signs of stabilization, maintaining its higher value. The benchmark also experienced growth, but at a slower rate compared to the strategy portfolio. The strategy portfolio has a value of $28,929 while the benchmark portfolio has a value of $16,028 at the end, showing a return of 189.3% and 60.28%, respectively.

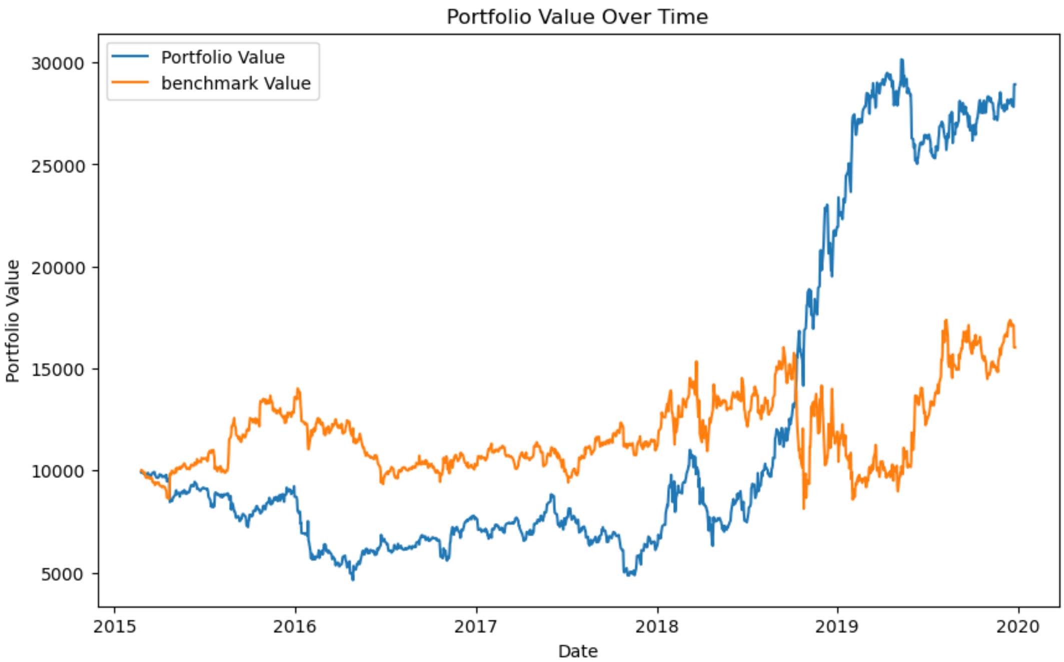

Figure 5.12: Portfolio Value using Price Predictions

Second, the portfolio strategy that leverages sentiment predictions will be assessed in figure 5.13. In the initial period (2015-2016), the portfolio value started closely aligned with the benchmark but soon dropped below it. Throughout 2016, both the portfolio and the benchmark experienced some volatility, with the benchmark performing relatively better. From 2016 to 2018, the benchmark consistently maintained a higher value compared to the strategy portfolio, although both exhibited fluctuations. However, the portfolio strategy showed a significant decline during this period, underperforming relative to the benchmark. In the late period (2018-2020), the strategy portfolio continued to struggle, experiencing further declines and even dipping into negative territory. By the end of the period, the benchmark portfolio reached a higher value ($16,028), while the strategy portfolio remained significantly lower and unstable (-$4,665). In summary, the portfolio and the benchmark showed a return of -146.65% and 60.28%, respectively.

Third, the portfolio strategy that leverages both price predictions and sentiment predictions will be assessed in figure 5.13. In the initial period (2015-2016), the portfolio value

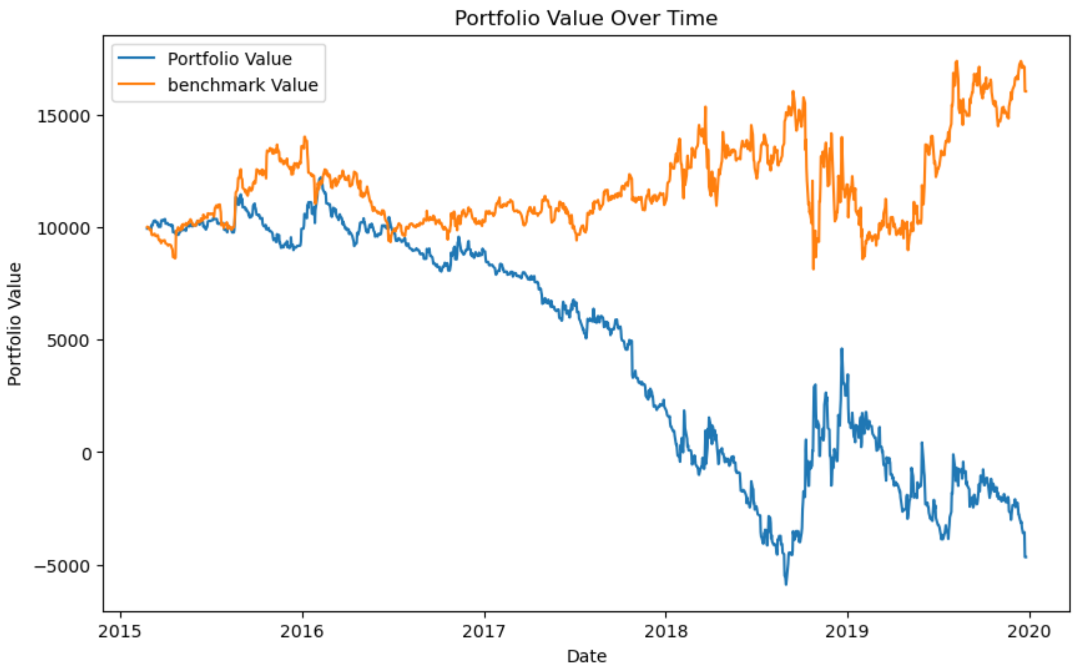

Figure 5.13: Portfolio Value using Sentiment Predictions

re-emerged closely aligned with the benchmark, but soon dropped below it. Throughout 2016, both the portfolio and the benchmark experienced some volatility, with the benchmark performing relatively better. From 2016 to 2018, the benchmark consistently maintained a higher value compared to the strategy portfolio, although both exhibited fluctuations. However, the portfolio strategy showed a significant decline during this period, underperforming relative to the benchmark. In the late period (2018-2020), the strategy portfolio continued to struggle, experiencing further declines and even dipping into negative territory. In contrast, the benchmark showed a more stable and upward trend, continuing to grow at a steady rate. By the end of the period, the benchmark portfolio reached a higher value, while the strategy portfolio remained significantly lower and unstable. At the end of the period, the strategy portfolio had a value of -\$12,160 indicating a loss, while the benchmark portfolio had a value of \$16,028, reflecting growth. In summary, the portfolio and the benchmark showed a return of -221.60% and 60.28%, respectively.

5.4.2 Rebalancing Portfolio Weights

In this subsection, the focus will be on the portfolio strategy that performed the best in the previous analysis. The strategy using price predictions consistently outperformed the other approaches. Although rebalancing portfolio weights resulted in less substantial

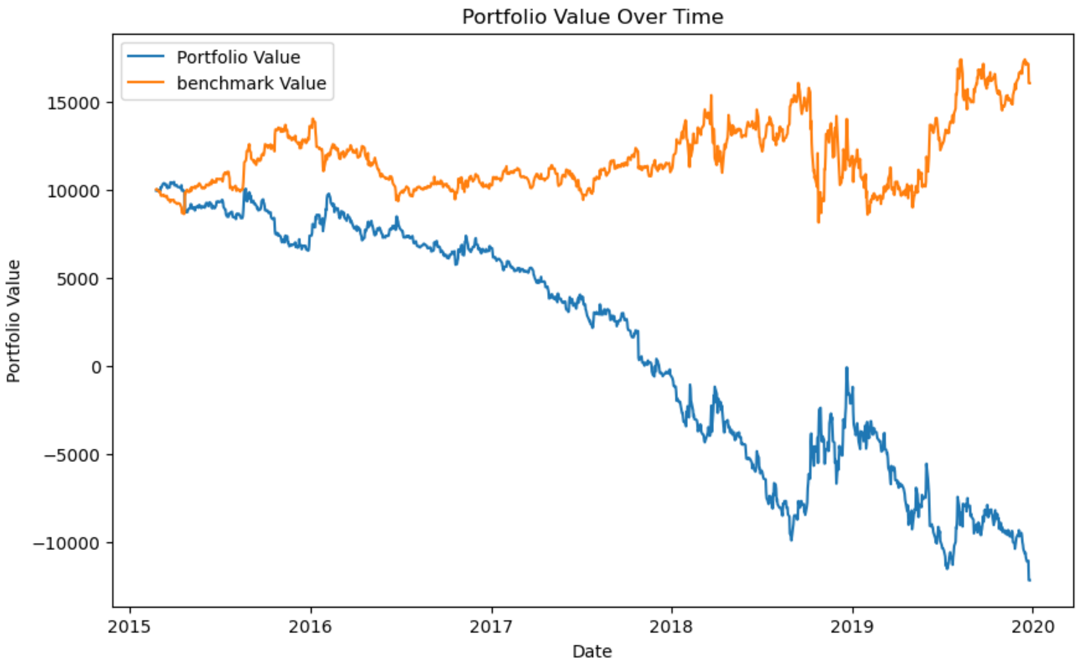

Figure 5.14: Portfolio Value using Price- and Sentiment Predictions

losses compared to sentiment-based and combined price and sentiment approaches, the price prediction strategy demonstrated superior overall performance. This highlights its potential effectiveness in portfolio management.

Figures 5.15, 5.16 and, 5.17 illustrate the results of using daily, weekly, and monthly intervals for rebalancing the portfolio weights through mean-variance portfolio optimization

The daily rebalancing strategy demonstrated impressive performance, achieving a strategy portfolio value of $13,866.04 compared to the benchmark portfolio value of $10,903.08. This translates to a total return of 38.66% for the strategy portfolio, significantly outpacing the benchmark's return of 9.03%. The consistent adjustment of portfolio weights on a daily basis allowed the strategy to capitalize on short-term market movements, resulting in superior returns.

Similarly, the weekly rebalancing approach also showed strong results, with the strategy portfolio reaching a value of $14,029.69 against the benchmark's $11,024.02. The total return for the strategy portfolio was 40.30%, slightly higher than the daily rebalancing interval, while the benchmark achieved a 10.24% return. This indicates that the weekly interval was particularly effective, offering the highest return among the three rebalancing frequencies. The balance between capturing market trends and minimizing transaction

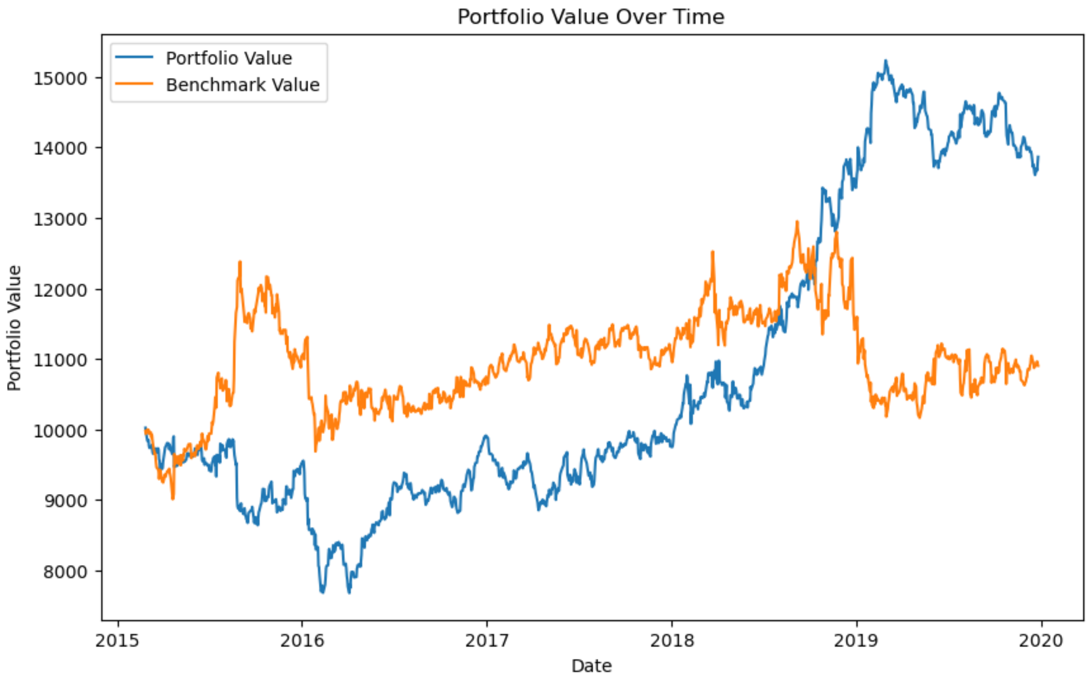

Figure 5.15: Portfolio Value using Daily Rebalancing

costs may have contributed to its outstanding performance.

On the other hand, the monthly rebalancing strategy, while still positive, resulted in a lower total return of 15.30% with a strategy portfolio value of $11,530.39. Despite this, it still outperformed the benchmark, which had a portfolio value of $9,008.74 and a negative return of -9.91%. The less frequent rebalancing may have led to missed opportunities for capturing shorter-term market gains, but it also avoided frequent transaction costs, maintaining a solid performance relative to the benchmark.

In summary, the strategy portfolios with daily and weekly rebalancing intervals performed substantially better than the benchmark, with the weekly rebalancing interval achieving the highest total return. The daily rebalancing interval also showed significant outperformance, highlighting the benefits of frequent adjustments to portfolio weights. Although the monthly rebalancing interval yielded lower returns, it still managed to outperform the benchmark by a considerable margin, underscoring the overall effectiveness of the mean-variance optimization approach.

When these results are compared with the initial portfolio strategy that utilizes price predictions, it is evident that the rebalancing frequency plays a critical role in portfolio performance. The price prediction strategy, with its rebalancing approach, demonstrated

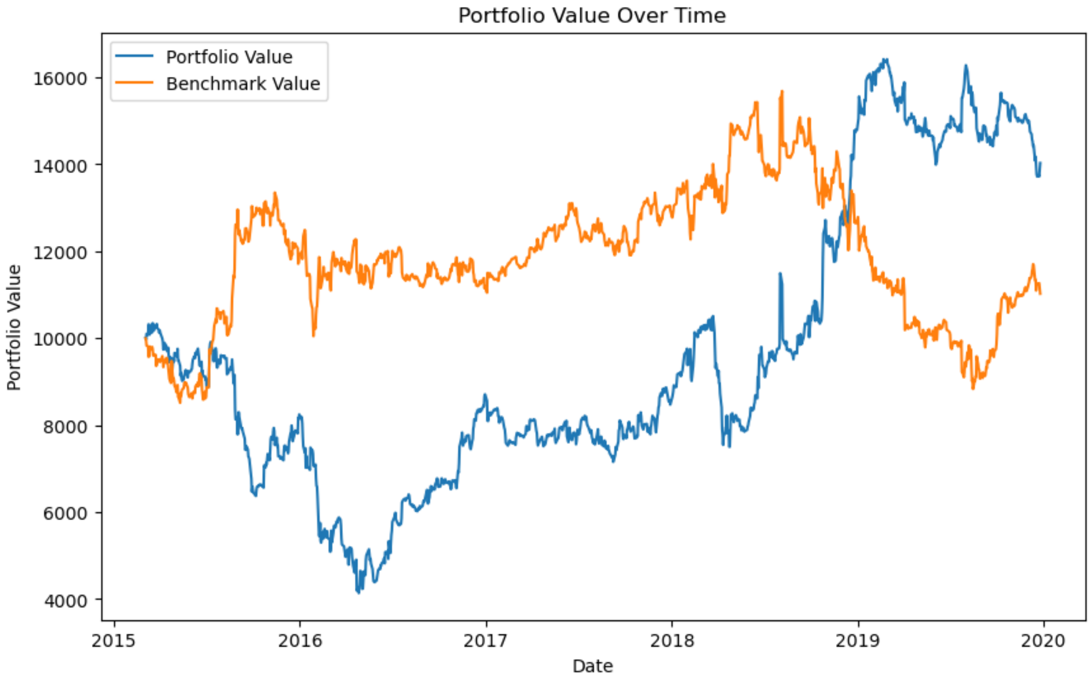

Figure 5.16: Portfolio Value using weekly Rebalancing

superior returns in the period 2015-2018, particularly when rebalanced weekly. During this period the rebalancing approach yielded smaller decrease in portfolio value compared to the initial portfolio strategy. However, the increase from 2018 until 2020 was not large in the rebalancing approach compared to the initial strategy. These findings highlight the importance of frequent rebalancing in optimizing portfolio performance and maximizing returns, as shown in the performance of the strategy portfolio across different rebalancing intervals.

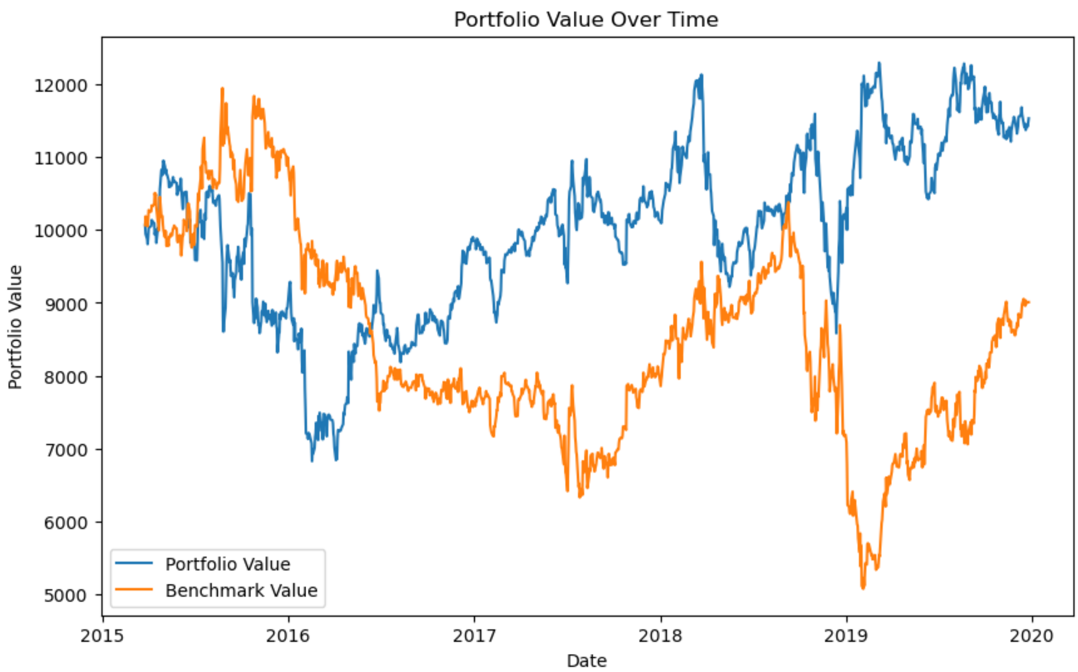

Figure 5.17: Portfolio Value using Monthly Rebalancing

6

Conclusion

6.1 Recapitulation of Objectives and Methods

This study aimed to enhance the accuracy of stock market forecasts by integrating diverse data sources and sophisticated analytical methods. The primary objective was to develop a model capable of accurately predicting stock prices for the top 25 large-cap stocks in the SP 500. These objectives converged in the research question: "How does incorporating multiple data sources, different Machine- /Deep Learning techniques, and sentiment analysis with Natural Language Processing enhance the accuracy of stock price predictions?"

A secondary objective was to test the validity of traditional financial theories, such as the EMH and RW theory, by incorporating these various data sources. Through this multifaceted approach, the research aimed not only to refine stock price predictions but also to assess the ongoing relevance of established financial theories.

6.2 Summary of Key Findings

This research explored the predictive performance of various models in forecasting stock prices across three distinct phases, aiming to enhance accuracy and robustness by incorporating diverse data sources. In all phases a rolling window approach was employed to mitigate lookahead bias, testing various window sizes (6 months, 1 year, 2 years) to capture short-term versus long-term trends. In **Phase 1**, using basic stock data (open, high, low, close, volume) with models like Naive, LR, RF, and SVR, simpler models like Naive and LR performed well, while SVR lagged significantly. In **Phase 2**, adding technical indicators and fundamental data, the models used were RF, SVR, and LSTM networks. Despite the increased dataset complexity, RF did not improve over Phase 1, SVR continued to underperform, and LSTM showed potential but had significant fluctuations. In **Phase 3**, incorporating sentiment analysis data, models used were Naive, RF, SVR, and LSTM. Sentiment data yielded mixed results; it improved LSTM's performance for specific stocks but introduced noise for others, particularly SVR.

The **Trading Simulation** underscored the importance of evaluation metrics beyond MAE, MSE, RMSE, MAPE, and R-squared. Key findings include: Pre-determined portfolio weights with price predictions showed substantial growth, especially from 2018 to 2020, with a return of 189.3% versus the benchmark's 60.28%. The sentiment prediction strategy underperformed at -146.65%, and combining the price and sentiment predictions resulted in -221.60%. For rebalancing portfolio weights, price prediction strategies outperformed others. Daily rebalancing achieved a return of 38.66%, weekly rebalancing 40.30%, and monthly rebalancing 15.30%, all exceeding the benchmarks. These results highlight the effectiveness of frequent rebalancing and the superior performance of price prediction strategies over sentiment-based approaches in portfolio management. Although sentiment analysis can improve model performance in specific cases, it can also introduce volatility and noise, affecting overall predictive accuracy.

6.3 Interpretation of Results

The results of this study have significant implications for the prediction of stock prices and trading strategies. They highlight the strengths and limitations of various predictive models and data sources, offering valuable insights for both academic research and practical applications.

The superior performance of simpler models like Naive and LR in Phase 1 suggests that basic stock data can still be highly effective for forecasting purposes. This aligns with the existing literature that emphasizes the utility of straightforward models under certain market conditions (e.g. (1)). However, the underperformance of the SVR model, even with enhanced datasets, indicates that more complex models may not always produce better predictions, diverging from theoretical expectations that increased complexity should improve accuracy (83).

In Phase 2, the addition of technical indicators and fundamental data did not significantly enhance the performance of the Random Forest model, challenging the assumption that incorporating more data inherently leads to better predictions (77). The Long Short-Term Memory (LSTM) model showed potential but faced instability, suggesting that while advanced neural networks can capture complex patterns, they require careful tuning and robust training data (80).

Phase 3's inclusion of sentiment analysis data yielded mixed results. Although sentiment data improved the performance of the LSTM for certain stocks, they introduced noise for others, particularly affecting the SVR model. This finding aligns with literature acknowledging the potential of sentiment analysis to enhance predictions but also highlights its volatility and context-dependent nature (8). It suggests that sentiment analysis can be a double-edged sword, improving forecasts in some scenarios while adding unpredictability in others.

The results of portfolio optimization emphasize the practical utility of combining price predictions with frequent rebalancing strategies. Price prediction-based strategies significantly outperformed sentiment-based ones, particularly with daily and weekly rebalancing. This finding underscores the importance of responsive and adaptive portfolio management techniques in real-world trading (82).

The results from the stock price prediction models are in line with the EMH and RW theory by demonstrating that it is not possible to achieve better performance than the benchmark. However, the results of portfolio optimization demonstrated that it is possible to achieve substantial returns through systematic prediction and rebalancing strategies (9). They also caution against over-reliance on sentiment data, which can lead to poor performance if not carefully managed.

In practical terms, the study suggests that integrating diverse data sources and employing frequent rebalancing can enhance the effectiveness of trading strategies. While sentiment analysis offers additional insights, its application should be context-specific and carefully evaluated to avoid introducing excessive noise. Overall, the research highlights the need for a balanced approach that leverages both traditional financial indicators and modern data analytics to achieve optimal stock price predictions and trading performance.

7

Discussion

7.1 Practical Implications

The findings of this research have several potential impacts on financial market practices and trading strategies. Based on the results of the study, several key recommendations for practitioners can be highlighted to improve current financial models and trading frameworks.

First, the choice of data is crucial. Incorporating a wide range of data sources does not inherently improve predictive performance. Practitioners should carefully assess the profitability and relevance of the data being used. The most beneficial data tend to be new or unique, providing an edge over competitors. Therefore, detailed methods for selecting and validating data sources should be established to ensure that only the most valuable information is utilized.

Second, sentiment data should be treated with caution. Although it can improve model performance in specific scenarios, it also has the potential to introduce significant noise. Practitioners should employ robust filtering and validation techniques to ensure that sentiment data contributes positively to prediction accuracy. This involves distinguishing between meaningful sentiment signals and irrelevant noise from tweets or other forms of text data.

Third, a portfolio optimization strategy should always be included. High prediction scores are beneficial, but ultimately meaningless, if they do not translate into increased portfolio value. Practitioners should focus on integrating predictive models with robust portfolio management strategies to maximize returns. The study showed that frequent rebalancing, particularly on a daily or weekly basis, significantly outperformed less frequent rebalancing approaches.

7.2 Limitations & Future Research

This study encountered several limitations and constraints that may have influenced the results. It is important to acknowledge these factors to provide a comprehensive understanding of the research results and their potential impact.

First, the feature selection methods used in this study were regular and straightforward. While they provided a baseline for identifying important features, more advanced feature selection techniques might have yielded a different and potentially more effective set of features. This limitation aligns with the practical implications discussed earlier, emphasizing the need for detailed methods for selecting data sources.

Second, the machine learning methods, particularly deep learning models, require large amounts of data to train effectively. The performance of these models may have been adversely affected by the data constraints, potentially leading to less accurate predictions. Ensuring access to extensive and high-quality datasets is crucial for the successful application of these advanced methods.

Third, the sentiment data used in this study could have been more recent. The timeliness of sentiment data is critical for capturing current market sentiments accurately. Utilizing more up-to-date sentiment data could have improved the performance of the sentiment analysis models.

Fourthly, the portfolio optimization strategy employed in this study was relatively basic, relying on a simple rule-based model. While this provided valuable insights, there is a need to enhance the portfolio optimization approach by incorporating different performance metrics such as the Sharpe ratio, alpha, beta, and others. These metrics can provide a more comprehensive evaluation of the portfolio's performance and help refine the optimization strategy.

References

[1] E. F. Fama. **Efficient capital markets: II.** *The Journal of Finance, 46(5), 1575-1617*, -1991. 1, 6, 78

[2] Fisher L. Jensen M. C. Roll R. Fama, E. F. **The adjustment of stock prices to new information.** *International Economic Review, 10(1), 1-21*, -1969. 1, 6

[3] Jorge García Francisco Guijarro Arévalo, Rubén and Alfred Peris. **A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting.** *Expert Systems with Applications*, **81**:177–92, 2017. 1

[4] R. J. Shiller. **Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?** *The American Economic Review, 71(3), 421–436.*, 1981. 2

[5] Frank M. Z. Antweiler, W. **Is all that talk just noise? The information content of internet stock message boards**. *Journal of Finance, 59(3), 1259-1294*, -2004. 2

[6] Whitelaw R. F. Tumarkin, R. **News or noise? Internet postings and stock prices**. *Financial Analyst Journal, 57(3), 41-51*, -2001. 2

[7] Chen H. Schumaker, R.P. **Textual Analysis of stock market prediction using breaking financial news: the afzin text system**. *ACM Transactions on information Systems, 27(2), 12:1-12:19*, -2009. 2, 14

[8] Mao H. Zeng X. Bollen, J. **Twitter mood predicts stock market**. *Journal of Computational Science, 2(1), 1-8*, -2011. 2, 11, 79

[9] Burton G. Malkiel. *A random walk down wall street.* -1973. 6, 79

[10] J. C. Van Horne and G. G. Parker. **The random-walk theory: an empirical test**. *Financial Analysts Journal, 23(6), 87-92*, 1967. 6

[11] P. H. Cootner. **The random character of stock market prices**. *Cambridge, MA: MIT Press*, -1964. 6

[12] E. F. Fama. **The bahvior of stock market prices.** *The journal of business, 38(1), 34-105*, 1965. 6

[13] Kenneth R. French. **Stock returns and the weekend effect**. *Journal of Financial Economics, 8 (1), 55-69*, 1980. 7, 30

[14] D.B. Keim. **Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence.** *Journal of Financial Economics, 12, 13-32.*, 1983. 7, 30

[15] R. Werner F. M. De Bondt, Thaler. **Does the Stock Market Overreact?** *The Journal of Finance, 40(3), 793–805.*, 1985. 7

[16] Zhu N. Dhar, R. **Up Close and Personal: Investor Sophistication and the Disposition Effect**. *Management Science, 52(5), 726–740.*, 2006. 7

[17] Mark Han, Bing NMI1 Grinblatt. **The Disposition Effect and Momentum**. *Yale School of Management*, 2001. 7

[18] Statman M. Shefrin, H. **The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence.** *The Journal of Finance, 40(3), 777–790.*, 1985. 7

[19] Jiang Wang Andrew W. Lo, Harry Mamaysky. **Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation**. *The Journal of Finance, 55 (4), 1705-1765*, 2000. 8

[20] Alexander Arkadyevich Safronov and Alexey Ivanovich Sazonov. **Assessing the Investment Attractiveness of Shares: The Joint Use of Fundamental and Technical Analysis**. *Universal Journal of Accounting and Finance, 9(5), 908-915*, 2021. 9

[21] Kaouther Flifel. **Financial Markets between Efficiency and Persistence: Empirical Evidence on Daily Data**. *Asian Journal of Finance Accounting, 4(2)*, 2012. 9

[22] S. Jakpar, M. Tinggi, A. H. Tak, and W. Y. Chong. **Fundamental analysis vs technical analysis: The comparison of two analysis in malaysia stock market**. *UNIMAS Review of Accounting and Finance, 2(1)*, 2018. 9

[23] DEV SHAH, HARUNA ISAH, AND FARHANA ZULKERNINE. **Stock market analysis: A review and taxonomy of prediction techniques**. *International Journal of Financial Studies*, **7**(2):26, 2019. 10

[24] JENKINS G. M. REINSEL G. C. LJUNG G. M. BOX, G. E. **Time series analysis: forecasting and control**. *John Wiley Sons*, -2015. 10

[25] T. BOLLERSLEV. **A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return**. *The Review of Economics and Statistics, 69(3), 542-547*, 1987. 10

[26] P. H. FRANSES AND D. VAN DIJK. **Forecasting stock market volatility using (non-linear) Garch models**. *Journal of forecasting, 15(3), 229-235*, 1996. 10

[27] KITA E. ZUO, Y. **Stock price forecast using bayesian network.** *Expert Systems with Applications: An International Journal, 39(8), 6729-6737*, 2012. 10

[28] PRIYANK THAKKAR K KOTECHA JIGAR PATEL, SAHIL SHAH. **Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques**. *Expert Systems with Applications, 42(1), 259-268*, -2015. 10, 12

[29] R. S. TSAY. *Analysis of financial time series.* John Wiley Sons, 2005. 10

[30] PRIYANK THAKKAR K KOTECHA JIGAR PATEL, SAHIL SHAH. **Predicting stock market index movement using fusion of machine learning techniques**. *Expert Systems with Applications, 42(4), 2162-2172*, -2015. 10, 12

[31] CHANG S. HA Q. T. COLLIER N VU, T. T. **An experiment in integrating sentiment features for tech stock prediction in Twitter**. *24th international conference on computational linguistics (pp. 23-38)*, -2012. 11

[32] S VANAJA AND RAMESHKUMAR KRISHNASWAMY. **Performance Analysis of Classification Algorithms on Medical Diagnoses-a Survey**. *Journal of Computer Science*, **113050**:30–52, 09 2014. 11

[33] C. M. BISHOP. *Pattern recognition and machine learning.* Springer, 2006. 11

[34] G. E. HINTON, S. OSINDERO, AND Y. W. TEH. **A fast learning algorithm for deep belief nets**. *Neural computation, 18(7), 1527-1554*, 2006. 11

[35] J. J. Murphy. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications.* Penguin, 1999. 11, 22

[36] K. H. Kim. **Dollar exchange rate and stock price: evidence from multivariate cointegration and error correction model**. *Review of Financial economics, 12(3), 301-313*, 2003. 12

[37] W. Huang, Y. Nakamori, and S. Y. Wang. **Forecasting stock market movement direction with support vector machine**. *Computers operations research, 32(10), 2513-2522*, 2005. 12

[38] M. R. Hassan, B. Nath, and M. Kirley. **A fusion model of HMM, ANN and GA for stock market forecasting**. *Expert Systems with Applications, 33(1), 171-180*, 2007. 12

[39] F C. Park E. Chong, C. Han. **Deep leamingnetwotks for stock market analysis and prediction: methodology, data representations, and case studies**. *Expert Systems with Apptications, 83, 187-205*, 2017. 12

[40] K. P. Valavanis. G. S. Atsalakis. **Forecasting stock matket short-tenn trends using a neuro-fu:z:zy based methodology,**. *Expert Systems with Applications, 36 (7), 10696-10707*, 2009. 12

[41] D. A. Georgoutsos. S. D. Bekiros. **Evaluating direction-of-change forecasting: Neurofuzzy models vs. neural networks**. *Mathematical and Computer Modelling, 46 (1), pp. 3846,*, 2007. 12

[42] S. J. Lee C. C. Wei, T. T. Chen. **A k-NN Based Neuro-Fuzzy System for Time Series Prediction**. *14th ACIS International Conference on Software Engineering, Artificial Intelligence, Netwotking and Parallel/Distributed Computing, 569-574*, 2013. 12

[43] Rao Y Bao W, Yue J. **A deep learning framework for financial time series using stacked autoencoders and long-short term memory**. *PLoS ONE*, 2017. 12, 43

[44] Dirk Hespeels Nathalie Gryp Ruben. Ballings, Michel Van den Poel. **Evaluating multiple classifiers for stock price direction prediction.** *Expert Systems with Applications. 42.*, 2015. 12

[45] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT Press, 2016. 12

[46] Y. LeCun, Y. Bengio, and G. Hinton. **Deep learning**. *Nature, 521(7553), 436-444*, 2015. 12

[47] A. M. Rather, A. Agarwal, and V. N. Sastry. **Recurrent neural network and a hybrid model for prediction of stock returns**. *Expert Systems with Applications, 42(6), 3234-3241*, 2015. 12

[48] K. Chen, Y. Zhou, and F. Dai. **A LSTM-based method for stock returns prediction: A case study of China stock market**. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2823–2824. IEEE, 2015. 12

[49] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. **Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction**. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, 12 2017. 12

[50] Ha Young Kim Yujin Baek. **ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module**. -2018. 12

[51] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. **Temporal Relational Ranking for Stock Prediction**. *ACM Transactions on Information Systems*, **37**:1–30, 03 2019. 13

[52] Bing Liu. **Sentiment Analysis and Opinion Mining**. In *Synthesis Lectures on Human Language Technologies*, **5**, 05 2012. 13

[53] Bo Pang and Lillian Lee. **Opinion Mining and Sentiment Analysis**. *Foundations and Trends in Information Retrieval*, **2**:1–135, 01 2008. 13

[54] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li. **Sentiment Classification of Internet Restaurant Reviews Written in Cantonese**. *Expert Systems with Applications*, **38**(6):7674–7682, 2011. 13

[55] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang. **Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach**. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 1031–1040, 10 2011. 14

[56] M. Kraus and S. Feuerriegel. **Decision Support from Financial Disclosures with Deep Neural Networks and Transfer Learning**. *Decision Support Systems*, **104**:38–48, 2017. 14, 15

[57] Tim Loughran and Bill McDonald. **Textual Analysis in Accounting and Finance: A Survey**. *Journal of Accounting Research*, **54**:1187–1230, 2016. 14

[58] Unknown Pagolu. **Deep Learning Semantics Using N-Grams**. 2016. 14

[59] Kiyoaki Shirai Julien Velcin Nguyen, T. H. **Sentiment analysis on social media for stock movement prediction**. *Expert Systems With Applications, 42, 9603-9611*, 2015. 14, 21

[60] Nan Jing, Zhao Wu, and Hefei Wang. **A Hybrid Model Integrating Deep Learning with Investor Sentiment Analysis for Stock Price Prediction**. *Expert Systems with Applications*, **178**:115019, 04 2021. 14

[61] Jiahong Li, Hui Bu, and Junjie Wu. **Sentiment-aware stock market prediction: A deep learning method**. In *2017 international conference on service systems and service management*, pages 1–6. IEEE, 2017. 15

[62] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara. **Deep Learning for Stock Prediction Using Numerical and Textual Information**. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pages 1–6. IEEE, 06 2016. 15

[63] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar. **Big Data: Deep Learning for Financial Sentiment Analysis**. *Journal of Big Data*, **5**(1):1–25, 2018. 15

[64] Bernhard Lutz, Nicolas Pröllochs, and Dirk Neumann. **Sentence-Level Sentiment Analysis of Financial News Using Distributed Text Representations and Multi-Instance Learning**. Technical Report arXiv:1901.00400, arXiv, 2018. 15

[65] Andre Freitas Macedo Maia and Siegfried Handschuh. **FinSSLx: A Sentiment Analysis Model for the Financial Domain Using Text Simplification**. In *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, pages 318–319. IEEE, 2018. 15

[66] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. 15, 16, 49

[67] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. **RoBERTa: A Robustly Optimized BERT Pretraining Approach**. *arXiv preprint arXiv:1907.11692*, 2019. 16

[68] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. **DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter**. *arXiv preprint arXiv:1910.01108*, 2019. 16

[69] Yi Yang, Mark Christopher Siy Uy, and Allen Huang. **FinBERT: A Pretrained Language Model for Financial Communications**, 2020. 16

[70] Ahmed Hazourli. **FinancialBERT - A Pretrained Language Model for Financial Text Mining**, 02 2022. 17

[71] John Bollinger. *Bollinger on Bollinger Bands*. McGraw-Hill, New York, 2002. 22

[72] Joseph E. Granville. *New Key to Stock Market Profits*. Unknown Publisher. 22

[73] George C. M.D. Lane. **Lane's Stochastics**. *Stocks Commodities*, **2**(3):87–90, 1984. 22

[74] Jr. J. Welles Wilder. *New Concepts in Technical Trading Systems*. Trend Research, Greensboro, NC, 1978. Includes the Average Directional Index (ADX) indicator. 22

[75] Helmut Luetkepohl and Fang Xu. **The role of the log transformation in forecasting economic variables**. *Empirical Economics*, **42**, 03 2009. 36

[76] Ishwaran H. . Chen X. **Random forests for genomic data analysis**. *Genomics*, **99(6)**:323–9, 2012. 45

[77] L. Breiman. **Random forests.** *Machine learning*, **45**:5–32, 2001. 45, 78

[78] Pallavi Deshpande. **Heart Disease Prediction using Random Forest**. 04 2023. 46

[79] Harris Drucker, Christopher Burges, Linda Kaufman, Alexander Smola, and V. Vapnik. **Support vector regression machines**. *Adv Neural Inform Process Syst*, **28**:779–784, 01 1997. 46

[80] Sepp Hochreiter and Jürgen Schmidhuber. **Long short-term memory**. *Neural computation*, **9**(8):1735–1780, 1997. 47, 79

[81] Zhengmin Kong, Yd Cui, Zhou Xia, and He Lv. **Convolution and Long Short-Term Memory Hybrid Deep Neural Networks for Remaining Useful Life Prognostics**. *Applied Sciences*, **9**:4156, 10 2019. 48

[82] Harry Markowitz. **Portfolio Selection**. *The Journal of Finance*, **7**(1):77–91, 1952. 51, 79

[83] Vladimir N Vapnik. *Statistical learning theory.* Wiley, 1998. 78

Appendix

A Data

Table 1: Top 25 Components of the SPDR S&P 500 Trust ETF (SPY) by Weight

#	Company	Ticker	Weight
1	Apple Inc.	AAPL	7.05%
2	Microsoft Corp	MSFT	6.54%
3	Amazon.com Inc	AMZN	3.24%
4	Nvidia Corp	NVDA	2.79%
5	Alphabet Inc. Class A	GOOGL	2.13%
6	Tesla Inc.	TSLA	1.95%
7	Alphabet Inc. Class C	GOOG	1.83%
8	Berkshire Hathaway Class B	BRK.B	1.83%
9	Meta Platforms, Inc. Class A	META	1.81%
10	UnitedHealth Group	UNH	1.28%
11	Exxon Mobil	XOM	1.27%
12	Eli Lilly & Co.	LLY	1.21%
13	JPMorgan Chase	JPM	1.18%
14	Johnson & Johnson	JNJ	1.07%
15	Visa Class A	V	1.05%
16	Procter & Gamble	PG	0.99%
17	Mastercard Class A	MA	0.93%
18	Broadcom Inc.	AVGO	0.92%
19	Home Depot	HD	0.85%
20	Chevron Corporation	CVX	0.81%
21	Merck	MRK	0.75%
22	AbbVie	ABBV	0.75%
23	Costco	COST	0.67%
24	PepsiCo	PEP	0.67%
25	Adobe	ADBE	0.65%
		Combined Weight:	44.22%

Table 2: Computation of the Average Directional Index (ADX)

Step	Formula/Description
1. Calculate True Range (TR)	$\text{TR} = \max[(H-L), \lvert H-C_{\text{prev}}\rvert, \lvert L-C_{\text{prev}}\rvert]$
	H: Current high, L: Current low, C_{prev} : $Previousclose$
2. Calculate Directional Movement (DM)	$\text{DM}_{+} = H - H_{\text{prev}} \quad \text{if } H - H_{\text{prev}} > L_{\text{prev}} - L, \text{ otherwise } 0$
	$\text{DM}_{-} = L_{\text{prev}} - L \quad \text{if } L_{\text{prev}} - L > H - H_{\text{prev}}, \text{ otherwise } 0$
3. Calculate Smoothed True Range (ATR)	$\text{ATR} = \left(\sum_{i=1}^{n} TR_i\right)/n \quad$ or a smoothed moving average
4. Calculate Smoothed Directional Indicators (DI)	$\text{DI}_{+} = 100 \times \left(\frac{\text{Smoothed DM}_{+}}{ATR}\right) \quad$ for n periods
	$\text{DI}_{-} = 100 \times \left(\frac{\text{Smoothed DM}_{-}}{ATR}\right) \quad$ for n periods
5. Calculate Directional Index (DX)	$\text{DX} = \left(\frac{\lvert DI_{+}-DI_{-}\rvert}{DI_{+}+DI_{-}}\right) \times 100$
6. Calculate Average Directional Index (ADX)	ADX = Smoothed Moving Average of DX over n periods

B Model Results

B.1 Sentiment Model

Class	Precision	Recall	F1-Score	Support
Negative (0)	0.08	0.01	0.01	188
Neutral (1)	0.61	0.99	0.76	879
Positive (2)	0.00	0.00	0.00	358
Accuracy			0.61	1425
Macro Avg	0.23	0.33	0.26	1425
Weighted Avg	0.39	0.61	0.47	1425

Table 3: Evaluation Metrics for BERT Model

	Pred. Negative (0)	Pred. Neutral (1)	Pred. Positive (2)
Actual Negative (0)	1	187	0
Actual Neutral (1)	12	867	0
Actual Positive (2)	0	358	0

Table 4: Confusion Matrix for BERT Model

Class	Precision	Recall	F1-Score	Support
Negative (0)	0.88	0.23	0.37	188
Neutral (1)	0.68	0.97	0.80	879
Positive (2)	0.82	0.30	0.44	358
Accuracy			0.70	1425
Macro Avg	0.79	0.50	0.54	1425
Weighted Avg	0.74	0.70	0.65	1425

Table 5: Evaluation Metrics for RoBERTa Model

	Pred. Negative (0)	Pred. Neutral (1)	Pred. Positive (2)
Actual Negative (0)	44	144	0
Actual Neutral (1)	5	851	23
Actual Positive (2)	1	251	106

Table 6: Confusion Matrix for RoBERTa Model

Class	Precision	Recall	F1-Score	Support
Negative (0)	0.13	1.00	0.23	188
Neutral (1)	0.00	0.00	0.00	879
Positive (2)	0.14	0.00	0.01	358
Accuracy			0.13	1425
Macro Avg	0.09	0.33	0.08	1425
Weighted Avg	0.05	0.13	0.03	1425

Table 7: Evaluation Metrics for DistilBERT Model

B.2 Phase 1

	Pred. Negative (0)	Pred. Neutral (1)	Pred. Positive (2)
Actual Negative (0)	188	0	0
Actual Neutral (1)	873	0	6
Actual Positive (2)	357	0	1

Table 8: Confusion Matrix for DistilBERT Model

Class	Precision	Recall	F1-Score	Support
Negative (0)	0.96	0.99	0.97	360
Neutral (1)	0.92	0.99	0.96	197
Positive (2)	1.00	0.97	0.98	868
Accuracy			0.98	1425
Macro Avg	0.96	0.98	0.97	1425
Weighted Avg	0.98	0.98	0.98	1425

Table 9: Evaluation Metrics for FinBERT Model

	Pred. Positive (0)	Pred. Negative (1)	Pred. Neutral (2)
Actual Positive (0)	355	5	0
Actual Negative (1)	1	196	0
Actual Neutral (2)	14	12	842

Table 10: Confusion Matrix for FinBERT Model

Class	Precision	Recall	F1-Score	Support
Negative (0)	0.99	0.97	0.98	194
Neutral (1)	1.00	0.99	0.99	890
Positive (2)	0.96	0.99	0.98	341
Accuracy			0.99	1425
Macro Avg	0.98	0.99	0.98	1425
Weighted Avg	0.99	0.99	0.99	1425

Table 11: Evaluation Metrics for FinancialBERT Model

B.3 Phase 2

	Predicted Negative (0)	Predicted Neutral (1)	Predicted Positive (2)
Actual Negative (0)	189	1	4
Actual Neutral (1)	0	881	9
Actual Positive (2)	2	1	338

Table 12: Confusion Matrix for FinancialBERT Model

B.4 Phase 3

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	2.9916	1.7296	1.0222	0.0124	0.9991
MSFT	10.0407	3.1687	1.8767	0.0117	0.9991
AMZN	4.5924	2.143	1.296	0.0143	0.9983
NVDA	29.8367	5.4623	2.4992	0.0201	0.9983
GOOGL	2.2133	1.4877	0.9118	0.0122	0.9984
TSLA	29.927	5.4706	2.6519	0.0242	0.9975
GOOG	2.2193	1.4897	0.9072	0.0121	0.9985
BRK-B	7.1109	2.6666	1.7957	0.0083	0.9986
META	24.615	4.9614	2.8553	0.0154	0.9966
UNH	21.898	4.6795	2.8714	0.0107	0.999
XOM	1.1565	1.0754	0.747	0.012	0.9969
LLY	16.3125	4.0389	2.0202	0.0112	0.9992
JPM	2.637	1.6239	1.0559	0.0113	0.998
JNJ	1.9686	1.4031	0.9297	0.0077	0.9979
V	6.3129	2.5125	1.5704	0.0105	0.9986
PG	1.4911	1.2211	0.7766	0.0077	0.9986
MA	20.2441	4.4993	2.7321	0.0116	0.9986
AVGO	69.0584	8.3101	4.7928	0.0154	0.9988
HD	11.185	3.3444	2.0681	0.0102	0.9986
CVX	3.0339	1.7418	1.1606	0.0122	0.9968
MRK	0.8054	0.8974	0.5966	0.0093	0.9983
ABBV	1.7758	1.3326	0.8715	0.0115	0.9988
COST	21.2707	4.612	2.6845	0.0091	0.9992
PEP	1.9306	1.3895	0.8789	0.0076	0.9985
ADBE	53.1882	7.293	4.2555	0.0139	0.9983

Table 13: Phase 1 Naive Model

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	3.7012	1.9239	1.2077	0.0127	0.9989
MSFT	12.4456	3.5278	2.2388	0.012	0.9988
AMZN	5.6804	2.3834	1.5431	0.0143	0.9971
NVDA	37.2778	6.1056	3.1044	0.0218	0.998
GOOGL	2.7022	1.6438	1.053	0.0124	0.9978
TSLA	37.3544	6.1118	3.2419	0.0254	0.9971
GOOG	2.7108	1.6465	1.0495	0.0124	0.9979
BRK-B	8.4223	2.9021	1.9858	0.0085	0.998
META	30.0733	5.4839	3.2778	0.0156	0.9947
UNH	26.8982	5.1864	3.3347	0.0107	0.9986
XOM	1.3056	1.1426	0.7959	0.0126	0.9971
LLY	20.1507	4.4889	2.3636	0.0114	0.9991
JPM	3.1856	1.7848	1.1989	0.0115	0.9968
JNJ	2.303	1.5176	1.013	0.0077	0.9961
V	7.6925	2.7735	1.8057	0.0106	0.9976
PG	1.7843	1.3358	0.8669	0.0079	0.9982
MA	24.9558	4.9956	3.1998	0.0118	0.9977
AVGO	85.2241	9.2317	5.6356	0.015	0.9985
HD	13.6667	3.6969	2.383	0.0105	0.9978
CVX	3.5136	1.8745	1.2521	0.0123	0.9964
MRK	0.9387	0.9689	0.649	0.0093	0.9979
ABBV	2.0753	1.4406	0.9529	0.0109	0.9985
COST	26.1986	5.1185	3.1377	0.0095	0.9989
PEP	2.3052	1.5183	0.9739	0.0078	0.9978
ADBE	66.0628	8.1279	5.092	0.0144	0.9975

Table 14: Phase 1 Naive Model (Window Size: 500)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	3.311	1.8196	1.1107	0.0127	0.999
MSFT	11.121	3.3348	2.0449	0.0119	0.999
AMZN	5.0869	2.2554	1.4143	0.0144	0.9978
NVDA	33.1307	5.7559	2.7694	0.0211	0.9982
GOOGL	2.4424	1.5628	0.9806	0.0124	0.9982
TSLA	33.2141	5.7632	2.9128	0.0247	0.9973
GOOG	2.4496	1.5651	0.9764	0.0123	0.9982
BRK-B	7.7383	2.7818	1.8942	0.0085	0.9984
META	27.1338	5.209	3.0584	0.0155	0.9958
UNH	24.2062	4.92	3.1092	0.0109	0.9988
XOM	1.233	1.1104	0.7738	0.0124	0.997
LLY	18.08	4.2521	2.1979	0.0116	0.9992
JPM	2.8952	1.7015	1.1277	0.0115	0.9975
JNJ	2.1246	1.4576	0.9706	0.0077	0.9973
V	6.9572	2.6377	1.6909	0.0107	0.9982
PG	1.6294	1.2765	0.822	0.0079	0.9985
MA	22.3553	4.7281	2.9503	0.0117	0.9982
AVGO	76.5065	8.7468	5.2248	0.0155	0.9987
HD	12.3348	3.5121	2.2273	0.0104	0.9982
CVX	3.2697	1.8082	1.2117	0.0125	0.9968
MRK	0.868	0.9317	0.6229	0.0094	0.9982
ABBV	1.9238	1.387	0.9141	0.0113	0.9987
COST	23.5173	4.8495	2.9079	0.0094	0.999
PEP	2.1062	1.4513	0.9262	0.0077	0.9982
ADBE	58.9322	7.6767	4.6393	0.0141	0.998

Table 15: Phase 1 Naive Model (Window Size: 250)

Stock	**MSE**	**RMSE**	**MAE**	**MAPE**	**R^2 Score**
AAPL	3.1442	1.7732	1.0674	0.0126	0.9991
MSFT	10.5544	3.2487	1.9594	0.0119	0.999
AMZN	4.8244	2.1964	1.3512	0.0143	0.9981
NVDA	31.3844	5.6022	2.6265	0.0206	0.9983
GOOGL	2.3202	1.5232	0.9438	0.0123	0.9983
TSLA	31.4719	5.61	2.7744	0.0244	0.9974
GOOG	2.3267	1.5254	0.9394	0.0122	0.9984
BRK-B	7.4253	2.7249	1.8488	0.0084	0.9985
META	25.7842	5.0778	2.9471	0.0153	0.9962
UNH	23.0077	4.7966	2.9917	0.0109	0.9989
XOM	1.2006	1.0957	0.7648	0.0123	0.997
LLY	17.1468	4.1409	2.107	0.0114	0.9992
JPM	2.7626	1.6621	1.0927	0.0114	0.9978
JNJ	2.0512	1.4322	0.9544	0.0078	0.9976
V	6.6225	2.5734	1.6299	0.0106	0.9984
PG	1.5596	1.2489	0.801	0.0078	0.9985
MA	21.2452	4.6093	2.8361	0.0117	0.9984
AVGO	72.6076	8.521	5.0098	0.0156	0.9988
HD	11.7441	3.427	2.153	0.0104	0.9984
CVX	3.1697	1.7804	1.1947	0.0125	0.9968
MRK	0.8374	0.9151	0.6106	0.0094	0.9983
ABBV	1.8538	1.3615	0.8966	0.0115	0.9987
COST	22.3508	4.7277	2.7974	0.0093	0.9991
PEP	2.0174	1.4204	0.9033	0.0076	0.9983
ADBE	55.8954	7.4763	4.4381	0.014	0.9982

Table 16: Phase 1 Naive Model (Window Size: 125)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	2.3666	1.5384	0.9202	0.0126	0.9988
MSFT	7.3800	2.7166	1.6550	0.0112	0.9989
AMZN	4.3140	2.0770	1.3004	0.0129	0.9981
NVDA	11.6696	3.4161	1.7819	0.0206	0.9976
GOOGL	1.6034	1.2663	0.8088	0.0113	0.9983
TSLA	25.9668	5.0958	2.2826	0.0254	0.9975
GOOG	1.5960	1.2633	0.8007	0.0113	0.9984
BRK-B	7.2063	2.6845	1.7772	0.0086	0.9961
META	23.8161	4.8802	2.9024	0.0139	0.9951
UNH	19.6525	4.4331	2.7796	0.0110	0.9977
XOM	0.7052	0.8398	0.6139	0.0123	0.9924
LLY	6.3921	2.5283	1.4412	0.0112	0.9980
JPM	3.0624	1.7500	1.1320	0.0118	0.9964
JNJ	2.1584	1.4692	0.9576	0.0078	0.9946
V	7.2161	2.6863	1.6799	0.0108	0.9973
PG	1.4904	1.2208	0.7704	0.0079	0.9978
MA	23.5321	4.8510	2.9754	0.0121	0.9974
AVGO	34.6022	5.8824	3.8772	0.0148	0.9975
HD	10.7221	3.2745	2.0045	0.0100	0.9979
CVX	2.2744	1.5081	1.0166	0.0122	0.9848
MRK	0.7740	0.8798	0.5772	0.0095	0.9940
ABBV	1.6024	1.2659	0.8241	0.0116	0.9964
COST	14.2560	3.7757	2.3554	0.0092	0.9987
PEP	1.9179	1.3849	0.8692	0.0078	0.9960
ADBE	53.0234	7.2817	4.3558	0.0134	0.9980

Table 17: Phase 1 LR (Window Size: 500)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	2.1139	1.4539	0.8467	0.0128	0.9989
MSFT	6.4841	2.5464	1.5017	0.0114	0.9990
AMZN	3.8140	1.9529	1.1813	0.0133	0.9985
NVDA	10.2223	3.1972	1.5567	0.0202	0.9979
GOOGL	1.4107	1.1877	0.7512	0.0115	0.9986
TSLA	22.5805	4.7519	2.0194	0.0249	0.9976
GOOG	1.4053	1.1855	0.7423	0.0115	0.9986
BRK-B	6.3601	2.5219	1.6885	0.0086	0.9971
META	21.1555	4.5995	2.7105	0.0142	0.9963
UNH	17.6734	4.2040	2.6005	0.0114	0.9982
XOM	0.7067	0.8407	0.6115	0.0121	0.9912
LLY	5.8425	2.4171	1.3807	0.0116	0.9981
JPM	2.7498	1.6583	1.0572	0.0118	0.9972
JNJ	1.9966	1.4130	0.9253	0.0080	0.9965
V	6.3703	2.5239	1.5566	0.0110	0.9979
PG	1.4066	1.1860	0.7564	0.0081	0.9981
MA	20.5138	4.5292	2.6890	0.0120	0.9980
AVGO	30.5953	5.5313	3.6090	0.0156	0.9980
HD	9.7496	3.1224	1.8872	0.0102	0.9983
CVX	2.1651	1.4714	1.0086	0.0127	0.9891
MRK	0.7285	0.8535	0.5602	0.0098	0.9953
ABBV	1.4928	1.2218	0.8043	0.0122	0.9969
COST	13.3062	3.6478	2.2271	0.0093	0.9988
PEP	1.7136	1.3090	0.8191	0.0077	0.9969
ADBE	47.3541	6.8814	3.9512	0.0136	0.9984

Table 18: Phase 1 LR (Window Size: 250)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	2.1053	1.4510	0.8398	0.0133	0.9988
MSFT	6.2745	2.5049	1.4771	0.0119	0.9991
AMZN	3.7099	1.9261	1.1448	0.0137	0.9986
NVDA	9.9566	3.1554	1.4959	0.0204	0.9979
GOOGL	1.3600	1.1662	0.7351	0.0117	0.9986
TSLA	22.1748	4.7090	1.9462	0.0249	0.9975
GOOG	1.3468	1.1605	0.7215	0.0116	0.9987
BRK-B	6.3189	2.5137	1.6829	0.0087	0.9971
META	20.7009	4.5498	2.6919	0.0146	0.9966
UNH	17.3946	4.1707	2.5232	0.0115	0.9983
XOM	0.7530	0.8677	0.6334	0.0124	0.9904
LLY	5.8103	2.4105	1.3652	0.0118	0.9981
JPM	2.6675	1.6333	1.0264	0.0119	0.9974
JNJ	1.9842	1.4086	0.9314	0.0082	0.9967
V	6.2873	2.5074	1.5270	0.0112	0.9981
PG	1.2955	1.1382	0.7399	0.0081	0.9982
MA	20.2975	4.5053	2.6444	0.0123	0.9981
AVGO	29.3231	5.4151	3.4987	0.0160	0.9982
HD	9.5783	3.0949	1.8902	0.0106	0.9984
CVX	2.2027	1.4841	1.0324	0.0130	0.9884
MRK	0.7262	0.8521	0.5678	0.0101	0.9954
ABBV	1.5304	1.2371	0.8120	0.0127	0.9969
COST	13.7362	3.7062	2.1983	0.0095	0.9988
PEP	1.6742	1.2939	0.8173	0.0079	0.9971
ADBE	46.0243	6.7841	3.8072	0.0137	0.9984

Table 19: Phase 1 LR (Window Size: 125)

Stock	**MSE**	**RMSE**	**MAE**	**MAPE**	**R^2 Score**
AAPL	5.1859	2.2772	1.3761	0.0187	0.9973
MSFT	14.6911	3.8329	2.4468	0.0166	0.9978
AMZN	8.1592	2.8564	1.8262	0.0184	0.9964
NVDA	29.3470	5.4173	2.7180	0.0316	0.9941
GOOGL	3.1351	1.7706	1.1426	0.0157	0.9967
TSLA	63.0171	7.9383	3.3680	0.0358	0.9938
GOOG	2.9694	1.7232	1.1067	0.0154	0.9970
BRK-B	12.4940	3.5347	2.3512	0.0112	0.9932
META	39.9808	6.3230	3.9521	0.0188	0.9918
UNH	36.9495	6.0786	4.0277	0.0159	0.9957
XOM	1.8311	1.3532	0.8538	0.0174	0.9803
LLY	14.3933	3.7939	2.1145	0.0161	0.9955
JPM	4.9946	2.2349	1.4919	0.0155	0.9941
JNJ	3.3977	1.8433	1.3024	0.0107	0.9915
V	11.0593	3.3255	2.2163	0.0145	0.9959
PG	2.6503	1.6280	1.1042	0.0113	0.9961
MA	36.3247	6.0270	3.9725	0.0168	0.9960
AVGO	73.0081	8.5445	5.5937	0.0207	0.9948
HD	23.7946	4.8780	3.0064	0.0147	0.9954
CVX	4.6370	2.1534	1.3548	0.0164	0.9690
MRK	1.2322	1.1101	0.7665	0.0126	0.9905
ABBV	3.6631	1.9139	1.2314	0.0167	0.9917
COST	31.0073	5.5684	3.6349	0.0139	0.9972
PEP	3.0465	1.7454	1.1865	0.0106	0.9936
ADBE	95.2746	9.7609	6.1465	0.0195	0.9965

Table 20: Phase 1 RF (Window Size: 500)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	4.5339	2.1293	1.2436	0.0184	0.9975
MSFT	12.6763	3.5604	2.1844	0.0165	0.9981
AMZN	7.3298	2.7074	1.6708	0.0194	0.9971
NVDA	24.9617	4.9962	2.3978	0.0321	0.9949
GOOGL	2.7993	1.6731	1.0610	0.0160	0.9971
TSLA	53.9824	7.3473	2.9653	0.0349	0.9942
GOOG	2.6544	1.6292	1.0341	0.0158	0.9974
BRK-B	12.0775	3.4753	2.2581	0.0114	0.9944
META	40.1074	6.3330	3.8232	0.0199	0.9929
UNH	32.8850	5.7345	3.6888	0.0158	0.9967
XOM	1.8131	1.3465	0.8749	0.0175	0.9775
LLY	12.9445	3.5978	2.0212	0.0166	0.9958
JPM	4.7673	2.1834	1.4355	0.0160	0.9951
JNJ	3.4051	1.8453	1.2571	0.0109	0.9940
V	9.6087	3.0998	2.0344	0.0145	0.9969
PG	2.4426	1.5629	1.0623	0.0116	0.9966
MA	31.1794	5.5839	3.5557	0.0163	0.9969
AVGO	71.8678	8.4775	5.3003	0.0223	0.9953
HD	21.0881	4.5922	2.7908	0.0147	0.9962
CVX	4.9203	2.2182	1.4036	0.0176	0.9752
MRK	1.2018	1.0963	0.7553	0.0132	0.9923
ABBV	3.4864	1.8672	1.2132	0.0178	0.9928
COST	26.6153	5.1590	3.2990	0.0135	0.9977
PEP	2.9625	1.7212	1.1488	0.0108	0.9946
ADBE	82.1131	9.0616	5.4440	0.0189	0.9972

Table 21: Phase 1 RF (Window Size: 250)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	4.2812	2.0691	1.2202	0.0193	0.9976
MSFT	11.8134	3.4371	2.0778	0.0164	0.9982
AMZN	7.9094	2.8124	1.6940	0.0204	0.9970
NVDA	23.1911	4.8157	2.2429	0.0313	0.9952
GOOGL	2.6547	1.6293	1.0258	0.0160	0.9973
TSLA	51.3998	7.1694	2.8559	0.0363	0.9942
GOOG	2.6619	1.6315	1.0209	0.0160	0.9974
BRK-B	12.9534	3.5991	2.3230	0.0119	0.9941
META	60.8289	7.7993	4.0276	0.0214	0.9899
UNH	31.0414	5.5715	3.5657	0.0161	0.9970
XOM	1.9038	1.3798	0.9026	0.0179	0.9758
LLY	13.3214	3.6499	2.0214	0.0173	0.9957
JPM	5.1032	2.2590	1.4307	0.0164	0.9951
JNJ	3.2780	1.8105	1.2578	0.0111	0.9946
V	10.6924	3.2699	2.0460	0.0151	0.9967
PG	2.5404	1.5939	1.0652	0.0118	0.9964
MA	34.2463	5.8520	3.5899	0.0170	0.9967
AVGO	69.5288	8.3384	5.1803	0.0232	0.9956
HD	25.0622	5.0062	2.8269	0.0157	0.9957
CVX	4.6709	2.1612	1.3918	0.0175	0.9754
MRK	1.3075	1.1434	0.7640	0.0136	0.9917
ABBV	3.3600	1.8330	1.1974	0.0182	0.9932
COST	26.0536	5.1043	3.2839	0.0141	0.9977
PEP	3.4384	1.8543	1.1645	0.0111	0.9940
ADBE	88.3784	9.4010	5.4194	0.0193	0.9970

Table 22: Phase 1 RF (Window Size: 125)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	760.948	27.585	19.543	0.2374	0.5997
MSFT	2354.711	48.525	39.077	0.2574	0.6400
AMZN	938.687	30.638	23.727	0.2550	0.5848
NVDA	2894.881	53.804	36.500	0.4219	0.4158
GOOGL	544.284	23.330	16.052	0.1939	0.4288
TSLA	5801.713	76.169	40.840	0.3197	0.4336
GOOG	507.750	22.533	15.807	0.1947	0.4900
BRK-B	1043.664	32.306	25.320	0.1143	0.4361
META	2604.444	51.034	41.251	0.1932	0.4682
UNH	3620.035	60.167	50.521	0.1986	0.5770
XOM	60.217	7.760	5.232	0.1090	0.3517
LLY	1452.013	38.105	26.499	0.1751	0.5494
JPM	431.608	20.775	16.660	0.1637	0.4885
JNJ	198.306	14.082	12.300	0.1001	0.5057
V	825.416	28.730	25.602	0.1756	0.6938
PG	225.533	15.018	12.281	0.1196	0.6678
MA	2768.732	52.619	45.065	0.1994	0.6989
AVGO	6686.048	81.768	62.863	0.2213	0.5229
HD	2047.074	45.245	36.743	0.1741	0.6044
CVX	155.119	12.455	9.594	0.1110	-0.0386
MRK	57.915	7.610	5.701	0.0936	0.5528
ABBV	274.866	16.579	12.686	0.1603	0.3765
COST	4042.247	63.579	49.627	0.1761	0.6397
PEP	193.844	13.923	11.674	0.1010	0.5942
ADBE	9642.112	98.194	80.918	0.2649	0.6443

Table 23: Phase 1 SVR (Window Size: 500)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	229.118	15.137	10.134	0.1453	0.8756
MSFT	651.793	25.530	18.569	0.1316	0.9033
AMZN	307.455	17.534	11.643	0.1475	0.8792
NVDA	1142.396	33.799	20.540	0.2787	0.7676
GOOGL	165.168	12.852	8.545	0.1133	0.8317
TSLA	2044.340	45.214	21.767	0.2122	0.7797
GOOG	150.546	12.270	8.305	0.1123	0.8532
BRK-B	408.150	20.203	14.889	0.0713	0.8119
META	949.470	30.813	23.735	0.1239	0.8323
UNH	1143.635	33.818	27.766	0.1194	0.8839
XOM	45.279	6.729	4.886	0.0989	0.4392
LLY	510.967	22.605	14.958	0.1114	0.8350
JPM	218.714	14.789	10.517	0.1092	0.7772
JNJ	74.158	8.611	6.915	0.0585	0.8693
V	238.192	15.433	12.940	0.0968	0.9230
PG	91.041	9.542	7.260	0.0768	0.8742
MA	772.261	27.790	22.279	0.1094	0.9239
AVGO	2309.319	48.055	34.760	0.1382	0.8485
HD	680.299	26.083	20.048	0.1032	0.8781
CVX	104.044	10.200	7.820	0.0967	0.4756
MRK	26.072	5.106	3.839	0.0676	0.8319
ABBV	114.135	10.683	7.944	0.1095	0.7641
COST	1604.612	40.058	26.927	0.0997	0.8590
PEP	74.180	8.613	6.623	0.0591	0.8655
ADBE	3394.044	58.258	41.110	0.1414	0.8828

Table 24: Phase 1 svr (Window Size: 250)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	95.552	9.775	6.315	0.1008	0.9471
MSFT	210.555	14.511	9.955	0.0767	0.9687
AMZN	118.660	10.893	7.052	0.0919	0.9557
NVDA	444.163	21.075	11.451	0.1622	0.9084
GOOGL	51.125	7.150	4.892	0.0705	0.9488
TSLA	1042.077	32.281	14.197	0.1587	0.8822
GOOG	50.279	7.091	4.854	0.0702	0.9516
BRK-B	205.647	14.340	10.639	0.0532	0.9062
META	462.181	21.498	15.313	0.0824	0.9233
UNH	433.410	20.819	15.630	0.0718	0.9585
XOM	22.935	4.789	3.400	0.0684	0.7081
LLY	215.448	14.678	9.065	0.0734	0.9302
JPM	97.672	9.883	6.718	0.0735	0.9056
JNJ	39.614	6.294	5.063	0.0443	0.9344
V	98.583	9.929	7.701	0.0594	0.9699
PG	39.541	6.288	4.883	0.0541	0.9444
MA	326.643	18.073	13.402	0.0672	0.9689
AVGO	1024.541	32.008	22.111	0.0969	0.9358
HD	333.556	18.264	12.630	0.0694	0.9428
CVX	54.191	7.361	5.519	0.0679	0.7144
MRK	11.328	3.366	2.583	0.0464	0.9280
ABBV	53.073	7.285	5.433	0.0798	0.8925
COST	590.398	24.298	16.652	0.0686	0.9486
PEP	29.800	5.459	4.280	0.0402	0.9484
ADBE	1519.951	38.987	24.567	0.0839	0.9484

Table 25: Phase 1 SVR (Window Size: 125)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	7.1209	2.6685	1.7417	0.0184	0.9978
ABBV	5.4981	2.3448	1.5345	0.0149	0.9948
ADBE	124.3205	11.1499	7.3068	0.0213	0.9953
AMZN	11.2442	3.3532	2.1909	0.0209	0.9943
AVGO	253.1454	15.9105	8.7970	0.0206	0.9951
BRK-B	17.5211	4.1858	2.7945	0.0117	0.9959
COST	65.1361	8.0707	5.1446	0.0144	0.9972
CVX	10.5331	3.2455	1.9631	0.0170	0.9848
GOOG	4.9484	2.2245	1.4810	0.0174	0.9961
GOOGL	4.9201	2.2181	1.4777	0.0174	0.9960
HD	27.7662	5.2694	3.4332	0.0139	0.9950
JNJ	4.3526	2.0863	1.4833	0.0102	0.9883
JPM	6.7007	2.5886	1.8293	0.0159	0.9919
LLY	70.7876	8.4135	4.1139	0.0174	0.9968
MA	43.7399	6.6136	4.6163	0.0172	0.9960
META	74.7306	8.6447	5.0200	0.0245	0.9869
MRK	2.0637	1.4366	0.9981	0.0127	0.9936
MSFT	21.3135	4.6167	3.0999	0.0164	0.9979
NVDA	126.5226	11.2482	5.2987	0.0358	0.9932
PEP	3.8799	1.9697	1.3720	0.0099	0.9947
PG	3.5829	1.8928	1.3111	0.0110	0.9953
TSLA	79.7632	8.9310	4.6258	0.0374	0.9937
UNH	46.8457	6.8444	4.7096	0.0149	0.9974
V	12.5749	3.5461	2.4794	0.0145	0.9960
XOM	3.9564	1.9891	1.2859	0.0183	0.9900

Table 26: Phase 2 RF (Window Size: 500)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	6.5295	2.5553	1.6447	0.0189	0.9980
ABBV	5.6317	2.3731	1.5497	0.0159	0.9949
ADBE	119.4916	10.9312	6.9191	0.0213	0.9960
AMZN	11.8105	3.4366	2.1069	0.0221	0.9950
AVGO	244.8252	15.6469	8.5611	0.0224	0.9954
BRK-B	17.0762	4.1323	2.7250	0.0119	0.9964
COST	64.4071	8.0254	5.0205	0.0147	0.9973
CVX	10.6573	3.2645	2.0219	0.0181	0.9849
GOOG	4.8757	2.2081	1.3972	0.0175	0.9965
GOOGL	5.0181	2.2401	1.4229	0.0178	0.9962
HD	27.4988	5.2439	3.3529	0.0144	0.9957
JNJ	4.5282	2.1280	1.4824	0.0106	0.9916
JPM	7.6607	2.7678	1.8982	0.0170	0.9924
LLY	61.9343	7.8698	3.8147	0.0176	0.9970
MA	38.2326	6.1832	4.1170	0.0163	0.9970
META	75.6348	8.6968	4.9115	0.0252	0.9882
MRK	2.3039	1.5179	1.0360	0.0138	0.9934
MSFT	21.8812	4.6777	3.0252	0.0170	0.9980
NVDA	92.8296	9.6348	4.5915	0.0352	0.9949
PEP	4.1084	2.0269	1.3719	0.0102	0.9949
PG	3.5967	1.8965	1.3182	0.0116	0.9957
TSLA	84.4903	9.1919	4.5375	0.0391	0.9932
UNH	43.6701	6.6083	4.4611	0.0152	0.9978
V	11.4802	3.3882	2.3166	0.0144	0.9969
XOM	4.1174	2.0291	1.3507	0.0189	0.9884

Table 27: Phase 2 RF (Window Size: 250)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	6.7236	2.5930	1.6556	0.0197	0.9979
ABBV	6.3565	2.5212	1.6668	0.0174	0.9944
ADBE	127.3319	11.2841	6.9035	0.0216	0.9959
AMZN	11.8042	3.4357	2.0765	0.0222	0.9953
AVGO	232.6461	15.2527	8.4099	0.0233	0.9957
BRK-B	18.4376	4.2939	2.8465	0.0128	0.9962
COST	72.4251	8.5103	5.2342	0.0161	0.9970
CVX	9.9635	3.1565	2.0819	0.0185	0.9852
GOOG	5.4435	2.3331	1.4341	0.0183	0.9962
GOOGL	5.2965	2.3014	1.4182	0.0181	0.9961
HD	33.2160	5.7633	3.5329	0.0157	0.9951
JNJ	4.7762	2.1855	1.5489	0.0112	0.9917
JPM	8.3290	2.8860	1.9238	0.0176	0.9923
LLY	66.8249	8.1746	3.9801	0.0192	0.9967
MA	44.9163	6.7020	4.3818	0.0179	0.9966
META	75.7376	8.7027	4.9571	0.0253	0.9888
MRK	2.3942	1.5473	1.0597	0.0143	0.9931
MSFT	21.6485	4.6528	2.9546	0.0172	0.9980
NVDA	119.4316	10.9285	5.0038	0.0381	0.9934
PEP	4.4750	2.1154	1.4151	0.0107	0.9947
PG	4.2950	2.0724	1.4143	0.0127	0.9948
TSLA	95.8000	9.7877	4.6504	0.0417	0.9921
UNH	43.5627	6.6002	4.4270	0.0157	0.9979
V	12.6816	3.5611	2.3814	0.0153	0.9969
XOM	4.4415	2.1075	1.4327	0.0198	0.9872

Table 28: Phase 2 RF (Window Size: 125)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	257.0731	16.0335	12.5498	0.1416	0.9196
ABBV	174.1265	13.1957	10.5112	0.0999	0.8346
ADBE	4899.0306	69.9931	61.1305	0.1957	0.8140
AMZN	347.3804	18.6381	16.0425	0.1786	0.8239
AVGO	7907.0305	88.9215	70.2789	0.1776	0.8482
BRK-B	612.0738	24.7401	21.1049	0.0885	0.8570
COST	2681.4323	51.7825	41.7031	0.1211	0.8835
CVX	53.6665	7.3257	5.7515	0.0495	0.9224
GOOG	166.0992	12.8879	10.4999	0.1325	0.8699
GOOGL	168.4690	12.9796	10.5026	0.1314	0.8627
HD	1031.5598	32.1179	27.1252	0.1165	0.8143
JNJ	84.8004	9.2087	7.9679	0.0554	0.7719
JPM	201.8412	14.2071	12.3252	0.1067	0.7553
LLY	2576.0204	50.7545	35.2801	0.1435	0.8819
MA	1429.2243	37.8051	33.1124	0.1345	0.8693
META	1920.3017	43.8212	37.6640	0.1917	0.6635
MRK	56.6464	7.5264	6.0470	0.0756	0.8250
MSFT	1148.5700	33.8906	28.8332	0.1691	0.8852
NVDA	3441.8438	58.6672	39.8131	0.3060	0.8161
PEP	108.5375	10.4181	9.1951	0.0666	0.8505
PG	43.7513	6.6145	5.7831	0.0524	0.9422
TSLA	1640.3772	40.5016	24.4114	0.1980	0.8711
UNH	2248.2994	47.4162	41.2598	0.1410	0.8734
V	456.9818	21.3771	19.4272	0.1233	0.8546
XOM	29.2223	5.4058	4.2950	0.0623	0.9263

Table 29: Phase 2 SVR (Window Size: 500)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	176.7071	13.2931	10.1106	0.1206	0.9457
ABBV	130.7346	11.4339	9.1828	0.0934	0.8828
ADBE	3543.3229	59.5258	46.2784	0.1467	0.8804
AMZN	228.3505	15.1113	11.8663	0.1379	0.9026
AVGO	5495.5315	74.1319	54.3807	0.1401	0.8976
BRK-B	419.2033	20.4745	16.3286	0.0695	0.9120
COST	1384.1002	37.2035	28.1580	0.0834	0.9419
CVX	155.1733	12.4569	9.7741	0.0876	0.7803
GOOG	141.8861	11.9116	8.9702	0.1074	0.8971
GOOGL	140.7994	11.8659	8.9604	0.1071	0.8937
HD	672.2255	25.9273	21.1519	0.0928	0.8944
JNJ	59.2379	7.6966	6.5396	0.0472	0.8901
JPM	181.2251	13.4620	11.2289	0.0995	0.8205
LLY	1462.1973	38.2387	24.4865	0.1069	0.9297
MA	821.0864	28.6546	23.9089	0.1012	0.9349
META	1678.7492	40.9725	32.7012	0.1679	0.7383
MRK	37.9126	6.1573	4.7705	0.0624	0.8908
MSFT	730.8289	27.0338	21.2016	0.1195	0.9316
NVDA	2673.0805	51.7018	32.1861	0.2550	0.8538
PEP	67.5782	8.2206	6.8292	0.0507	0.9166
PG	70.9312	8.4221	6.9365	0.0623	0.9143
TSLA	1431.6294	37.8369	21.5913	0.1771	0.8840
UNH	1121.3818	33.4870	28.2699	0.1002	0.9439
V	270.4581	16.4456	14.2871	0.0923	0.9279
XOM	72.6980	8.5263	6.8700	0.0995	0.7955

Table 30: Phase 2 SVR (Window Size: 250)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	103.8566	10.1910	7.2147	0.0894	0.9681
ABBV	87.2636	9.3415	7.3453	0.0756	0.9226
ADBE	2004.0848	44.7670	31.7080	0.0986	0.9349
AMZN	144.7219	12.0300	8.6884	0.1003	0.9427
AVGO	2996.7988	54.7430	37.6169	0.0998	0.9452
BRK-B	250.2393	15.8190	12.3441	0.0537	0.9482
COST	825.1247	28.7250	21.0172	0.0653	0.9658
CVX	105.4549	10.2691	7.8369	0.0693	0.8432
GOOG	73.0659	8.5479	6.2546	0.0755	0.9487
GOOGL	71.4501	8.4528	6.2498	0.0760	0.9479
HD	437.5461	20.9176	16.0671	0.0709	0.9359
JNJ	46.9968	6.8554	5.7949	0.0421	0.9184
JPM	113.6639	10.6613	8.2167	0.0748	0.8955
LLY	835.4439	28.9040	16.8302	0.0754	0.9591
MA	454.8498	21.3272	16.6842	0.0709	0.9658
META	1029.2847	32.0825	23.6585	0.1209	0.8483
MRK	23.4999	4.8477	3.7248	0.0494	0.9318
MSFT	364.7712	19.0990	13.8233	0.0794	0.9664
NVDA	1424.3405	37.7404	21.7925	0.1716	0.9209
PEP	43.2368	6.5755	5.3337	0.0404	0.9484
PG	47.5184	6.8934	5.6627	0.0513	0.9419
TSLA	1036.8752	32.2005	17.5320	0.1533	0.9143
UNH	573.5831	23.9496	19.2934	0.0694	0.9727
V	148.7165	12.1949	9.8498	0.0640	0.9632
XOM	44.1057	6.6412	5.3255	0.0758	0.8732

Table 31: Phase 2 SVR (Window Size: 125)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	62.8812	7.9298	6.0362	0.0617	0.9760
ABBV	33.3093	5.7714	4.1897	0.0375	0.9619
ADBE	812.5449	28.5052	22.2855	0.0588	0.9496
AMZN	65.7649	8.1096	6.2472	0.0536	0.9394
AVGO	1821.6026	42.6802	29.0005	0.0727	0.9646
BRK-B	136.4863	11.6827	9.0940	0.0362	0.9594
COST	658.1691	25.6548	18.7343	0.0525	0.9631
CVX	47.4823	6.8907	5.1737	0.0419	0.9422
GOOG	40.3705	6.3538	4.8084	0.0522	0.9615
GOOGL	38.0874	6.1715	4.7319	0.0557	0.9624
HD	237.4816	15.4104	11.7514	0.0458	0.9342
JNJ	23.1915	4.8158	3.6008	0.0242	0.9116
JPM	50.6220	7.1149	5.2600	0.0424	0.8989
LLY	989.6219	31.4583	19.2682	0.1059	0.9560
MA	384.3395	19.6046	14.9071	0.0521	0.9196
META	384.5136	19.6090	14.2977	0.0665	0.9291
MRK	14.9464	3.8661	2.8086	0.0337	0.9427
MSFT	303.8732	17.4320	12.6823	0.0646	0.9584
NVDA	824.9644	28.7222	19.0863	0.1793	0.9579
PEP	32.6882	5.7174	4.2806	0.0303	0.9465
PG	23.7454	4.8729	3.8697	0.0310	0.9602
TSLA	400.0830	20.0021	14.0888	0.1912	0.9683
UNH	443.6770	21.0636	15.9788	0.0455	0.9651
V	104.2043	10.2080	7.7677	0.0421	0.9234
XOM	17.4472	4.1770	3.1558	0.0426	0.9655

Table 32: Phase 2 LSTM (Window Size: 500)

Stock	**MSE**	**RMSE**	**MAE**	**MAPE**	**R^2 Score**
AAPL	47.5678	6.8969	5.0669	0.0711	0.9851
ABBV	37.3434	6.1109	4.3936	0.0454	0.9645
ADBE	743.1607	27.2610	20.0569	0.0731	0.9718
AMZN	55.8677	7.4745	5.4859	0.0630	0.9717
AVGO	1203.9912	34.6986	24.8650	0.0733	0.9767
BRK-B	133.8750	11.5704	8.1648	0.0356	0.9686
COST	542.4784	23.2912	17.1119	0.0604	0.9764
CVX	38.2927	6.1881	4.3641	0.0376	0.9446
GOOG	26.9083	5.1873	3.6590	0.0473	0.9789
GOOGL	28.5605	5.3442	3.9096	0.0524	0.9767
HD	208.7608	14.4486	10.5383	0.0475	0.9624
JNJ	20.8770	4.5691	3.4012	0.0239	0.9440
JPM	36.6922	6.0574	4.5984	0.0414	0.9555
LLY	499.1659	22.3420	14.0142	0.0865	0.9770
MA	261.7247	16.1779	11.6252	0.0507	0.9761
META	246.6788	15.7060	11.4853	0.0564	0.9565
MRK	9.5393	3.0886	2.3328	0.0304	0.9704
MSFT	217.9798	14.7641	10.7544	0.0775	0.9782
NVDA	548.7317	23.4250	15.4350	0.3364	0.9704
PEP	19.8788	4.4586	3.3651	0.0250	0.9726
PG	20.1961	4.4940	3.1637	0.0275	0.9733
TSLA	348.4420	18.6666	11.3143	0.2319	0.9726
UNH	330.9058	18.1908	13.7112	0.0481	0.9814
V	78.7866	8.8762	6.6256	0.0424	0.9749
XOM	14.9406	3.8653	2.8891	0.0373	0.9623

Table 33: Phase 2 LSTM (Window Size: 250)

Stock	MSE	RMSE	MAE	MAPE	R^2 Score
AAPL	50.9809	7.1401	4.7113	0.0658	0.9843
ABBV	23.7968	4.8782	3.5822	0.0377	0.9786
ADBE	549.9527	23.4511	16.3250	0.0674	0.9814
AMZN	55.2921	7.4359	5.4095	0.0751	0.9764
AVGO	1510.2680	38.8622	24.5023	0.0762	0.9717
BRK-B	94.6963	9.7312	7.2756	0.0329	0.9801
COST	415.5707	20.3856	14.1434	0.0509	0.9825
CVX	35.0556	5.9208	4.3162	0.0369	0.9503
GOOG	21.1273	4.5964	3.3245	0.0488	0.9847
GOOGL	23.8083	4.8794	3.5287	0.0495	0.9820
HD	173.8137	13.1838	9.5955	0.0449	0.9727
JNJ	17.4845	4.1814	3.2300	0.0235	0.9676
JPM	28.9484	5.3804	4.0406	0.0385	0.9713
LLY	511.8202	22.6234	12.9184	0.0915	0.9752
MA	242.9708	15.5875	11.8005	0.0548	0.9807
META	213.8090	14.6222	10.5195	0.0595	0.9665
MRK	10.0306	3.1671	2.3393	0.0322	0.9710
MSFT	151.0715	12.2911	8.9170	0.0751	0.9858
NVDA	502.3594	22.4134	13.3284	0.4866	0.9722
PEP	20.3102	4.5067	3.3882	0.0261	0.9749
PG	14.2980	3.7813	2.9063	0.0257	0.9827
TSLA	210.1645	14.4971	9.2889	0.2025	0.9830
UNH	296.6599	17.2238	12.7664	0.0479	0.9852
V	93.8787	9.6891	6.8673	0.0510	0.9750
XOM	13.9752	3.7383	2.7928	0.0360	0.9606

Table 34: Phase 2 LSTM (Window Size: 125)

Ticker	Window Size	MSE	RMSE	MAE	MAPE	R^2
AAPL	251	0.4237	0.6509	0.4393	0.0106	0.9967
AAPL	125	0.4093	0.6397	0.4363	0.0111	0.9968
AMZN	251	1.4317	1.1965	0.7664	0.0119	0.9974
AMZN	125	1.3073	1.1434	0.7263	0.0122	0.9979
GOOG	251	0.5328	0.7299	0.4913	0.0098	0.9938
GOOG	125	0.5305	0.7283	0.4869	0.0102	0.9949
MSFT	251	1.689	1.2996	0.8744	0.0097	0.9981
MSFT	125	1.5803	1.2571	0.8407	0.0099	0.9984
TSLA	251	0.3044	0.5517	0.3800	0.0204	0.9794
TSLA	125	0.2929	0.5412	0.3741	0.0204	0.9791

Table 35: Phase 3 Naive Model Various Window Sizes

Ticker	Window Size	MSE	MAE	MAPE	R^2
AAPL	251	1.010	0.709	0.0168	0.9922
AAPL	125	1.001	0.712	0.0177	0.9922
AMZN	251	2.446	1.083	0.0168	0.9956
AMZN	125	2.626	1.094	0.0183	0.9958
GOOG	251	0.908	0.684	0.0136	0.9895
GOOG	125	0.946	0.695	0.0146	0.9909
MSFT	251	3.010	1.261	0.0138	0.9967
MSFT	125	2.729	1.190	0.0139	0.9972
TSLA	251	0.554	0.529	0.0287	0.9627
TSLA	125	0.674	0.560	0.0312	0.9521

Table 36: Phase 3 RF Various Window Sizes

Ticker	Window Size	MSE	MAE	MAPE	R^2
AAPL	251	19.226	3.873	0.0982	0.8505
AAPL	125	16.040	3.335	0.0831	0.8756
AMZN	251	63.616	6.768	0.1176	0.8862
AMZN	125	36.193	5.004	0.0905	0.9417
GOOG	251	16.030	3.581	0.0744	0.8153
GOOG	125	9.140	2.575	0.0545	0.9123
MSFT	251	120.980	9.874	0.1082	0.8675
MSFT	125	53.838	6.228	0.0709	0.9450
TSLA	251	1.658	1.043	0.0584	0.8884
TSLA	125	3.062	1.330	0.0754	0.7824

Table 37: Phase 3 SVR Various Window Sizes

Ticker	Window Size	MSE	MAE	MAPE	R^2
AAPL	251	9.668	2.177	0.0492	0.8689
AAPL	125	4.997	1.693	0.0427	0.9609
AMZN	251	26.611	3.807	0.0560	0.9284
AMZN	125	24.881	3.517	0.0603	0.9555
GOOG	251	6.172	1.992	0.0378	0.8665
GOOG	125	6.943	1.910	0.0405	0.9198
MSFT	251	41.335	4.763	0.0502	0.9394
MSFT	125	30.086	3.866	0.0474	0.9669
TSLA	251	2.099	1.150	0.0577	0.7834
TSLA	125	1.921	1.039	0.0561	0.8698

Table 38: Phase 3 LSTM Various Window Sizes

Ticker	Window Size	MSE	MAE	MAPE	R^2
AAPL	251	0.9332	0.6883	0.0163	0.9927
AAPL	125	0.9546	0.6989	0.0175	0.9926
AMZN	251	2.4309	1.0731	0.0167	0.9957
AMZN	125	2.5872	1.0779	0.0180	0.9958
GOOG	251	0.9144	0.6891	0.0137	0.9895
GOOG	125	0.9324	0.6911	0.0146	0.9911
MSFT	251	2.9264	1.2384	0.0136	0.9968
MSFT	125	2.6517	1.1667	0.0137	0.9973
TSLA	251	0.5699	0.5359	0.0289	0.9616
TSLA	125	0.6348	0.5517	0.0308	0.9549

Table 39: Phase 3 RF Various Window Sizes, Without Sentiment

Ticker	Window Size	MSE	MAE	MAPE	R^2
AAPL	251	19.0477	3.8649	0.0980	0.8519
AAPL	125	15.7579	3.3149	0.0827	0.8777
AMZN	251	62.6174	6.7343	0.1172	0.8880
AMZN	125	35.6971	4.9759	0.0902	0.9425
GOOG	251	15.3742	3.4861	0.0719	0.8228
GOOG	125	9.0972	2.5684	0.0543	0.9127
MSFT	251	119.6830	9.8334	0.1079	0.8689
MSFT	125	53.5654	6.2178	0.0708	0.9453
TSLA	251	1.5315	0.9806	0.0545	0.8969
TSLA	125	2.9957	1.3078	0.0740	0.7872

Table 40: Phase 3 SVR Various Window Sizes, Without Sentiment

Ticker	**Window Size**	**MSE**	**MAE**	**MAPE**	**R^2**
AAPL	251	9.9697	2.2382	0.0504	0.8648
AAPL	125	8.5362	1.8041	0.0481	0.9332
AMZN	251	33.7742	4.5559	0.0663	0.9092
AMZN	125	21.5762	3.4299	0.0580	0.9614
GOOG	251	10.4418	2.3294	0.0446	0.7741
GOOG	125	5.3953	1.7823	0.0365	0.9377
MSFT	251	33.5280	4.2385	0.0438	0.9508
MSFT	125	37.7577	4.3244	0.0506	0.9585
TSLA	251	2.6148	1.2632	0.0635	0.7302
TSLA	125	1.7568	0.9978	0.0537	0.8809

Table 41: Phase 3 LSTM Various Window Sizes, Without Sentiment