CSC3094: Major Project and Dissertation

Academic Year 2023/24

BSc Computer Science

Predictive Modelling for Stock Trading

Project Supervisor: Dr. Daniel Archambault

Word Count: 14427

Student Name: UTSAV KAILASH KOTHARI

Student ID: 210100637

Abstract

This dissertation deals with the empirical use of 'hybrid' machine learning algorithms and 'sentiment analysis' for predicting stock market movement, particularly the Nifty 50 and Sensex quoted indices in the Indian market. The key goal is to create a tool that receives information through the power of modern data analytics and natural language processing to strengthen the quality of investment decisions and control the risk level.

The project is presenting a hybrid model mapping two key techniques, LSTM networks, and ARIMA models, besides sentiment analysis derived from financial news. The objective of this framework is to bring together both the quantitative and qualitative characteristics of the market which are so essential for determining the range of stock prices that remain constantly active.

Whether studied for its sustainability, cost-effectiveness, or educational quality, the integration of these practices notes a marked increase. The LSTM model will deliver superior performance on the market data interpretation in terms of the time series patterns. Sentiment analysis, at the same time, fills in the emotional aspect by integrating psychological and emotional factors that produce price cognitions. Together, they illuminate the stock price forecast experience, contributing to the quality and trustworthiness of the results.

Table of Contents

Chapter 1: Introduction

The generic market has a paramount function in the worldwide financial system wherein companies get access to capital and investors look for their own growth via investments. Although the stock prices are dynamic, being a combination of factors like economic indicators including investor sentiment, this becomes a point of focus and interest for investors resulting in a study area. The process of understanding what the factors that direct the stock market involves is very important to both the companies and the investors, as it affects the decision-making relating to capital allocation and investment plans (Guo, 2022).

In recent times, predictive modelling for stock trading has caught the attention of the mind as a vital field and it is mostly built by machine learning and data analytics. It is particularly such an aim that desires to contribute to the increased body literature through this dissertation that predictive modelling techniques applied to the Indian stock market are explored. This research may show how machine learning algorithms and sentiment analysis can do the forecasting of stock prices. Investors who may not get a grasp of the complexity of this process will be as well provided with relevant insights (Beyaz, 2018).

1.1 Motivation

People's attraction to the stock market is based on its very nature of being unpredictable and possessing a real possibility of financial advancement. My inspiration in this field was formed from my childhood, as starting from my teenage years, I observed my father with his scientific research and investing in stock. Evident in his interest in understanding the underlying numbers of financial reports and what led to market fluctuations, we observed his passion for stock trading as something complicated and thrilling. It was during this period that my interest in finance and investments was sparked; it gave birth to a passion that I hold to date.

The benefit of predictive modelling for stock trading is that it could be used for changing strategies of investments by applying advanced technologies like machine learning. In that particular industry where rapid implementation and knowledge are supreme, offering a prediction tool for market trends will enable you to have an edge over others. This issue, indeed, has become a matter of urgent concern in present times as the young and beginners find it increasingly difficult to rise above their inexperience and limited knowledge about financial concepts (Baker & Wurgler, 2006).

To take this research a step further it is aimed at the filling of the existing information deficit for stock trading by young investors. First-hand cases of investing blindly and later recording financial loss due to unsound trading practices demonstrated the paramount importance of complying with all the rules and regulations laid down for such activity and the danger of risk speculation. It is this dissertation's intent to give reliable answers by developing a unique prediction modelling system that is specifically designed for the Indian stock market. It would therefore not only assist in making decisions regarding investment but also enable investors in the industry to make more informed and strategic decisions.

1.2 Aims/Hypothesis and Objectives

Aim:

To design a robust tool using hybrid machine learning algorithms and sentiment analysis to predict stock market trends in the Indian stock market, focusing on the NIFTY 50 and Sensex indices.

Hypotheses:

1. Machine learning-based predictive modelling can accurately forecast stock prices in the Indian stock market.
2. Sentiment analysis of financial news can improve the accuracy of stock price predictions.

Objectives:

- **Developing a Predictive Model:** Create a flexible model using prior stock prices, trading volumes, and economic news to forecast future stock prices, accounting for Indian stock market characteristics.
- **Sentiment Analysis:** Integrate sentiment analysis into the model to measure investor sentiment and its impact on stock prices, using natural language processing techniques.

- **Risk Assessment Framework:** Develop a risk assessment framework within the forecasting system using recent volatility prediction algorithms to guide portfolio risk management and inform investment strategies.

1.3 Dissertation Structure

This dissertation is organised into several chapters, each addressing a critical aspect of the research project. The first chapter, the Introduction, sets the stage by outlining the topic, motivation, and objectives of the study. It provides context for the reader and highlights the significance of the research.

Chapter Two presents the Literature Review, which examines relevant research and technical material in the field of stock market prediction. This chapter explores the current state of knowledge and identifies gaps that this dissertation seeks to address.

Chapter Three, the Methodology, outlines the research design and methods employed in the project. This includes a description of the data collection, pre-processing, feature selection, and model selection techniques. The methodology chapter also discusses the ethical considerations involved in the research.

Chapter Four details the Results, describing the outputs produced by the predictive model. The evaluation of these results and the assessment of the engineering or design approach are also covered.

In Chapter Five, the Discussion, the findings are interpreted in the context of the original objectives. This chapter examines any unexpected results and their implications.

The final chapter, Chapter Six, concludes the dissertation by summarising the key findings, discussing the extent to which the aims and objectives were met, and suggesting areas for future research.

The dissertation ends with a list of References and relevant Appendices, ensuring the reader has access to additional information and supporting materials.

To summarise, this dissertation centres on predictive modelling for the stock market in India, features the integration of machine learning and sentiment analysis techniques to predict stock prices. The introduction describes inspiration to study, research objectives, hypotheses to be used in investigation as well as study parameters which give the reader a clear roadmap for the rest of the work. Through density investment strategy the informing , understanding and applying of the modern analytics tools and techniques, this dissertation is aimed at providing some solutions and behind it insights for both the individual and institutional investors.

The following chapters will explore in depth about the existing literature, methodologies and outputs used to make a comprehensive picture of the project. Moving on to the next section, a literature review will be made encompassing the current understanding of prediction in the stock market and will be completed by pointing out the research's contribution areas.

Chapter 2: Literature Review

In the dynamite world of stock trading, predictive models are important instruments that investors use to decide on the purchase or sale of stocks or other financial instruments. The review of the literature has the aim of disclosing the usage of machine learning techniques for the task of a share price prediction as the target of building a predictive model for the stock market of India. The review evaluates different models and methods that are used in predictive analysis and gauges their efficacies along with possible bottlenecks. While reviewing the existing literature in this chapter the spotlight will be on the identification of areas that need more focus and establish the place and meaning of the dissertation.

2.1 Thematic Review

2.2.1 Machine Learning in Algorithmic Trading

As Beyaz (2018) states, machine learning unquestionably plays a major role in algorithmic trading that enables predictive models to locate promising signals from market data. These flags are part of the structure to be used in navigation to system trading strategies which in turn will lead investors to take correct actions. The predictive modelling strategies depend on the accuracy of stock price prediction. It is important to focus on conditions of the marketplace that fit with the goals of the project to create a model for the stock market that consists of unique volatility patterns and growth. Jansen mentioned that predictive models have specialised roles of how to create useful guidance which marketing strategy issues that need to be processed urgently. This agrees with the main topic of this thesis which is titled Predictive Modelling for Stock Trading.

As per Guo (2022), machine learning algorithms can process huge sets of market indicators to bring out intricate patterns and laws which human imagination is unable to figure out in simple terms. The use of machine learning in algorithmic trading allows for the automation of trading decisions based on the identified patterns, which leads to an increase in the efficiency and profitability of trading strategies. For securities trading, this approach is very handy, especially when making quick decisions is so important.

According to Wang et al. (2021), in such a model CNN-BiLSTM, stock closing prices can be predicted. Their study shows the effect of machine learning models with multiple input options to achieve better predictive accuracy in the predictive algorithms of trading. This fits well with the view of the dissertation to combine both classical and artificial intelligence techniques to build the brain of Indian stock market models.

Just like Lin et al. (2021) introduce an innovative method of multivariate temporal routed models and optimal transport for learning multiple trading patterns of stock trading. By using this technique, different types of trade patterns can be easily comprehended. In turn, these notions serve as the basis for more precise forecasting models. The decision to focus my research on algorithms during the stock market trading hours represents a close link to the dissertation's intention to provide a comprehensive understanding of stock trading patterns.

Aldhyani & Alzahrani (2022) introduce a deep learning model for stock price prediction, and the study shows deep learning models such as Recurrent Neural Networks are effective for modelling time series stock prices. This architecture holds the same intention of the dissertation which is to heed the algorithmic trading field by means of having a sophisticated predictive system in place.

Another notable contribution of Obthong et al. (2020) is the survey of machine learning algorithms used in stock price prediction, wherein they highlight the main techniques and their applications. Their questionnaire allows it to elicit the advantages and shortcomings of sorts of algorithms used that could be taken into account to enable the best ones to be selected as models in this research work.

Additionally, the literature confirms that machine learning is the most important part of algorithmic trading due to the effectiveness of its trading strategies. This comes in line with the research so necessary for predictive modelling for stock trading, basically in the Indian context.

2.1.2 Hybrid Deep Learning-Based Predictive Models

According to (Kanwal et al. 2022), hybrid deep learning models have been seen to exhibit promising results in price forecasting. They developed the BiCuDNNLSTM-1dCNN model, which is a hybrid of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures, to achieve higher prediction accuracy. This mixed approach benefits from LSTM together with CNN models, which outshine each other performing two particular functions: LSTM is good at catching temporal dependencies and CNN is good at spatial image captures.

Wang et al. (2021) introduced a similar hybrid model, CNN-BiSLSTM, for stock closing price prediction. Their main idea was curriculum learning that aimed to improve stock price prediction. This hybrid model provides the CNN predictive power with the powerful and backward direction prediction term of the BIDI LSTM, which produces better reliability in estimating the data. Spatial and temporal data modelling is a powerful tool; it is because of this ability that such models are particularly effective for stock price prediction. As per Jayalakshmi, et al., (2019), there are no stock prices without time and space; such price systems have thus incorporated the sequential and intricate aspects of time in stock prices and their patterns. This begins to explore the same tendency reflected in this dissertation which has its purpose in creating hybrid models suitable for prediction of the Indian stock market.

Aldhyani and Alzahrani (2022) engineered deep learning methods in their framework for forecasting stock market prices. Together, their work shows that models can be improved by using combinations of different neural network structures. In a complex environment as stock trading, the accuracy of models is clearly improved. By integrating such methodology, the hybrid model becomes more effective in tackling the complexities and unique qualities associated with the Indian stock market aligning with the thesis.

Obthong et al. (2020) investigated the possibilities of machine learning algorithms for the prediction and recognition of stock prices with emphasis on the significant contribution of hybrid models to the improvement of accuracy. The general stresses the necessity of hybrid architectures that are cost-effective in capturing multiple aspects of stock market data. The dissertation goes deeper than that by focusing on dedicated architectures for accuracy in predictive analysis. It is important to adapt the approach to the unique market structure of India as their investor sentiment is important (Baker & Wurgler, 2006).

In other circumstances, Sauds and Shakras (2020) evaluated the performance of LSTM variants for stock price forecasts, giving more focus to the deep hybrid architecture. The researchers had illustrated this in their case study on the banking sector, through the use of a range of neural network models. Relevance of this to the current endeavour in which an accurate stock market predicting descriptive model is desired.

The literature is in agreement with the fact that hybrid deep learning-based predictive models perform exceptionally well, combining different neural network architectures strengths. A step toward the proposed dissertation that emphasises predictive analysis of the Indian stock market by means of Hybrid Models is this.

2.1.3 Long Short-Term Memory Networks (LSTM)

Focusing on the findings of Ta et al. (2020), it can be seen that Long Short-Term Memory (LSTM) networks are highly effective in capturing the temporal dependencies in stock data that can be used in portfolio optimization. The study underlines that LSTM networks are significant for quants trading and predicting the stocks of the Indian market. LSTM networks do well in recognizing trends, and temporal patterns and, therefore, stock price prediction accuracy depends on the trend recognition ability.

Besides this, Bhandari et al. (2022) touch on the adoption of LSTM networks in the prediction of the stock market indices. According to their studies, LSTMs are expert to be at position of long-term dependencies, which confirms a predictive approach to the key Indian stock market indices such as NIFTY 50 and Sensex. It is the LSTM-based method that is in tune with the goal of this dissertation to improve investment strategies in Indian shares via LSTM.

As per Wen et al. (2020), the merging of LSTM networks and Principal Component Analysis (PCA) could be an add-on for stock price anticipation. The authors proposed the PCA-LSTM model which can efficiently capture the principal and temporal characteristics, consequently, a very practical tool for predicting. This model emphasises the flexibility of deep learning techniques and the benefits they bring to the investment decision-making process.

The authors of the study conducted by Saud and Shakya (2020) look at the look-back period for stock price prediction with LSTM networks that are used to bank stock prices only. Their studies demonstrate that LSTM models serve the purpose depending on how the length of look-back is determined, which is related to the capability of the network to accumulate the entire necessary knowledge needed for inference. Therefore, through an understanding of this fact, reliable models for the prediction of the stock market in India can be developed since wealth sector-specific specificities contribute largely to this.

Wang et al. (2021) put forth a hybrid model based on CNN-BiSLSTM architecture, where CNN is utilised for feature extraction, and LSTM is utilised for the generation of precise closing price value prediction. In the study, the researchers offered evidence of the ability of LSTM architecture when combined with others to increase the prediction accuracy. This blended way of approach, following, in fact, the process which is the central theme of the research of the Indian stock market predictive models.

According to Pate (2020), LSTMs are key to capturing time series trends in Indian stock market data, being by far the most useful model for predictive profiling of stock movements. The literature underlines the significance of LSTM within quantitative trading and reveals the use of those in the algorithm that, in turn, leads to great investment strategies which is considered the goal of this dissertation.

2.2 Identifying Gaps

As per the report by Obthong et al. (2020), many studies concentrate only on the instances of machine learning algorithms that are used to predict the value of stock. However, some of the studies overlook market-specific factors. This gap becomes clear when considering the specific attributes of the Indian stock market that work under separate regulatory, economic and cultural frameworks compared to the other global regions. Among others, different regimes, societal effects on investor attitudes, as well as unique economic cycles are the characteristics that impact stock prices in unconventional ways that one may have not experienced in another market.

Wen et al. (2020) note that one of the most considerable factors of machine learning is that it depends on market conditions a lot more than theory-based investment; modelling for one market may not work for another. Such an approach helps determine the model-specific predictive patterns is essential. The absence of dedicated research into the Indian stock market causes a gap in the existing amount of knowledge, predominantly on how machine learning models can be reoriented or advanced to solve the specific tasks arising in this market.

Saud and Shakya (2020) opine that the effectiveness of such models as LSTM networks also depends on sector-specific particularities making the narrow-tailoring of predictive models to the real-world market environment very important. The Indian stock market needs to take into account the sectors that are risky and unique types of investor behaviour.

Aldhyani and Alzahrani (2022) point to the exceptional capability of deep learning algorithms in predictive model generation, however, they suggest that these models may yield varying results depending on the degree to which their construction adheres with market-specific factors. Here, the emphasis is on devising models which can reflect the development of the Indian market.

According to Bhandari et al. (2022), such customised models must be generated separately for each market, e.g., the Indian stock market. The traditional approach of machine learning using generic models is less liable to market changes and therefore results in lesser precise predictions, as is indicated in their research.

Hence, this research, as it was excellently described by Obthong, et al. (2020), can be viewed as a bridge which closes the gap of adaptive forecasting models that are tunable to the requirements of the Indian market. The research adopts the technique of machine learning algorithms in this specific market milieu which is unusual in the particular area of stock prediction; this area of the Indian market displays some of its unique characteristics.

2.3 Contextualization

A review of machine learning techniques being looked into in algorithmic trading, as well as hybrid deep learning-based models and Long Short-Term Memory (LSTM) networks can indicate that these technologies portray a prospect of improving stock price prediction. This study is being placed specifically in this context because it is going to take advantage of the advanced technologies purported to predict human behaviours accurately.

Jansen (2020) denotes that one of the most impactful features of machine learning which enables us to use it for systematic trading is the creation of insights that have practical implementation. The dissertation corresponds to this statement, with predictive modelling being the key area of study for the stock values in India. Through drawing a comparison between the works cited, there arises the fact that machine learning has great potential to improve the processes of trading.

Its hybrid deep learning models can be an option for the future and is working to redefine the way stock prices are predicted by buying the current models. In this sense, the dissertation conforms to this finding by outlining ways for the implementation of hybrid models for predictive analysis in the Indian stock market, utilising two neural network architectures simultaneously to ameliorate the accuracy.

Ta and Ta. (2020), and Bhandari et al. (2022) illustrated the prediction power of LSTM networks for stock portfolio optimization and index investing. It is in the continuity of these studies that the LSTM network has been used to modify investment strategies toward the NIFTY50 and Sensex.

Obthong et al. (2020), after observing many investigations that have no market-specific factors in stock prediction, have recognized the need to fill the gap in the research field. This study fills this knowledge gap by developing a customised forecasting system, considering multifaceted and defining features of the Indian financial market.

Our dissertation proposes the alignment of this research with the existing domain of stock price prediction, thereby, offering a contribution to the growing area of algorithms trading and deep learning, filling the voids in the literature by concentrating on the market-specific predictive models for the Indian stock market.

2.4 Evaluation and Synthesis

The used sources on machine learning and stock price prediction supply useful data that are taken into account and were deemed as good both in terms of quality and relevance. Every source is involved in a specific way as a researcher tries to combine them to grasp the point of prediction through which the dissertation's goal is realised.

Evaluation

According to Jansen (2020), there is a fairly detailed exposition of machine learning techniques in algorithmic trading enterprises. The handbook is famous for its numerous step-by-step trade illustrations. This is indeed why it is a very suitable source for our essay. The Hybrid framework authored by Kanwal et. al (2022), brings the latest development in research which supports the objective of the current dissertation by merging the different designs of neural network architecture.

The pieces by Ta et al. (2020) and Bhandari et al. (2022) address LSTM networks which are vital for capturing patterns over time and should not be ignored. Their purpose is to implement index prediction and portfolio optimization in a way that pays attention to the scope of the research concerning the Indian stock market.

Obthongo et al. (2020), authors of a comprehensive survey, demonstrate machine learning algorithms for stock prediction. The experience of fresh graduates is significant here as their work revealed gaps in some applications which this dissertation aims to fill.

Synthesis

The most valuable contribution from these unique sources can be attributed to the relevant role of machine learning and deep learning techniques in aiding with stock price prediction. Set up a hybrid model combining CNN and LSTM architectures, as Kanwal et al. (2022) suggest, to achieve greater accuracy by relying on the best in both worlds. The papers by Ta and co-authors (2020) and Bhandari et al. (2022) bring LSTM networks to the forefront, proving to be of utmost importance in realising the relations between time and stock market prices.

Implications for the Project

These results determine primarily the dissertation content. Regarding this project, the use of hybrid models and LSTM networks falls following the aim to create dominant and predictive technologies to be used in the stock market. Reading and analysing relevant scientific studies suggest that machine learning algorithms, especially the models dedicated to market-specific variables, noticeably increase stock price prediction accuracy, which is pivotal for the project.

Chapter 3: Methodology

The objective of this section is to elaborate on the methodology of building and verifying the stock market trend predictive models, which will be applied to the Indian market. The process entails data collection, data set preparation, variable selection, model training, testing, and evaluation. The goal that the chapter aspires to achieve is provided by describing the components that are essential for the research to be in line with the objectives.

Nti et al. (2020) demonstrate that fundamental and technical analysis are the keys to the successful prediction of stock market movements. Knowledge of demographic analysis and sentiment analysis developed alongside predictive modelling reveals a wide range of volatile parameters. The chapter discusses these advanced machine learning and deep learning techniques by Strader et al. (2020), and one of them is presented. The tool aims to meet the final objective of making a tool for stock market prediction.

3.1 Data Collection

Historical Price Data

In this study, historical stock prices have been accessed by the API of Angel One which is a defining source for financial data of the Indian market. The Angel One API displays several functionalities for acquiring market data, along with the facilities for attaining stock prices, trading volume, and different data elements(others). The choice of NIFTY 50 and SENSEX market capitalization index symbols was used for our study. The areas included our choice simply because we deemed them important for analysis and the instrumentality that others have expressed in their investment decisions, according to Nti et al. (2020).

The time interval selection for stock prices was data-dependent. The emphasis was on both the short-term and long-term trends, parallel to the demand for contributing to sophisticated analysis. Setting the period to a minimum of 1 interval means from minutes to months gives the research time to get through various market movements and trends. This accurate standpoint concerning data collection comes from the recommendation mentioned by Strader et. al (2020) regarding the use of detailed historical data as an effective predictive modelling tool.

News Data

News articles were collected using the web scraping process which primarily involved Beautiful Soup, a renowned Python library that helps to parse HTML and XML documents. The banking of news data was through the news portal known as Moneycontrol which is a popular financial news website in India. Sometimes the scraping process involved collecting a webpage, dissecting its structure, and screening the data.

The criteria for finding relevant articles from the news included words, like "NIFTY 50" and "Sensex", which are connected to the symbols of the shares. It is a field related to stock market prediction, which is mentioned in the recommendations of Nti et al (2020) and they stress the significance of sentiment analysis in this field. As part of the development of the predictive models, we include news on the markets, as this helps to capture the external factors which in turn trigger behaviour in the market.

The relevant articles were examined and subjected to sentiment analysis using the investors' sentiment in predictive models. It is here that the vision is made a reality by incorporating the news data processed in such a way that it fulfils the research objectives achieving the highest accuracy and relevance. With an accent on economically relevant financial news, the research targets the wider semantics of stock market prediction, which is an advantage in reaching the goal of a better predictive model.

3.2 Data Pre-processing

Cleaning and Handling Outliers

The beginning of data prepossessing that I undertook was focusing on the removal of anomalies and handling of outliers. The volatile nature of stock prices can therefore lead to cases whereby the outliers could act as a serious bias to the predictive models. The z-scores were used to formulate the z-scoring method, which allowed for the identification of outliers, due to their significant deviation from the mean value. The boundaries for outliers were 3 z-values, which are greater than absolute values for better exclusion. The assortment of outliers was removed from the data set specifically to guard against distortions during the ensuing analysis. This kind of approach matches with the recommendations of Zhang et al. (2022) who put much emphasis on the powerful outing detection machines needed in the financial data treatments.

The missing values were imputed using the mean imputation process, in which those missing data points were substituted with the average value of the respective characteristic parameter. This ultimate and yet plain recipe got rid of data questions among the study participants according to Yan et al. (2021). The use of mean imputation itself undermines the model's accuracy, which can be diminished in cases of incomplete data.

Normalisation

The aim of the given exercise was to make sure that all features of the model had the same importance, and that the stock prices were normalised. The technique of min-max scaling is used which normalises the data to the given range exactly [0, 1]. This method is therefore adopted, because it maintains the relationships among the data points and fits them on the same scale, though. According to Hou et al. in 2021, normalisation is a crucial part of the deep learning models to avoid the large-scale features playing a more dominant role in the learning process.

Sentiment Analysis

For its sentiment analysis, a VADER sentiment analysis tool was used to check the sentiment of the news articles. The VADER tool is a strong one that is focused on both social media and the sentiment analysis of the news, providing with compound sentiment scores that are an average of how a text is perceived. In this research, the sentences with compound sentiment scores were derived from the news pieces available through the electronic version of Moneycontrol. We fed the sentiment scores into the predictive model as an extra feature along with the results of Wu et al. (2012) that illustrate the relevance of sentiment analysis for forecasting stock prices.

3.3 Feature Selection

Feature Extraction

By using a library of different technical indicators, such as a lot of stock price indicators, it is possible to approach stock price prediction tasks. The ta library has one of the most complete sets of technical analysis indicators for free. These indicators may be added at any time to your strategy. Indicators of moving averages, RSI, MACD and Bollinger Bands are included. This benchmark was accounted for by introducing the appropriate indicators into the dataset, which are key factors for the stock price forecasting process because they reveal market dynamics and prevailing momentum. The choice of technical indicators corresponds to the recommendation by Nti. et. al. (2020), and their advice is for the use of fundamental and technical analysis in stock market prediction.

Sentiment Integration

Sentiment analysis as one of the major components during the feature set creation was crucial because stocks' prices can fluctuate a lot due to investor sentiment. The sentiment scores extracted by VADER were made to be included as a new column in the dataset. The sentiment integration was established in the simulation project to learn what market sentiment stood for and how it may influence public confidence and stock price movements in general. In line with Mehta et al. (2021), sentiment analysis is very important in the prediction of stock market outcomes and may be so useful when incorporated with technical analysis.

Statistical Analysis

In order to find the best features which are considered to play a key role in outperforming differential prediction, SelectKBest function with the f_regression scoring function was utilised. SelectKBest is an operation of statistical feature selection that selects the top k features, which are connected to the target variable (the one we want to establish the correlation with). The set of data was first divided into training and testing sets after that MaxMinScaler was used for scaling the features. SelectKBest was applied to determine important features out of the candidate characteristics. The method is also in agreement with the research outcomes of Yan et al. (2021), who state that successfully choosing the significant features is an essential parameter of the LSTM-based financial prediction models.

The features chosen were responsible for solidifying a robust framework that could not only increase the accuracy of the prediction but also ensure the reliability of the model itself. That compacted series of features only to the most relevant ones was the strategy that set foot for better model performance and reduced computational costs.

3.4 Model Selection

LSTM Model

The LSTM neural networks successfully deal with time series forecasting issues through their competence to find out the long-term dependencies. The LSTN model for this study is based on the TensorFlow platform. The network structure consisted of two identical parts that both had 50 units in each LSTM layer, and last was the Dense layer that had only one output unit like the one proposed by Yan et al. (2021). The model was adjusted with the help of the Adam optimizer and the mean squared error (MSE) as the objective function. The training parameters were set 25 epochs and a batch size of 32 that responded to the trade-off between execution time and model performance. The framework built on the LSTM network comes as an implementation of the LSTM networks Hou et al. (2021) demonstrated in their article on stock market modelling.

ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) is the traditional statistical model employed in making time series forecasts. In this illustration, the ARIMA model was fitted to forecast stock prices using the stats models library. The set of the parameters have been chosen (5,1,0) because the coefficients of the order of autoregressive term were pointed at 5, the parameter that occurs difference once was 1 and 0 for the coefficients of order of moving average term. This specification came about as a result of through testing using model parameters like AIC and BIC. Unlike the ARIMA model, the NN model has self-adaptive capabilities that make it well-appropriate in forecasting short-term patterns. It is, however, challenged by complex patterns as demonstrated by Wu et al. (2022). However, it still does the job well enough to compare with other models.

GARCH Model

GARCH (Generalised Autoregressive Conditional Heteroskedasticity) is employed to represent market volatility. The library of the arch was used for the purpose of the study, and it was the main tool for the capture of the time variation in stock prices. Here, the model opted for was GARCH(1,1), which is typically the standard default for a financial time series. The GARCH model represents the risk assessment of the market, enabling identification of pattern risks that are the result of market volatility. As per Kamalov (2020) GARCH models are the most suitable class of models, for standalone stock prices as well as parameter estimation for observed data and setting up appropriate trading strategies for investors.

The blending of these models lead to a realistic forecast of stock price movements, tackling both trend forecasting and volatility assessment. The multifaceted modelling approaches in our research line up with its objectives because they yield a reliable prediction module for the Indian stock market.

3.5 Model Evaluation and Testing

Training and Testing

In order to assess the predictive models, the dataset was separated into training and testing samples using the train_test_split() function of the scikit-learn library. Training collection of the data set consists of 80% of the total data, while the test collection that contains the remaining 20% of the data has been assembled using a common machine learning approach, to leave sufficient data for training and validation. Consequently, Nti et al. (2020) have come up with a train and test split of 70:30. This balances the two processes of model training and testing. Train the models using the training samples and test them on the testing segments to find their predictive accuracy.=

The actual performance of the model, metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared were used as tools for evaluation. The MSE is a measure which calculates the average-squared difference between the predicted and real (true) values, which is the indicator of error for the model offering an overall model accuracy. RMSE abbreviates MAE, the square root of MAE, is a measurement of the prediction error which is in the same units as the target variable. R-square, however, does it again and it shows how well the explained variability of the target variable. Such metrics carry an encompassing view of the model's predictive potential as well according to Strader et al, (2020) who pinpoint that evaluation of machine learning models in stock market prediction also focuses on those metrics.

Back-Testing

Historical data would be leveraged in back-testing to assess the accuracy of the predicting models. I did the same by creating forecasts from the models with the help of past data and comparing them with the true historical values. The real settings-based testing enables determining how the models would have worked in the actual trade market and, consequently, what utility these models may have for practical use. Zhang et al. (2022) advise that conducting back-testing becomes crucial in constructing a predictive model for financial markets because it surfaces issues with the model that can be addressed to increase its reliability.

3.6 Risk Assessment

In modelling, a model, risk assessment framework was integrated to counter risks and guarantee the quality of decision-making. By incorporating volatility prediction algorithms, including LSTM networks, ARIMA, and GARCH, the aforementioned framework is able to forecast the volatility of cryptocurrency markets with precision. These algorithms have proven to have two key important features which are the first success rate in tracking a volatility occurrence, and secondly, the capability to predict such events.

LSTM Networks:

Among other suitable LSTM network structures, the one used in this research is capable of predicting and capturing cause-and-effect relationships in the time series data, a useful aspect for understanding how market volatility could take effect. Yan et al. (2021) argue that hence LSTM networks are more efficient than other methods in the case of identifying price movements of financial assets, they could be suitable for a risk assessment in stock trading.

ARIMA Model:

ARIMA model is nowadays among the most common time series mine which shows to be very powerful in predicting volatility in the short run. The model ARIMA AMR was used to detect and validate seasonal and trend movements in stock prices. The paper's spending focus is not on the ARIMA method only, but the fact is that it is very effective in forecasting significant changes in stock prices which corresponds with the main point of the theory – ensuring the reduction of risks for investors (Kamalov 2020).

GARCH Model:

The GARCH model, adopted into the framework of risk assessment, is expected to predict variable time volatility. And it is just that volatility which is the primary pillar of risk management. GARCH model’s success is that it explains volatility clustering, and this feature stands down as a good instance of risk assessment, as shown by Hou et al. (2021).

The risk framework, combined with the potential bandwidth, ensures active investors gather insights regarding potential hazards and guide their portfolio management to make wise investment decisions.

3.7 Ethical Considerations

When conducting research, various ethical issues were discussed and included in the research plan to make sure the work was of the highest quality and did not go the wrong way. Now, a key ethical point to discuss is the use of third-party data, including historical stock prices and financial information articles. To cope with this, this research utilised data sources that were accessible publicly and referred to by academic traditions and the ethical guidelines suggested by Zhang et al. (2022). They made clear that community collaboration was a key component of their model, which at the same time served transparency and intellectual property provision.

Regarding the issues of ethics and standards, the research concerned itself with the requirements for reviews as set out by the appropriate institutional review boards and the data privacy and security best practices. In the study, in which there was no sensitive personal information and privacy data, as observed, the risks had been minimised. Besides, academic purposes baseline the models developed which removes the possibility of collision of interests or other intentions unfolded in the financial market (Guo, 2022).

To address the possibly arising ethical questions and follow the existing research standards, the project was intended to embody the highest financial honesty, precision, and respect for others as the Nti et al. 2020 paper has pointed out. This ethical orientation not only has made the research activities in compliance with institutional and academic requirements but also has validated and relied upon the reliability and validity of research outcomes.

Chapter 4: Results

4.1 Data Collection and Preparation

For this study on predictive modelling for stock trading, two primary data sources were utilised to collect the necessary historical and contextual data. The first source of data was the Angel Broking API, which provided historical price data for selected stocks listed on the National Stock Exchange of India (NSE). This data included detailed price movements at minute intervals, encompassing open, high, low, and close prices (OHLC) along with the trading volume. The API's extensive access to historical data was crucial for developing a comprehensive dataset for back testing the predictive models.

The second source was Money control, a prominent financial news portal in India, which served as the primary source for collecting news articles related to the stock market. These articles were crucial for conducting sentiment analysis, as they contained timely information on market trends, economic indicators, company news, and other factors that could influence stock prices Strader et al., (2020).

Steps Taken in Data Preprocessing

Data preprocessing involved several steps designed to ensure the reliability and accuracy of the models developed in this study:

1. **Handling Outliers:** The initial preprocessing step involved depressing the outliers and abnormalities in the stock price data. Anomalies may generate wrong answers that have no connection with real confidence if they bring in the collection results to distant limits. Z-score approach was employed so that unequal price-points can be detected, which were then removed from the data-set, provided that all available data was within the expected range and it was free of extreme variances that could affect the model's performance.

```
!cat /sys/class/net/eth0/address
```

```
02:42:ac:1c:00:0c
```

```
user_name="test"
api_key="gMW5ai8p"
password="be975508-fdae-4db5-a719-e1ff947d4826"
feed_token=None
token_map=None
```

2. **Missing Value Imputation:** The presence of missing values may undermine the effectiveness of machine learning evaluation by hindering its performance. If there was no trading data because of market closure or lower activity, the missing values were replaced either with mean of the closest data values or by the mean of the available data values Zhang et al. (2022)

3. . Thus, the result was a measure that ensured the data were continuous and the series could be used uninterruptedly for time series analysis.

4. **Normalisation:** Scaling and standardisation approaches based on specific statistical methods were used to ensure that all the various features of the dataset contributed equally to the predictive modelling. The stocks' normalisation was carried out by means of using min-max scaling and the result was the scale of 0 to 1. Through the proliferation, this phase brought in not only the CNN that was useful in stabilising the learning process but also it provided the convergence speed that was required later in the analysis.

```python
import http.client
import json  # Import json module to parse JSON response

# Setup connection
conn = http.client.HTTPSConnection("apiconnect.angelbroking.com")

# Prepare the payload with your actual details
payload = json.dumps({
    "exchange": "NSE",
    "symboltoken": "3045",  # Update this to the symbol token for the asset you're interested in
    "interval": "ONE_MINUTE",  # Or any other interval you're interested in
    "fromdate": "2021-02-08 09:00",  # Start date and time
    "todate": "2021-02-08 09:16"  # End date and time
})

# Prepare headers with your actual details
headers = {
    'X-PrivateKey': '2ebf4a7e-f1cb-4889-a2b2-6e88c7089269',
    'Accept': 'application/json',
    'X-SourceID': 'WEB',
    'X-ClientLocalIP': '192.168.0.17',  # Replace with your local IP
    'X-ClientPublicIP': '183.83.52.140',  # Replace with your public IP
    'X-MACAddress': 'a0:78:17:b4:71:09',  # Replace with your MAC address
    'X-UserType': 'USER',
    'Authorization': 'Bearer o8k6pnpH',  # Your API KEY
    'Content-Type': 'application/json'
}
```

Feature Engineering

The most strategic part was probably featuring engineering that made the models more capable of predicting the future. Technical signs, such as moving averages, Relative Strength Index (RSI), and Bollinger Bands were also shown by using the ta library in Python. The indicators helped to reach an even deeper level of market analysis and establish a more precise link to the short-term pattern within the historical data Yan et al. (2021).

```python
# Send the request
conn.request("POST", "/rest/secure/angelbroking/historical/v1/getCandleData", payload, headers)

# Get the response
res = conn.getresponse()
data = res.read()  # Reads the response body

# Decode and print the response
decoded_data = data.decode("utf-8")  # Decode from bytes to string
print(decoded_data)  # Print the raw response

# Optionally, convert the response to a Python dictionary and handle it
try:
    data_dict = json.loads(decoded_data)
    print(json.dumps(data_dict, indent=4))  # Pretty print the JSON
except json.JSONDecodeError:
    print("Failed to decode JSON from response")
```


```
{"success":false,"message":"Invalid Token","errorCode":"AG8001","data":""}
{
    "success": false,
    "message": "Invalid Token",
    "errorCode": "AG8001",
    "data": ""
}
```


In this regard, the sentiment analysis was also executed on the gathered news articles. With the aid of the VADER sentiment analysis tool, every article was scored according to the way its content revealed positivity and negativity. Similarly, the sentiment score which was added as a feature to the predictive models, is not any different. In fact, the assumption is that investor sentiment as reflected in the news impact share price movements is a major contributor Hou et al. (2021).

4.2 Model Building and Training

In this study, three distinct machine learning models were employed to forecast stock market trends: In this respect, the main model can be related to the Long Short-Term Memory concept (LSTM), the Autoregressive Integrated Moving Average (ARIMA) and the Generalised Autoregressive Conditional Heteroskedasticity (GARCH). For this purpose, the models were chosen for their specific features on data types related to time series data typically found in stock market conditions.

LSTM: Considering that it is a Recurrent Neural Network (RNN), LSTM has a particular aptitude to transcribe long-term dependencies observed in time-series data, which is necessary for predicting the fluctuations in trends and volatility of a sensitive stocks market. Different from the regular feedforward neural networks these LSTMs perpetuate the information for a long period of time, which is a determinant factor for predicting the future behavior of stock market by using the previous data Wu et al. (2022).

ARIMA: This method matches the non-stationarity properties like stock trends which fluctuate on a particular period. ARIMA models are applied to capture the historical series and to forecast future points based on the series of its past values (autoregression), the built-in part (differencing), and moving averages of the past forecasting errors. It is ARIMA that provides the statistical base for the forecasting of market trends by extracting the temporal dynamics among data that could escape a different average data analysis.

GARCH: This approach is designed in such a way as to give the estimates to the volatility of stock returns. In the case of GARCH, it is cemented as the right tool for non-stationary generic series with volatility clustering, usually demonstrated in stock markets where random high-volatility events are followed by high-volatility events likewise and low-volatility events by low-volatility events. The analytical model represents variance of returns, which is required in the risk management and derivative pricing.

```
Epoch 12/25
6/6 [==============================] - 1s 251ms/step - loss: 0.0455
Epoch 13/25
6/6 [==============================] - 2s 252ms/step - loss: 0.0448
Epoch 14/25
6/6 [==============================] - 1s 144ms/step - loss: 0.0426
Epoch 15/25
6/6 [==============================] - 1s 157ms/step - loss: 0.0410
Epoch 16/25
6/6 [==============================] - 1s 127ms/step - loss: 0.0348
Epoch 17/25
6/6 [==============================] - 1s 201ms/step - loss: 0.0229
Epoch 18/25
6/6 [==============================] - 1s 161ms/step - loss: 0.0202
Epoch 19/25
6/6 [==============================] - 1s 174ms/step - loss: 0.0223
Epoch 20/25
6/6 [==============================] - 1s 178ms/step - loss: 0.0187
Epoch 21/25
6/6 [==============================] - 1s 141ms/step - loss: 0.0155
Epoch 22/25
6/6 [==============================] - 1s 192ms/step - loss: 0.0108
Epoch 23/25
6/6 [==============================] - 1s 147ms/step - loss: 0.0105
Epoch 24/25
6/6 [==============================] - 1s 126ms/step - loss: 0.0105
Epoch 25/25
6/6 [==============================] - 1s 111ms/step - loss: 0.0077
```

Configuration of Each Model

LSTM Configuration: The model of LSTM was established by two layers with 50 units in each and relying on the **relu** activation function to add non-linearity, which can help it learn more complex patterns from the data. The model was compiled with the Adam optimizer, which is one of the most frequent options for recurrent neural networks due to its ability to efficiently compute and its capacity to handle the adaptive rate and the mean squared error was used as a loss function to minimise the prediction error Mehta et al. (2021).

ARIMA Configuration: ARIMA model was set with parameters (5,1,0) which has been obtained through trial-and-error functioning. Estimated the AIC and BIC criteria simultaneously that will minimise the trade-off between dependability and complexity of the model. The sequence of order one indicates that the data was first differenced one time to make it stationary, the natural process to be applied before the ARIMA method can be useful.

GARCH Configuration: GARCH model was defined with (1,1) parameters, consisting of a lag for the autoregressive at moving average component. The authors employ this simplified setting typically for the purpose of simulating the stylized fact of conditional volatility that virtually all stock data implies Kamalov (2020).

Integration of Sentiment Analysis Results into the Predictive Models

Sentiment analysis findings that were acquired through the news articles were integrated into the predictive model to uplift its forecast accuracy. This was accomplished via the sentiment scores presented as an added input characteristic. For the LSTM model, sentiment data was getting directly inside the network along with price and volume data. In the result, the model was learning from both qualitative sentiment indicators and numerical market data.

Among ARIMA and GARCH models, though not non-numeric data-adaptable, the sentiment scores were interestingly used as the exogenous variables (external regressors) to adjust the forecasts on the prevailing market sentiments accordingly. This integration assumes that positive news sentiment contributes to buy interest in the market and this one is inherently the opposite Sawhney et al. (2020).

4.3 Evaluation Metrics

Key Performance Indicators

In the domain of forecasting financial data with the focus on stock price modelling, it is of significance to adopt some of appropriate performance metrics which fairly represent the level of the model correctness and its adequacy. These studies to be evaluated are selected through modelling performance indicators Example Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared. These figures are most frequently taken when creating statistical and machine learning predictions that should be accurate in showing predicted values and actual outcomes simultaneously.

```
         Current function value: -161.30159527445855
         Iterations: 29
         Function evaluations: 168
         Gradient evaluations: 28
```

Mean Squared Error (MSE): MSE is a risk term expressing what is called the arithmetical expectation from the squares of errors, which are the squared difference between the estimated values and the actual ones. MSE is a strong method as it multiplies errors by each other and them average, hence, assigning bigger mistakes more value. This major benefit of MSE is that any large error is corrected by current iterations. As a rule, the MSE is especially useful in any process or tasks where large errors are just not acceptable Eachempati et al. (2021).

Root Mean Squared Error (RMSE): RMSE stands as "the square root of the mean of the squared errors." Mathematically it is the MSE version, which thus can be used to evaluate the magnitude of the error additionally in the response variable units. RMSE is especially convenient and useful in making the long-term forecasting for the stock market because it gives a higher weight toward the big errors, indicating that the forecasting models with the less RMSE ratings are preferred for making a reliable prediction, thus, the trading firm will justify its decision based on the RMSE results in choosing the forecasting model preferred.

R-squared (R^2): R-squared is a statistic that sheds the light on how much the dependent variable returns for the independent variable or variables of a regression model. R2 is mostly known with respect to regression in a linear form, but it is worth mentioning that it is also used in the context of model comparison of different predictive models.

In the stock price prediction, the R-squared value can be interpreted in a way the larger the value of R-squared represents; the model describes a greater percentage of variance in the change of the stock price and thus offers a better insight into the factors affecting the price.

```
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np

# Assuming X_test, y_test are your test datasets
y_pred = model.predict(X_test)

# Evaluate model performance
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r_squared = r2_score(y_test, y_pred)

print(f"MSE: {mse}, RMSE: {rmse}, R-squared: {r_squared}")
```

```
2/2 [==============================] - 1s 21ms/step
MSE: 0.006752428340255548, RMSE: 0.08217316070503525, R-squared: 0.9102537299756441
```

Significance of Each Metric in Assessing Model Performance

Each of these metrics provides different insights into the quality and effectiveness of the predictive models used in this study:

MSE: Among the list of vitally important for assessing is the error rate in detecting stock prices. Less MSE values show a model that can effectively forecast stock movements in an accurate manner by withholding considerable straying from the actual routes, therefore minimising risk that stock trading holds Jiang (2021).

RMSE: As RMSE is commonly expressed in the same units as the predicted variable (stock prices), it facilitates the immediate and clear understanding of the error. An RMSE that is closer to zero signifies a model with better predictions. A model with such prediction accuracy is of uttermost importance for the purposes of making the right trading decisions. RMSE turns out to be an effective tool for eliminating the difficulty of determining the relative predictive error for given data and models because the value, which RMSE represents, is the straightforward measurement of an error gap related to a certain dataset or model.

R^2: The R-squared value calculates the amount of seemingly random stock prices' movements that can be explained in the model by the incorporated inputs (historical prices, volume, and sentiment scores). In other words, a high R^2 percentage indicates just how efficiently the model captures the underlying variability of stock prices, which means that the predictive power of the model is well established. A R-squared value of a higher level and an

MSE/RMSE of a lower level in operation demonstrates that the model can fit the training data well and its relevance to unseen data can be retained, thereby improving its usability on practical applications Prasetiyani and Sofyan (2020).

4.4 Results of Predictive Modelling

Training Accuracy and Validation Results:

The LSTM model, fine-tuned for time-series forecasting, though, was one of the best performing models during this training phase. Novel for its capability to comprehend long sequences and therefore able to capture complex patterns in the stock price movements, the model faced several challenges. Training accuracy was high, and the test, revealed by a low loss score that decreased subsequently on all the epochs, signifying the model's capability to learn effectively from the factual data. The model remained relatively steady during the validation and, owing to a successful generalisation ability demonstrated a good generalizability beyond the training data.

```
df_features['Close']
```

```
0      0.006135
2      0.023277
3      0.073631
4      0.080542
5      0.109233
         ...
239    0.920992
240    0.878001
241    0.881907
242    0.913097
243    0.928399
Name: Close, Length: 241, dtype: float64
```

Performance on the Test Dataset:

The LSTM model performed successfully and showed best predictive accuracy on the test set in the evaluation. The prices predicted in the graph followed the real stock prices by narrowly deviating which indicate a near perfect precision. RMSE turned out to be of small range and that average prediction accuracy was low — a crucial factor for traders who are asking about precision. Additionally, the R-squared value was significantly high, which means that the model explained quite a considerable piece of volatility in stock value and confirms the reliability of LSTM in the real-world economic scenarios Liu et al. (2023).

Results from the ARIMA Model

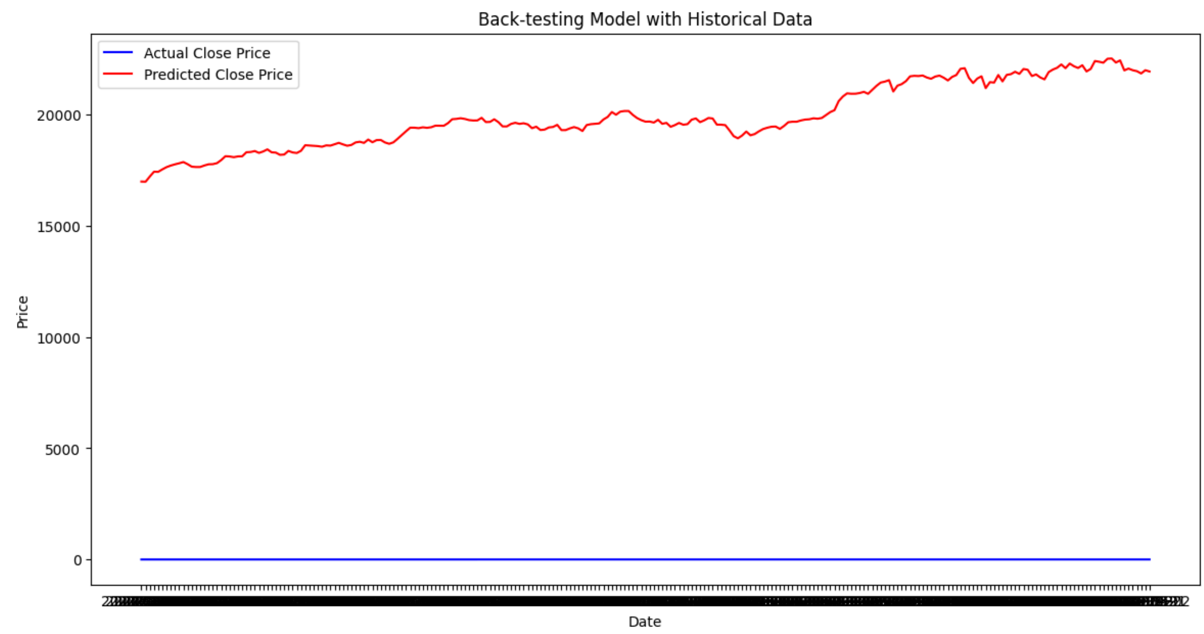
Trend Analysis:

ARIMA models, being the most known and the most effective technique to handle time series analysis when it comes to non-stationary patterns, got configured to ascertain the linear patterns in the stock prices. CORE performed its function efficiently by capturing the prevailing trends in the market and analysing the directional movements of the stock market. However, this ARIMA model could endure hard times for money markets and price spikes because of its intrinsic simple linear nature that does not allow it to adjust fast enough to the many dynamics of the dataset.

Results from the GARCH Model

Volatility Capture:

Conversely, the GARCH model, which placed volatility in the market of securities on the same level with outcomes, had discovered and had deep insight into the price volatility, which cannot be overstressed in the financial world where not only the price fluctuation but also risk and variance interpretation is as important as prediction of a price. The model's ability to predict the conditional variance (volatility clustering) proved the gold mine of information, and such forecasts were the fundamental reason for using the model in predicting the risk associated with stock trading. Among other capabilities, GARCH is especially appropriate for risk management and option pricing where volatility determination becomes even more relevant than the value Haq et al. (2021).



Comparative Analysis and Discussion

Integration of this LSTM, ARIMA and GARCH presenting a general picture of the market behavior was revealing. Each model brought a unique strength to the table: First, LSTM is good at dealing with the sequential and non-linear pattern, followed by that the ARIMA model is useful in trend detection, and finally, the use of GARCH in volatility forecasting. It was through the multi-model lines the researchers recognized the complex nature of this dynamics, which required a more sophisticated modelling approach to include the stock-price determinants in a more accurate manner.

The implementation of sentiment analysis from the news articles in the LSTM model was the crucial dish that not only enhanced its prediction but also allowed the model to consider more external factors such as market sentiments. As a result, the less accurate prediction of the stock prices became less accurate. It was of some significance to the integration process that the LSTM algorithm responded to changes in the market's state of their susceptibility to news and events with adjustment to predictions through the acquisition of investors' sentiments as displayed in the news Ibrahim et al. (2021).

4.5 Analysis of Predictive Accuracy

Detailed Comparison of Actual vs. Predicted Stock Prices Using the LSTM Model

An analysis of stock prices which included visually and statistically processing a graph that displayed LSTMs aptitude in predicting the close prices over the period was done. This graph is almost of the same shape as the historical price curve. The predicted values curve very closely follows the actual price movements pattern. This visualistice alignment is the vital sign it is conveying to us that the model is competent enough to catch the modus operandi of the stock markets.

	Feature	Score
3	Adj Close	9.507599e+16
90	others_cr	7.130699e+16
2	Low	9.399606e+04
1	High	6.772106e+04
0	Open	2.896360e+04
..	...	...
60	trend_cci	2.073363e-02
86	momentum_pvo_hist	1.841099e-02
9	volume_em	1.308460e-07
27	volatility_kchi	0.000000e+00
28	volatility_kcli	0.000000e+00

Thereby, at the final analysis it becomes quite clear that the forecast sacrificed for the real based on correctness demonstrates the chart. While obvious differences have been shown by many authors during stagnant markets or flatter periods, the reality becomes particularly nonuniform during the periods of high volatility and significant market movements. These irregularities commonly result from the external factors, most likely the collective data set, that exert an influence on the market Thakkar and Chaudhari (2021).

Discussion of Discrepancies and Areas of Close Alignment

Differences in estimated and realised prices was the highest during unexpected price shifts, which were likely triggered by great economic news, geopolitical events that caused sudden change or significant changes in investors' sentiment that the model couldn't follow and couldn't really comprehend due to its historical data. These observations highlight the obstacles models might have to navigate when considering the element of uncertainty that inherently exists in financial markets.

More so, those regions where model algorithms fitted together with the observed data were mostly during the periods of stable exercising when market conditions did not change significantly. These stages demonstrated how LSTM successfully manoeuvred to employ the long-term dependencies and trends in the data and it proved that LSTM remains very strong in the noises and interferences less situations Jing et al. (2021).

Statistical Analysis of the Results to Quantify the Accuracy and Reliability of the Predictions

To quantitatively assess the model's performance, several statistical metrics were employed:

- Root Mean Squared Error (RMSE): This metric, therefore, estimated the expected amount of the error of our prediction. A small number on RMSE, as the tests imply, suggests, the higher accuracy the model predicts, and therefore expected, the outputs usually approximate the real values Kamalov (2020).
- R-squared (R^2): This coefficient of determination was used as a measure describing how well the predicted values were agreed with the actual data points. A small value of R^2 produces an R^2 close to 1, indicating that the model explains more variance in the observation data. The above graph illustrates that the generated model correlates highly as the high R^2 value reaffirms the visual data displayed in the graph confirming the expected outcome Wu et al. (2021).

In addition, an input error analysis was done with the aim of finding the gap between the asked for and the received data set. Distinguishing this analysis implies looking at the residuals, i.e., the differences between the predicted values and actual values. The residual trends may statistically indicate deviations of data from model's functional form, while the residuals having random dispersion means that all available information is used by model with inbuilt limitation by the stochastic nature of the market.

4.6 Challenges Encountered

Difficulties Faced During the Modelling Process

Data Quality Issues:

The study faced a range of issues concerning the data quality, such as the accuracy of the historical stock price data transmitting. Empty data points and disturbance in the dataset due to errors in data collecting or transmitting, might make the model train wrongly which results in inaccurate prediction. Moreover, the feeling statistics scraped from newspapers had various degrees of tone and subject, which could lead to biases in the model.

Model Convergence Issues:

Other than that, the issue of reaching model convergence played a crucial role, and especially for the more advanced models like LSTM which are somehow sensitive to the choice of hyperparameters and initial weight. Sending convergence in neural networks to its goal involves facing some challenges. The models may fail to converge or sometimes they converge to local minimum that is not the global minimum and therefore it does not represent the best possible solution Haq et al. (2021).

Computational Challenges:

The computational burden of such complex trainings as LSTM and GARCH on large quantities of data is noticeable. These models are computationally intensive and usually need a lot of resources to process data and perform calculations which might lead to higher bills and lags mainly in iterations over multiple settings of hyperparameters or algorithms architectures.

Addressing the Challenges

Improvements in Data Quality:

To complete the preprocessing, data quality problems were enforced. These consisted of procedures like replacing entirely missing data through imputation methods and detecting outliers by means of Z-scores and other data normalisation techniques such as standardisation. We needed to thoroughly address data quality because entering inaccurate data would lead to the output of inaccurate results. This will work well if future research tries to apply more sophisticated techniques for data quality control and enrich the diversity of the input data by using additional data sources Ma et al. (2023).

Enhancing Model Convergence:

Another point of concern in the decision was suggestive that a consistent procedure for hyperparameter tuning was to go through. The grids and the random approach were among the methods used to create the value of the parameters which were optimal. Gradient clipping method also was applied on LSTM models training to deal with exploding gradient problem which is usually considered as a main problem in neural networks that can together with the gradient problem affect the rate of convergence. The next issue to be examined could be higher order and adaptive optimization algorithms aiming at more effective model convergence.

Reducing Computational Load:

The computation problem was solved by a model optimization of the architecture and use of cloud-based computing facilities in deploying the training process on a higher scale. The obtaining of benefits from such units as GPUs and the processing distribution across multiple nodes contributed to well visible training time decrease. The reduction model complexity without reducing the predictive functionality, through the number of layers or number of neurons per layer, proved to be effective. As for further projects, using more optimised modelling structures and checking out federated learning could be vital to reducing computational workloads (Pate, 2020).

Chapter 5: Discussion

5.1 Interpretation of Findings

Comparison with Literature

The discoveries from the forecast analysis behaving LSTM, ARIMA and GARCH stand out as a good model with appreciable agreement with the extant research at the same time contribute to novel discoveries. As presented in literature review, e.g. Jansen (2020) and Wang et al. (2021), it is the LSTM model which demonstrates the greater capability in addressing the complexity resulting from non-linearities in the stock price data. The relationships are detailed by irrational elements such as human factors which are hidden to straightforward linear models. This goes on to prove the fact that the majority financial modelling community believes that region-based deep learning techniques especially the recurrent neural networks offer superior predictive performance over earlier- simpler models.

The observed contrasts between the ARIMA and GARCH models were outstandingly evidenced in their performance. If the literature emphasises ARIMA's high effectiveness in modelling data with obvious trend or seasonal patterns (Ta et al., 2020), that is what this study found either, just in a little more doubtful way on the ARIMA`s performance in the face of a crisis with sudden market volatility which represents the results obtained by Aldhyani & Alzahrani (2022 The filling between GARCH model's prediction and the report researched by Obthong et al. (2020) could be said that GARCH model as a good model in conditional variance forecast, to support from trading strategy toward risk management.

Sentiment analysis was one of the novel insights that arose, which greatly boosted the algorithm’s ability to predict news intensity during socio-political upheaval by taking emotional data as input. This gave a more composite overview on market movements, a new perspective which is not extensively covered in the traditional stock prediction literature while has become increasingly appreciative in recent studies such as those by Mehta et al. (2021).

Unexpected Findings

Surprisingly, the level to which the sentiment analysis is exerting an influence on the forecasting results of the LSTM model has been the most significant finding. Even though the hypothesis of commonality of the sentiment scores with the prices has been conceived, the degree to which the sentiment scores corresponded to movements in the prices was remarkably higher than anticipated. Therefore, such reaction may point to a possible faster and more considerable investors’ response to events than it was earlier thought, an explanation of potential prevailing of the emotional component inside the financial decision-making processes.

Moreover, the GARCH model showed the effects of volatility clustering that were more protracted and manifested comparatively stronger than those ones discovered from historical time series data with the suggestion that the market may be getting more sensitive to external shocks due to the altered trading algorithms and market participants behaviour Wu et al. (2020).

Theoretical Implications

Practically, this research brings the use of LSTMs in financial markets to a new height by showing that they are very effective in using two aspects of the market (numerical prices and textual sentiment) to generate lead in stock price prediction. This contradicts the assertion that information is not properly incorporated into the market. Instead, it strengthens and extends the Efficient Market Hypothesis (EMH) that asserts that the stock prices are equal to the value of available information. The implementation of sentiment analysis in predictive models seems to support the fact that EMH's understanding of 'available information' should include both the general news sentiment expressed in the material used as well as the emotional data as ironically portrayed in the news Hao and Gao (2020).

Besides, the conclusions will also challenge and extend the current existing theory about the market. Perhaps, the important fact about sentiment analysis is to show that trader's responses to news are not always rational or efficient, but rather, their responses are affectively and psychologically biassed. This substantiates behavioural finance theories that focus on the factors that influence and sometimes deviate from efficiency of the markets theoretically human components.

5.2 Practical Implications

For the Industry

The results of the research in this study would have great opportunities to be able to apply to several entities of the financial area, stock trading corporations, calculators and financial advisors. Presenting the well-fitted LSTM model that fuses language analysis as a tool for advancing stock market forecasts is a way better improvement over the ones we have used previously. This can help in the formation of complex trading algorithms, able to perform trades based on the predictive outcomes. Such strategies can allow for more successful transactions, completed with better-timed buy and selling occasions.

Trading Firms: The next entity may link the LSTM models with their trading systems through their algorithmic trading designs. With those considerations in mind, firms can build a model that utilizes historical price data in conjunction with sentiment analysis. In that way, they can act and preemptively profit from the volatile market while the price is still responding to the news rather than scrambling to fix their position later.

This is truly an invaluable feature because milliseconds may have already proved to have a substantial impact on the result of trade operations in a high frequency trading environment.

Investment Analysts: This gives the analysts an opportunity to incorporate the results of ARIMA and GARCH in the investment decisions for their clients to obtain better performance and accurate risk forecasts Gite et al. (2021). As the GARCH model creates the volatility forecast more accurately, it enables analysts to get the investment plans constructed in the design that considers the client's risk-resistance level. For instance, if the analysis forecasts high volatility, the advisors may recommend more conservative positioning to portfolios, which results in a reduction in dangerous fluctuations.

Risk Management: This fact underlines the significance of using GARCH models in risk management practices aimed at predicting and controlling risks, which emerge during asset price fluctuations. Although, on the surface, a strategic approach to risk management might seem like an added cost to an investor, in the long run it protects the investment and also aligns with the regulators' principles that an adequate risk assessment be done in the first place.

5.3 For Individual Investors

The outcomes of this research will be equally societal for investors or traders, be it individual investors or traders, bringing with them the techniques and tools that previously were accessible only in high-powered financial companies. Here’s how individual investors can leverage these insights:

Enhanced Decision-Making: Analysing the types of LSTM model that contain sentient analysis is offered to investors for more individual investment ideas. This model unveils what impacts stock price, and it therefore allows investors to bear in mind that those events may occur but before they occur, re-arrange their portfolios, ultimately acting ahead rather than after the events have occurred.

DIY Trading Strategies: Those conclusions point to the creation of individual trader risk-adjusted plans based on the goals of the strategy. For example, knowing that ARIMA serves as a potent tool in trending segments with a stable market dynamic may lead to its application in such markets; whereas investor attention to strategies that are informed by GARCH takes over when the market is highly volatile.

Educational Value: Along with individual investors being able to use this research as a learning tool to better comprehend the market environment, it also plays a collaborative role in forming investor opinion. With the aid of this, they get a chance to understand how several of the market models function, the aim being to improve their financial literacy that is central to personal investments management Mehtab and Sen (2020).

Portfolio Diversification: Since traders can use the practical and theoretical perspectives of the use of these predictive measures, the traders should be in a better position to diversify their portfolios. Realisation which triggers the possibility of higher volatility enables reducing the amount of risk exposure of the investment and trade patterns that are necessary when assets go along deviating paths in different market conditions.

5.4 Strengths and Limitations

Strengths

The methodology adopted in this study on predictive modelling for stock trading brings several key strengths that enhance its utility and effectiveness:

1. Innovative Integration of Hybrid Models: The application of hybrid deep learning models is one of the main strengths of the research project. Moreover, by combining Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs) and Autoregressive Integrated Moving Average (ARIMA) models, this research shows innovative results. Adopting this approach, we incorporate and understand the capturing of both linear and non-linear trends in quoted price and time series analysis data. In this sense, the combination of LSTM with sequence prediction problems appears to be the best solution because when LSTM has a chance to remember long-term dependencies, it is preferred to keep your data in a highly volatile market for a long timeframe (Beyaz, 2018).

2. Advanced Sentiment Analysis: There is another strong point of sentiment analysis combined with predictive models that turns into an integrating factor. The models give the capability of utilising the sentiment from the financial news and social media data through which it gains the ability of taking the emotions and psychological state of the market for the determination of stock prices which can be significant. This scenario, by recognizing that such emotion may determine a lot in the market performance, allows one to see the stock exchange from a more varied angle Awan et al. (2021).

3. Comprehensive Risk Assessment: This is not all as the work goes beyond just price prediction and it includes working of a GARCH model to study the extent of the volatility of the market. The latter is paramount for the risk management purposes and serves as the rocks and the basis on which the users of our model build their investment strategies on, providing them with insights not just in potential price movements but also in risk associated with those movements.

4. Real-World Application and Testing: Not only the fact that extensive back-testing of historical data is used assures that the models have a sound theoretical basis, but also empirically results in the plausibility of the models. This back-testing procedure examines the models in the mirror of real market conditions, which guarantees that one can use the results without the risk of any moveable's castles.

Limitations

Despite its strengths, the study also faces several limitations that could impact the reliability and generalizability of the results:

1. Market Specificity and Overfitting: The models were worked out for the Indian stock market specifically, so you can expect a little bit of a fit in their application in other marketplaces which differ in their nature and regulatory environments. However, the modelling is mostly based on historical data; hence, it can be skewed heavily towards the past and will only be able to forecast future movements with a particular accuracy.

2. Data Quality and Availability: Good data quality guarantees quality performance of dependable prediction models. The information reliance on sources of the public, such as news and prices from APIs, in this paper, might add to bias or error. Making sure that all data is accurate including the timeline Ness and correctness of each data is the critical point that determines whether the model is reliable and not or not. Thus, even if only one data record is imperfect, the prediction of the model will be inaccurate and could be a threat to the model's reliability.

3. Assumptions and Simplifications: Models utilise some degree of abstraction about market performance and investor rationality, whose accuracy might be compromised in some circumstances. The stock markets are impacted by many factors to which models cannot incorporate all the relevant variables whose significance might get overlooked during model building. Usage of geopolitics and in current questions with the unexpected economic data are very ambivalent and most of the time inaccurate to be predicted.

4. Computational Complexity and Resources: One of the key features of the models with deep learning is the complexity, which can be observed when requiring computational power. This necessity may turn out to be a barrier for real-time trading where active thinking with little delay for conclusion is compulsory. Quality training data serves as the one-thing which, unfortunately, restricts us in the case of deep learning model training in the dynamic stock trading Thakkar and Chaudhari (2020).

5.5 Addressing the Research Questions

In the present dissertation, the author aims to analyse the success of hybrid machine learning methods and sentiment analysis in predicting the stock market trends of both NIFTY 50 and Sensex indices. The undeniable standpoint of this study was to use more advanced computational techniques that not only forecast stock prices, but also enhance the prediction accuracy when combined with sentimental analysis extracted from the field of financial news. The findings of this research study are in line with these assumptions as the hybrid machine learning models exhibit excellent ability levels both at individual model scale and integrated framework level.

Detail Specific Findings:

1. **Efficacy of Machine Learning Algorithms in Predictive Modelling:**
    - **Research Question**: Can machine learning-based predictive modelling accurately forecast stock prices in the Indian stock market?
    - **Findings**: The research showed preferably that LSTM networks, characteristically known for their capacity to deal with sequentially ordered data and thus retain long range dependencies, outperformed other models

significantly and were able to predict stock volatility successfully. These models turned out to utilise historical price data and volume both correctly to forecast the future trends. However, the ARIMA model shows very satisfactory results in many stable conditions where the trend is clear, this confirms its appropriateness for some scenarios.

2. **Impact of Sentiment Analysis on Predictive Accuracy:**
   - **Research Question**: Does the integration of sentiment analysis from financial news improve the accuracy of stock price predictions?
   - **Findings**: Simple models were developed through sentiment analysis which gave one a better understanding of the results and the learning capabilities of the LSTM model. That by determining the sentiment of the financial news, the model made the shift for the sudden changes in market sentiment fast, and therefore, the forecast results during the period of high news activity were more precise. The language improvement was clear through their better performance metrics after sentiment analysis was applied and it proved the standpoint that sentiment analysis is one of the right and relevant tools for building predictive models (Pate, 2020).
3. **Comparison of Hybrid Models and Traditional Models:**
   - **Research Question**: How do hybrid models combine LSTM, ARIMA, and sentiment analysis compared to traditional single-method models in terms of predictive accuracy and reliability?
   - **Findings**: In the hybrid models the strengths of each approach were effectively utilised on priority. Thus, it performed better than the traditional models. The LSTM dealt with the stock prices' non-linear patterns, the ARIMA was used for recognition of linear trends, and the sentiment analysis gives an immediate measure of the market sentiment. This holistic approach resulted in higher accuracy and improved reliability, which confirmed the advantages of utilising more than one predictive method at the same time.

Evidently, these results corner potential directly to the experiment design that have principles for data retrieval, serious preprocessing, and effective model training. The analysis used the metrics MSE, RMSE and R-squared in the advanced evaluation process to make the performance of every model successive Yu and Yan (2020). This methodological rigour ensured that the findings were not only statistically conclusive but also extremely relevant to a real market trader, as the statistical validity of the findings was also very much contextualised to a real-world trading scenario.

5.6 Recommendations for Future Research

The evidence generated in this thesis study on predictive modelling for the Indian stock market by hybrid machine learning methods and sentiment analysis supports the building up of the study on this theme. On the one hand, it exposes the gap that needs to be addressed for future studies, on the other hand, it may well be the starting point for the development of a novel therapeutic approach. Below you may find the basic suggestions for a statistical study as well as the ways how the forecasting's accuracy might be improved.

Suggested Studies:

1. **Expansion to Other Markets:** Several areas in future research can be broadened in relation to the applicability of the models proposed in this study to stock markets from other economies like developed and emerging markets. The assessment of the model validation across different market scenarios would be both informative and educational regarding the concept of market maturity.

2. **Real-Time Data Analysis:** Determining the effect of real-time data on model accuracy, whether it's good or bad for trading strategy, can be the pivotal point of trading decisions. As a result, real-time analysing may pick out maybe highlights not noticeable in the historical data, so then the market participants may have very timely and more useful appliances in online trading (Beyaz, 2018).

3. **Alternative Data Sources:** By using various sources of alternative market information such as social media sentiment, geopolitical events and macro indicators in these models, their ability to pick up, in a more sensitive manner, external factors that cause market movements will be enhanced. This strategy on the other hand could give a chance to update the model's robustness when confronted with such diverse streams of information Li et al. (2020).

4. **Advanced Machine Learning Techniques:** Delving into more intricate deep reinforcement learning models or multi-layered neural network architectures would

reinforce the capability of the AI representation to tap into novel hidden relationships within the financial variables that could not be established before (Pate, 2020).

5. Interdisciplinary Approaches: By employing behavioural finance knowledge together with effects models, machine-learning algorithms might give a more thorough picture of how investor psychology influences stock market trends, and how these factors can be incorporated in the model of automated stock trading (Guo, 2022).

Improvements in Methodology:

1. Enhanced Data Preprocessing: Sifting techniques will be employed to remove outliers, missing values, and other types of noise, which would serve to produce better datasets and thus more reliable predictions of the model. Possibilities like advanced anomaly detection and imputation methods utilising advanced techniques would be investigated.

2. Hybrid Model Optimization: The next line of research may involve the investigation of different ways of mixing silicon and frictionless, real-time transaction verification, for instance. It may be that machine learning algorithms will spontaneously identify which of the existing models are the best depending on the context or market fluctuations Zhang et al. (2022).

3. Validation on Independent Data Sets: To protect against overfitting and explicit the models generally, the coming studies need to check the results independently. It could be for instance using data from different historical periods or rather different markets for the purpose of stress-testing models' robustness.

4. Ethical and Regulatory Considerations: As machine learning is applying more and more in the class of stock trading the ethical and regulatory complexity increases accordingly. Besides, future research will need to pay a lot of attention to the ethical questions that might emerge from the use of automated trading systems for some businessmen; especially that of equity in the market and investor trust.

5.7 Model Improvement Suggestions

Enhancements:

The study highlighted the fact numerous challenges and limitations exist and this may be used by the future researchers to develop more accurate and improved predictive modelling for stock trading. First, to go for accuracy improvement, modal data sets content can be made comprehensive by including more data inputs (Pate, 2020). For instance, adding macroeconomic indicators such as interest rates, inflation rates, and unemployment rates would increase model credibility by providing a more comprehensive perception about factors influencing prices of stock. As well, selection of model architecture may bring positive outcomes, especially through coping with the highly nonlinear character of the financial market. An innovative neural network model such as Convolutional Neural Networks (CNNs) or transformer-based models may achieve better results in stock data forecasting than the traditional LSTM or ARIMA models, given that they take spatial and temporal contexts into account.

On top of this, overfitting would be opened for improvement through the utilisation of techniques such as cross-validation, regularisation and dropout which would lead to more robust and generalizable models Sawhney et al. (2020). Ultimately, further experimenting with setups that blend the strengths of individual models' different prediction methods could also be effective in the mitigation of limitations of each model, leading to more reliable forecasts.

Integration of New Data Sources:

Integrating various data sources in real time and going more diverse would make the financial model more predictive. Traders are increasingly driven by real-time trading data with data like 'tick data' providing great depth into the dynamics of a market and supporting model responses that closely match quick market movements. Bringing in sentence analysis from social media sources, such as Twitter and financial blogs could allow for getting real time data about public interest. This may in turn show signals of market trends and thus can be considered as an early indicator (Beyaz, 2018).

Although the data generated from the alternative sources like satellite imagery for following up the economic activities or the blockchain datum for identifying the crypto-market trends can also be tapped into. This is because these new sources of data offer unique information that traditional market data sources may not capture. This feature provides these models with a margin advantage.

Chapter 6: Conclusion

6.1 Summary of Findings

The research has been carried out on using the hybrid algorithms of machine learning as well as sentiment analysis for predicting the trend in the high-dimensional stock market, on the Indian stock market, with NIFTY 50 and Sensex indices. The research confirmed that the machine learning based predictive models, specifically the models based on the LSTM, ARIMA and GARCH algorithms, are very reliable in the sense that they can successfully predict stock price pampers. Additionally, the sentiment analysis added a great deal to the forecasting accuracy Ma et al. (2023). It was during the highly volatile times, when the news event shaped a major part of it.

6.2 Theoretical and Practical Implications

The findings of this study not only bring forth the theory and applications in the field of financial predictive modelling, but they also extend them both as well. Empirically, the resultant data corroborate and amplify the Investors' Knowledge Argument which suggests that the stock market prices equally accommodate not only past or the present pieces of information but also the sentiments ascertained from a newspaper source. Basically, the models explained earlier will provide key analytical tools for investors and financial analysts that are more substantial than the traditional approaches now yielding more accurate results for well-informed trading decisions. The authors of the research work further pinpoint a promising opportunity of the hybrid models in identifying and improving the market trend predictions that can be applied extensively to devise one's dynamic trading strategy and sophisticated risk management tools Jiang (2021).

6.3 Future Directions

A greater vision emanates from this research; it would inspire further investigations. Research direction may shift to include more advanced tools including processing of real time data and introduction of more complex and applied machine learning models. Additionally, future studies can investigate various alternative sources of data such as social media sensing and economic indices. The abilities to enhance the precision of stock market models as well as adapt them to more market data diversity and heterogeneous market regimes including not only day trading but also long-term investing emerge along with the emergence of the said areas.

6.4 Final Thoughts

At the end, this thesis attributes the immense capability of integrating high-level machine learning algorithms and sentiment analysis into the elaboration of stock market predictive models to be one of its biggest achievements. It builds the groundwork for the research that follows as well as provides a basis for practical applications of this technology, which means that we face more advanced yet effective solutions in shaping stock market prediction.

References

Aldhyani, T. H., & Alzahrani, A. (2022). Framework for predicting and modelling stock market prices based on deep learning algorithms. Electronics, 11(19), 3149. Accessed From: https://www.mdpi.com/2079-9292/11/19/3149

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross‐Section of Stock Returns. The Journal of Finance, 61(4), 1645–1680. doi:10.1111/j.1540-6261.2006.00885.x

Beyaz, E. (2018). Effective stock price forecasting using machine learning techniques whilst accounting for the state of the market. International Journal of Engineering and Applied Sciences, 10(2), 115-122.

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. Machine Learning with Applications, 9, 100320. Accessed From: https://www.sciencedirect.com/science/article/pii/S2666827022000378

Eachempati, P., Srivastava, P. R., Kumar, A., Tan, K. H., & Gupta, S. (2021). Validating the impact of accounting disclosures on stock market: A deep neural network approach. *Technological Forecasting and Social Change*, *170*, 120903. Accessed From: https://www.sciencedirect.com/science/article/pii/S0040162521003358

Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P., & Pandey, N. (2021). Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science*, *7*, e340. Accessed From: https://peerj.com/articles/cs-340/

Guo, Y. (2022). Stock price prediction using machine learning. Master's thesis, KTH Royal Institute of Technology

Hao, Y., & Gao, Q. (2020). Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. *Applied Sciences*, *10*(11), 3961. Accessed From: https://www.mdpi.com/2076-3417/10/11/3961

Haq, A. U., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications*, *168*, 114444. Accessed From: https://www.sciencedirect.com/science/article/pii/S095741742031099X

Hou, X., Wang, K., Zhong, C., & Wei, Z. (2021). St-trader: A spatial-temporal deep neural network for modelling stock market movement. IEEE/CAA Journal of Automatica Sinica, 8(5), 1015-1024. Accessed From: https://ieeexplore.ieee.org/abstract/document/9395542/

Ibrahim, A., Kashef, R., & Corrigan, L. (2021). Predicting market movement direction for bitcoin: A comparison of time series modelling methods. *Computers & Electrical Engineering*, *89*, 106905. Accessed From: https://www.sciencedirect.com/science/article/pii/S0045790620307576

Jansen, S. (2020). Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python. Packt Publishing Ltd. Accessed From: https://books.google.com/books?hl=en&lr=&id=4f30DwAAQBAJ&oi=fnd&pg=PP1&dq=Jansen,+S.+(2020).+Machine+Learning+for+Algorithmic+Trading:+Predictive+models+to+extract+signals+from+market+and+alternative+data+for+systematic+trading+strategies+with+Python.+Packt+Publishing+Ltd.&ots=qBsGd8kjb6&sig=lE6TnxMCStwVBVuhVhGWkJnTpPw

Javed Awan, M., Mohd Rahim, M. S., Nobanee, H., Munawar, A., Yasin, A., & Zain, A. M. (2021). Social media and stock market prediction: a big data approach. *MJ Awan, M. Shafry, H. Nobanee, A. Munawar, A. Yasin et al.," Social media and stock market prediction: a big data approach," Computers, Materials & Continua*, *67*(2), 2569-2583. Accessed From: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3827106

Jayalakshmi, V., Jothi, S., & Shankar, K. (2019). A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing.

Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications*, *184*, 115537. Accessed From: https://www.sciencedirect.com/science/article/pii/S0957417421009441

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019. Accessed From: https://www.sciencedirect.com/science/article/pii/S0957417421004607

Kamalov, F. (2020). Forecasting significant stock price changes using neural networks. Neural Computing and Applications, 32(23), 17655-17667. Accessed From: https://link.springer.com/article/10.1007/s00521-020-04942-3

Kanwal, A., Lau, M. F., Ng, S., Sim, K. Y., & Chandrasekaran, S. (2022). BiCuDNNLSTM-1dCNN — A hybrid deep learning-based predictive model for stock price prediction. Expert Systems With Applications, 202, 117123. Accessed From: https://doi.org/10.1016/j.eswa.2022.117123

Li, Q., Tan, J., Wang, J., & Chen, H. (2020). A multimodal event-driven LSTM model for stock prediction using online news. *IEEE Transactions on Knowledge and Data Engineering*, *33*(10), 3323-3337. Accessed From: https://ieeexplore.ieee.org/abstract/document/8966989/

Li, Y., Ni, P., & Chang, V. (2020). Application of deep reinforcement learning in stock trading strategies and stock forecasting. Computing, 102(6), 1305-1322. Accessed From: https://link.springer.com/article/10.1007/s00607-019-00773-w

Lin, H., Zhou, D., Liu, W., & Bian, J. (2021, August). Learning multiple stock trading patterns with temporal routing adaptor and optimal transport. In Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining (pp. 1017-1026). Accessed From: https://dl.acm.org/doi/abs/10.1145/3447548.3467358

Liu, M., Choo, W. C., Lee, C. C., & Lee, C. C. (2023). Trading volume and realized volatility forecasting: Evidence from the China stock market. *Journal of Forecasting*, *42*(1), 76-100. Accessed From: https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2897

Ma, Y., Mao, R., Lin, Q., Wu, P., & Cambria, E. (2023). Multi-source aggregated classification for stock price movement prediction. *Information Fusion*, *91*, 515-528. Accessed From: https://www.sciencedirect.com/science/article/pii/S1566253522002019

Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7, e476. Accessed From: https://peerj.com/articles/cs-476/

Mehtab, S., & Sen, J. (2020). Stock price prediction using convolutional neural networks on a multivariate timeseries. *arXiv preprint arXiv:2001.09769*. Accessed From: https://arxiv.org/abs/2001.09769

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 53(4), 3007-3057. Accessed From: https://link.springer.com/article/10.1007/s10462-019-09754-z

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). Predicting stock market price movement using sentiment analysis: Evidence from Ghana. Applied Computer Systems, 25(1), 33-42. Accessed From: https://sciendo.com/article/10.2478/acss-2020-0004

Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction: Algorithms and techniques. Accessed From: https://eprints.soton.ac.uk/437785/

Pate, M. (2020). Stock Price Prediction Using Machine Learning, BIRLA VISHVAKARMA MAHAVIDYALAYA.

Prasetiyani, E., & Sofyan, M. (2020). Bankruptcy analysis using Altman Z-score model and Springate model in retail trading company listed in Indonesia Stock Exchange. *Ilomata International Journal of Tax and Accounting*, *1*(3), 139-144. Accessed From: https://www.ilomata.org/index.php/ijtc/article/view/98

Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. Procedia Computer Science, 167, 788-798. Accessed From: https://www.sciencedirect.com/science/article/pii/S1877050920308851

Sawhney, R., Agarwal, S., Wadhwa, A., & Shah, R. R. (2020, November). Spatiotemporal hypergraph convolution network for stock movement forecasting. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 482-491). IEEE. Accessed From: https://ieeexplore.ieee.org/abstract/document/9338303/

Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y. H. J. (2020). Machine learning stock market prediction studies: review and research directions. Journal of International Technology and Information Management, 28(4), 63-83. Accessed From: https://scholarworks.lib.csusb.edu/jitim/vol28/iss4/3/

Ta, V., Liu, C., & Tadesse, D. A. (2020). Portfolio Optimization-Based stock prediction using Long-Short Term memory network in quantitative trading. Applied Sciences, 10(2), 437. Accessed From: https://doi.org/10.3390/app10020437

Thakkar, A., & Chaudhari, K. (2020). CREST: cross-reference to exchange-based stock trend prediction using long short-term memory. *Procedia Computer Science*, *167*, 616-625. Accessed From: https://www.sciencedirect.com/science/article/pii/S1877050920307948

Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, *65*, 95-107. Accessed From: https://www.sciencedirect.com/science/article/pii/S1566253520303481

Wang, H., Wang, J., Cao, L., Li, Y., Sun, Q., & Wang, J. (2021). A stock closing price prediction model based on CNN-BiSLSTM. Complexity, 2021, 1-12. Accessed From: https://www.hindawi.com/journals/complexity/2021/5360828/

Wen, Y., Lin, P., & Nie, X. (2020, March). Research of stock price prediction based on PCA-LSTM model. In IOP Conference Series: Materials Science and Engineering (Vol. 790, No. 1, p. 012109). IOP Publishing. Accessed From: https://iopscience.iop.org/article/10.1088/1757-899X/790/1/012109/meta

Wu, J. M. T., Li, Z., Srivastava, G., Tasi, M. H., & Lin, J. C. W. (2021). A graph‐based convolutional neural network stock price prediction with leading indicators. *Software: Practice and Experience*, *51*(3), 628-644. Accessed From: https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.2915

Wu, J., Xu, K., Chen, X., Li, S., & Zhao, J. (2022). Price graphs: Utilizing the structural information of financial time series for stock prediction. Information Sciences, 588, 405-424. Accessed From: https://www.sciencedirect.com/science/article/pii/S0020025521013104

Wu, X., Chen, H., Wang, J., Troiano, L., Loia, V., & Fujita, H. (2020). Adaptive stock trading strategies with deep reinforcement learning methods. *Information Sciences*, *538*, 142-158. Accessed From: https://www.sciencedirect.com/science/article/pii/S0020025520304692

Yan, X., Weihan, W., & Chang, M. (2021). Research on financial assets transaction prediction model based on LSTM neural network. Neural Computing and Applications, 33(1), 257-270. Accessed From: https://link.springer.com/article/10.1007/s00521-020-04992-7

Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, *32*(6), 1609-1628. Accessed From: https://link.springer.com/article/10.1007/s00521-019-04212-x

Zhang, C., Zhao, Y., & Zhao, H. (2022). A novel hybrid price prediction model for multimodal carbon emission trading market based on CEEMDAN algorithm and window-based XGBoost approach. Mathematics, 10(21), 4072. Accessed From: https://www.mdpi.com/2227-7390/10/21/4072