Applying Machine Learning Technique to Forecast Stock Exchange Markets

Ismael Ben Daoud

Supervisor: Jean Paul Tabone

June - 2023

A dissertation submitted to the Institute of Information and Communication Technology in partial fulfilment of the requirements for the degree of BSc (Hons) in Business Analytics

Authorship Statement

This dissertation is based on the results of research carried out by myself, is my own composition, and has not been previously presented for any other certified or uncertified qualification.

The research was carried out under the supervision of Jean Paul Tabone.

5th June 2023 Ismael Ben Daoud

Date Signature

Copyright Statement

In submitting this dissertation to the MCAST Institute of Information and Communication Technology, I understand that I am giving permission for it to be made available for use in accordance with the regulations of MCAST and the Library and Learning Resource Centre. I accept that my dissertation may be made publicly available at MCAST's discretion.

5th June 2023	Ismael Ben Daoud
Date	Signature

Acknowledgements

I would like to give my utmost gratitude to my supervisor, Mr Jean Paul Tabone, for guiding me on the right path, providing assistance and support throughout the research I sough out to accomplish. I am grateful that I have been chosen as one of her students.

I would like to also thank my family, friends and my partner for supporting me throughout these stressful months and encouraging me to never give up.

Abstract

This paper investigates the use of artificial intelligence, specifically machine and deep learning techniques, to predict Using historical price data, five financial markets' closing prices. The study focuses on five datasets corresponding to Apple, Amazon, Meta, Google, and Tesla's financial markets in order to acquire accurate price attribute prediction through the use of long-short term-memory classification artificial neural networks. These datasets include historical prices obtained from the official Yahoo Finance (YF) API, spanning the years 2018 through 2022. The dataset contains numerous attributes, including price. Date on which the share will be transacted on the exchange for the first time. Open The initial price of a stock is its price at the start of a trading session or market day. It is the initial price documented upon the opening of the stock market. High price indicates the stock's highest recorded price during a given trading session. It reflects the highest price the stock attained during that period. The low price is the stock's lowest recorded price during a particular trading session. It indicates the lowest price the stock reached during the specified period. The close price is the last price recorded for a particular stock during a trading session. It denoted the lowest price the stock reached during the specified time frame. Adjusted The close price takes into account events such as dividends, stock splits, and other corporate actions that impact the price of the stock. It is a modified closing price that takes these adjustments into consideration to provide a more accurate reflection of the stock's value. During a given trading transaction, volume refers to the total number of shares or contracts traded for a particular stock. It reflects the quantity of activity and liquidity on the stock's market. Higher volumes are frequently indicative of heightened investor interest and can be a significant factor in price fluctuations. To ensure data veracity, the Close attribute, which represents the last traded price of the day, will be factored into forecasts of what was utilized following proper data cleansing. The dataset was divided 70with the plurality allocated for training. In addition to evaluating the prototype, an online survey was conducted to determine what people know about the appli- cability of machine learning across various industries. The survey was designed based on LSTM model prototype performance and analysed using IBM SPSS to collect demographic data. Utilizing Chi-Squared, T-test, and ANOVA tests, infer- ential analysis was used to examine the survey data. Long-Short Term Memory (LSTM) is the chosen paradigm for this investigation, which necessitates a three- dimensional array. Four methods were used to evaluate the efficacy of the models: Accuracy, Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). The outcomes show that LSTM, with an accuracy rate of 90% or better across all five datasets, is the most successful AI model for price prediction.

Keywords: Price Prediciton, LSTM, Neural Network, Machine Learning, Deep Learning, Yahoo Fiance, RMSE, MSE, MAPE, Accuracy

Table of Contents

List of Figures

List of Tables

List of Abbreviations

NN	Neural Network
ML	Machine Learning
DL	Deep Learning
FCN	Fully Convolutional Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory

Chapter 1: Introduction

1.1 Research Background

This study centres on machine learning, a subfield of artificial intelligence that employs algorithms and data to replicate moral knowledge processes and enhance overall precision [5]. Machine learning can be categorised into three subcategories: supervised, unsupervised, and reinforcement learning [6]. This method offers the benefit of efficiently handling large amounts of data and identifying notable trends and patterns to achieve desired results. A drawback is the requirement of a substantial training dataset, which could result in delays while awaiting the generation of fresh data. Machine learning, as outlined in reference [7], seeks to develop models that can imitate and generalise data by distinguishing between different items to achieve the desired level of precision. The aim is to tailor the analysis to the distinct attributes of each item, resulting in more accurate and significant outcomes.

The ROI of the close price attribute (ROI) using supervised learning, particularly supervised machine learning, is the specific focus of this study. In this method, labelled datasets are used to train algorithms for precise outcome prediction and data categorization. The objective is to predict the stock market using neural networks, a prototype created using Python script, and stock data generated from historical trends. In order to track and understand market changes over time, this research uses a quantitative research methodology to analyse stock data that is already available. The study aims to investigate factors influencing market predictions and different market types using historical datasets. Python will be used to create a data scraping bot and produce various datasets for each market.

1.2 Research Purpose

This study's main goal is to illustrate how machine learning approaches might increase profitability by enhancing decision-making via precise forecasts based on historical data and current exchange rates. For analysis and classification, many machine learning techniques are used. Automatic analysis of user-generated views is now essential for successful decision-making based on prototype accuracy. Furthermore, this research aims to get input on prediction accuracy from subject-matter specialists in the business. Making better judgments as a result will eventually improve the anticipated results.

This study serves as a valuable reference for researchers interested in machine learning techniques and is accessible to anyone interested in this area of study.

1.2.1 Hypothesis and Research Questions

The Hypothesis of this research is:

”The implementation of machine learning techniques can maximize profitability by improving decision-making through accurate predictions based on past data and current exchange rates.” This study will make an effort to support or refute important stock prediction factors. This study aims to support or refute key stock forecasting factors. The main objective of this study is to enhance market value forecasting. Online surveys and experiments will be used to analyse the market's relevance using machine learning algorithms. The results will be put into practice and evaluated against earlier research. The study's research questions (RQ) are as follows: The research questions in this study are:

RQ1: Are movements in rates of stock exchange predictable when taking into consideration past datasets with exchange rates?

RQ2: Can Machine learning contribute to an optimal decision when dealing with stock exchange?

The target participants for the study (experts in the field) and the sample size of 16 participants will be determined by these research questions. An online survey made with Google Forms will be used to collect data, and Python programming will be used in the experiment section to demonstrate how a machine learning algorithm for forecasting particular foreign exchange markets can be built. Descriptive statistics will be used in data analysis to describe the characteristics of the datasets. The methodology will also include a review of related techniques used in various markets, an assessment of the available research, and the creation of a prototype using a variety of techniques. Market forecasting will be the main focus of the prototype, and there will be thorough explanations of data cleansing, model training/testing, and algorithm implementation. We'll also talk about metrics and accuracy measures.

1.2.2 Significance of this Research

This study aims to investigate the efficacy of machine learning in optimising profits through improved decision-making based on historical data and prevailing exchange rates. Lasod and Pawar [8] employ various machine learning methods for categorization analysis. Automated analysis of opinions shared on web platforms is crucial for making informed decisions.

Better predictions can be made by utilizing machine learning techniques, which allow for more informed decision-making. On the accuracy of the adopted solution, industry experts in the relevant field will offer their opinions. Additionally, this study can be used as a helpful resource for scientists working on related projects and as reading material for people interested in this field of research.

This study aims to advance our understanding of machine learning strategies used in the stock market and explore the potential of forecasting stock movements using exchange rates. It also aims to investigate the potential of machine learning in efficient stock exchange decision-making

1.3 Importance of Research

This study has great significance for a number of reasons. Firstly, it offers a chance to combine the business and software sectors, enabling the researcher to create a project that combines both facets. As a business analytics student, you have a great opportunity to investigate how forecasting affects the market using various analytical methods. Additionally, by highlighting its growth and offering insightful data on market performance, this project sheds light on the stock market's dynamic nature. Second, a new technology in the area of artificial intelligence is machine learning. The researcher is interested in learning about the inner workings of machine learning techniques, the prerequisites for developing successful systems, and their various approaches and uses. Although machine learning has been used in many fields, such as bio-informatics and weather forecasting, its use in stock market prediction has not received much attention. Consequently, this research offers a chance to learn more about this area of study. Additionally, this study can be used as a helpful resource for scientists working on related projects and as reading material for people interested in the use of machine learning techniques. Its conclusions and insights can add to the body of existing knowledge and offer dependable data for ongoing future research in this field.

1.4 Research Boundaries

This study focuses specifically on using machine learning to forecast the performance of the stock market. It examines how machine learning methods can assess vast quantities of data, discover trends and patterns, and help with predictions. It does not, however, include information other than stock exchange data. Furthermore, this project will not use machine learning techniques for reinforcement. In order to predict market prices for various markets, the research will primarily use a deep learning-based model for stock prediction.

1.5 Research Outline

This research will contain the following chapters:

1. Introduction: Provides an overview of the research problem, including a description of its background, purpose, hypotheses, and research questions, as well as its significance, boundaries, and outline.

2. Literature Review: Provides a summary of literature pertinent to the research. Contains chapters on dealing with stocks, various trading markets, various stock analysis techniques, machine learning, supervised learning, and unsupervised learning, as well as neural networks, deep learning networks, recurrent neural networks, long short-term memory networks, and natural language processing.

3. Research Methodology: This section discusses the employed research methodology, including the study plan, data collection techniques and tools, prototype development, pilot testing, error considerations, and ethical issues.

4. Analysis of Results and Discussion: Describes the data analysis techniques used throughout the study and provides an analysis and discussion of the acquired data. Includes analysis and discussion of the online survey, the experiment, the literature, the hypothesis and research questions, and a summary.

5. Conclusions and Recommendations: Summarises the research findings, offers suggestions, calls attention to the study's limitations, identifies areas requir-

ing further study, and concludes the investigation.

1.6 Conclusion

The main literature that was used to build a solid foundation for the research will be covered in depth in the following chapter. It will cover a variety of subjects, including stocks, different kinds of trading markets, and machine learning methods.

Chapter 2: Literature Review

This chapter will present a literature review pertaining to the selected research topic. This chapter covers the topics of stock trading, market types, stock analysis techniques, and machine learning methods such as supervised and unsupervised learning, neural networks, deep learning networks, recurrent neural networks, long short-term memory networks, and natural language processing (NLP).

2.1 Literature Map

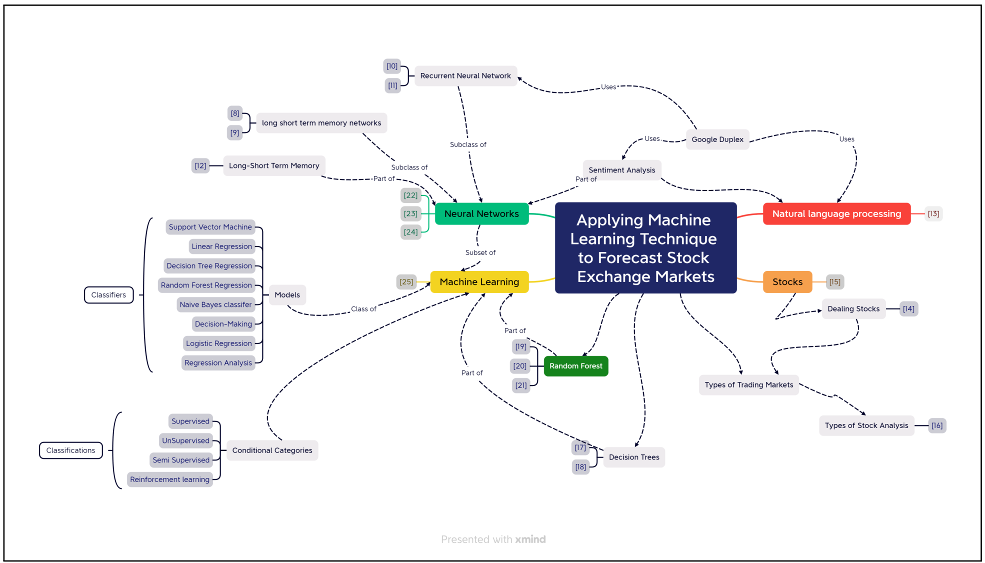


Figure 2.1: *Literature Map.*

2.2 Introduction

The unpredictable nature of the stock market poses a significant challenge for analysts in accurately forecasting price fluctuations. Before the emergence of computers, trading and investing in stocks and commodities were predominantly based on intuition. With the growth of trading and investment, people sought ways to increase profits and reduce risk. Conventional methods for predicting stock prices, including statistical, technical, fundamental analysis, and linear regression, were employed to anticipate market trends and profit from them. The efficacy of these techniques remains a subject of debate among analysts due to their inconsistent ability to serve as reliable predictive tools. Machine learning methods have shown higher accuracy and effectiveness than traditional approaches [9] and [10] for predicting price changes.

Stock prediction studies commonly utilise three main data sources. Financial news sources offer real-time updates on firms, sectors, economies, and financial markets, while historical stock information pertains to a stock's price and trading activity within a particular time frame. The dataset comprises open, high, low, and close prices, alongside volumes and adjusted pricing. Integrating the collective impact of these factors is an effective strategy for managing market volatility. Due to the large amount of data, researchers often use various techniques to efficiently handle and analyse it. LSTM, a type of artificial recurrent neural network, is widely used for stock prediction. The subsequent section will delineate diverse machine learning techniques and methodologies employed by researchers to predict stock prices and market trends.

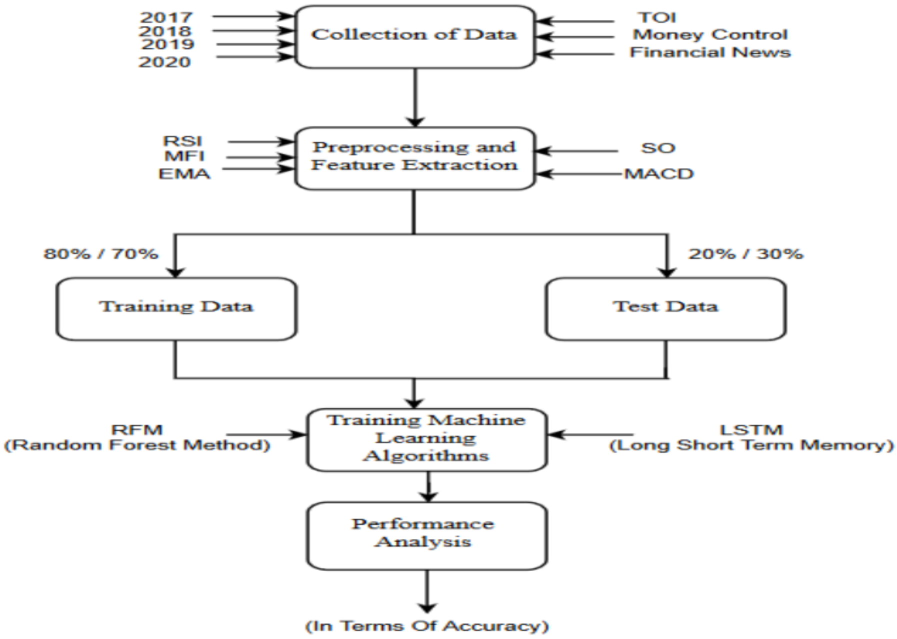


Figure 2.2: *Process of Stock Market Prediction [1]*

2.3 stocks

As per the information provided by Investor.gov [11], a stock denotes proprietorship in a corporation and signifies an entitlement to the company's assets and earnings. Equities is a term frequently used to refer to stocks. This financial instrument is commonly referred to as equity or shares. The possession of shares in a corporation endows specific entitlements, such as the capacity to participate in shareholder assemblies, obtain dividends, and transfer the shares. The extent of an individual's ownership in a company is directly proportional to their voting power and the level of indirect control they exercise over the organisation. The paramount focus of individuals who engage in stock speculation pertains to the intrinsic worth of the stock.

Stock trading takes place on stock exchanges, which serve as marketplaces for the buying and selling of securities by merchants. In contemporary times, a predominant proportion of transactions are executed through electronic means as opposed to physical trading platforms. The utilisation of electronic communication enables merchants to execute prompt transactions. There exist a multitude of stock exchanges across the globe, including but not limited to the London Stock Exchange (XLON), the Shanghai Stock Exchange (SSE), the Securities Dealers Automated Quotations (NASDAQ), and the New York Stock Exchange (NYSE). The contemporary markets are characterised by their interconnectivity, whereby digital linkages exist between them, thereby influencing each other's performance [12].

2.3.1 Deadling Stocks

The capacity to buy and sell securities for financial gain is a pivotal element for most stock traders, as previously mentioned. The initial valuation of a company's stock is determined by its overall worth, revenue, and other pertinent considerations. Hence, the stock's worth varies in accordance with its supply and demand. The purchase of a significant number of securities results in an increase in the stock's value. Conversely, the value of the stock will decrease if a significant proportion of it is sold and becomes accessible [13]. The stock markets tend to exhibit a positive trend owing to various factors, including inflation and the influx of new capital on a daily basis. The ability to generate profits does not necessarily imply a successful trading system. This is because an individual can simply purchase equities that exhibit strong performance and, over time, accrue profits. It is imperative for a trader to have a thorough understanding of the intrinsic value of a company's stock. The trader can make use of this knowledge to purchase equities at a price lower than their perceived value and sell them when their perceived value surpasses their actual value [14].

2.3.2 Types of Trading Markets

Zineb [15] identifies two primary trading strategies: long-term and short-term. As stated earlier, investing in established equities for a prolonged duration can yield profits. This approach entails lower time and labour investment from the trader and generally results in greater returns compared to bonds and other investment options. These techniques exhibit significant potential in the domain of forecasting owing to their ability to efficiently preserve and utilise data. Conversely, short-term trading involves seeking to gain from minor fluctuations in stock prices. Stock fluctuations can occur over short periods of time, such as intra-day trading, or over longer periods ranging from a day to several days, known as swing trading [16]. Short-term trading can demand significant time and attention, especially when executed manually. This is due to the need for the trader to closely track the latest stock price movements and market developments to optimise potential gains [17]. This thesis will focus on a trading strategy utilising long-short term memory. Due to the time-consuming nature of short-term trading, traders may benefit from employing machine learning algorithms for price prediction [18].

2.4 Types of Stock Analysis

The trend of stock market volatility is widely recognised. In contrast, it has been emphasised that the volatility of stock prices is attributed to various forces, which are the determinants of stock prices. The complexity of predicting stock prices is due to the multitude of factors involved. [19] Forecasting stock market prices involves considering a multitude of factors, including historical price trends, political climate, psychological sentiments towards a stock, global economics, and a company's financial condition (Introduction to the foreign exchange market). Violeta is a 19-year-old individual. Despite the abundance of literature on the subject, a comprehensive technique or algorithm for accurately predicting stock values remains elusive. Several studies suggest that accurate prediction of stock prices is highly challenging. Each of these components can be further segmented. This illustrates the intricate and varied elements that impact stock prices. As stated by reference [20], there exist two primary categories of analysis: fundamental and technical. It is infeasible for a singular algorithm to comprehend and assess the importance of these variables on the worth and market value of a particular investment. It is feasible to predict stock prices with a certain degree of error by analysing a subset of the various factors that can impact them. Diverse factors require different methods for analysing fluctuations in stock prices.

2.5 Machine Learning

Machine learning is a field of study that aims to replicate intelligent human behaviour through the use of machines, as noted by reference [21]. Both humans and machines employ memory and electrical signals to process information, allowing for data transmission, retrieval, and analysis. The subfield of machine learning within the domain of AI aims to replicate human behaviour. For machines to perform tasks like visual object classification or text analysis, they require a specific dataset for learning purposes. Contemporary machine learning research primarily centres on the creation of algorithms tailored to particular use cases. Machine learning employs data to develop a predictive model for making predictions. The primary difference between traditional computer code and machine learning code lies in the absence of hard-coded rules in the latter. This is a characteristic feature of machine learning. The model's principles are generated by the code itself [22]. Sufficient data access is crucial for machine learning, as data quality has a direct impact on algorithmic performance. The algorithm partitions the data into training and testing sets to identify relationships between the input and expected output. A model is developed based on the training data and subsequently assessed for its effectiveness using the testing data. Acquiring suitable data is a critical component of machine learning, where the significance and magnitude of data are pivotal [23].

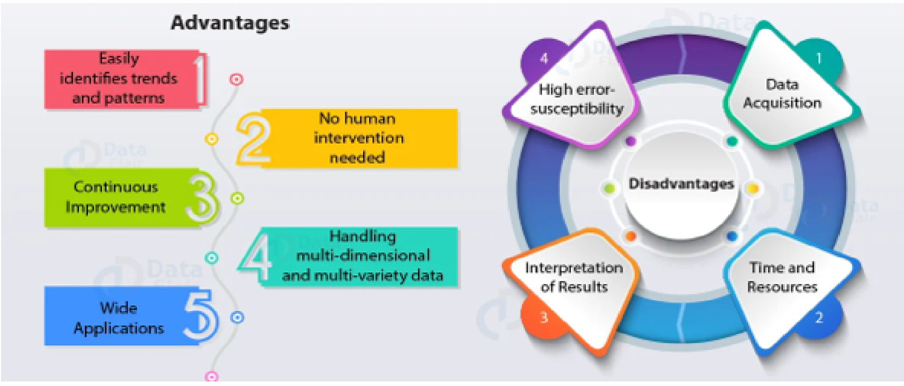


Figure 2.3: *Advantages and Disadvantages of Machine Learning [2]*

2.5.1 Supervised Learning

Supervised learning algorithms, as stated in reference [24], utilise training data to produce a function capable of mapping new samples. Supervised learning algorithms can yield remarkable outcomes and offer unforeseen insights in appropriate data and conditions, thereby augmenting the overall efficacy of an application. Supervised learning techniques are appropriate when labelled data is present, i.e., a dataset that contains explicit input and predicted output data. The anticipated result, commonly known as objective data, denotes the output that the machine learning algorithm is expected to obtain. The selection of data generation methods is contingent upon the nature of the data and its intended use. Financial applications commonly utilise input data in the form of stock prices. Financial applications commonly utilise target data that represents values within a specified timeframe. In the context of stock value forecasting, the predicted output is the closing price of the stock on the following day [25].

2.5.2 Unsupervised Learning

Unsupervised learning algorithms, as explained in reference [26], are employed in cases where the training data lacks information about the intended output. The main aim of unsupervised learning is to gain an understanding of the underlying structure of data through the identification of patterns or relationships. Unsupervised learning lacks a definitive answer to learn from, as opposed to supervised learning, due to the absence of explicit output in the training data. Unsupervised learning algorithms uncover interesting data patterns through self-guided learning from input data.

2.6 Neural Networks

According to [27], neural networks are another type of supervised learning that extract patterns from complex data, making them extremely valuable. Neurons, which are interconnected elements of neural networks, operate in parallel to solve problems. The neural organization of the human brain inspired the development of the first neural networks, which are still loosely founded on this concept. In recent years, neural networks have garnered considerable interest, accompanied by a surge in research papers [27]. The neural network architecture shown in Figure

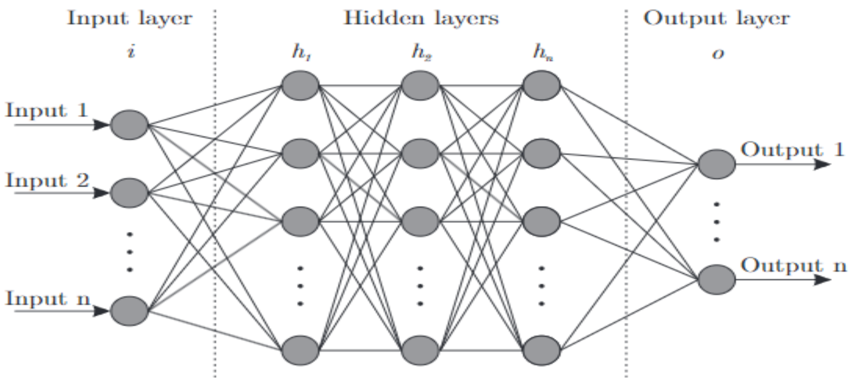


Figure 2.4: *Artificial Neural Network Architecture [3]*

2.4 involves a unidirectional flow of information during the computation of the optimal output. Neural networks consist of various elements, such as an input layer that normalises input data, a hidden layer that processes information, and an output layer that has an equivalent number of neurons as outputs [28].

2.7 Deep Learning Networks

Deep learning networks have been employed in various applications, such as speech and image recognition, despite potential issues such as vanishing gradients, as reported in [29]. Figure 4 illustrates that a shallow neural network comprises a single hidden layer, whereas a deep neural network consists of multiple hidden layers. According to reference [29], a deep neural network that has undergone extensive training can significantly surpass the performance of a shallow neural network in intricate scenarios. Deep neural networks exhibit significant potential for applications involving images and sequences. CNNs and their variations are highly effective in image-based tasks, while RNNs and their variations are well-suited for sequence-based tasks, such as leveraging the sequential nature of historical data for stock market forecasting. Deep learning networks exhibit remarkable performance in text and speech-based applications [30]. This thesis pertains to a specific recurrent neural network type, which will be elaborated upon in the subsequent chapter.

2.8 LSTM Recurrent Neural Networks

Recurrent neural networks are one of the few types of neural networks with feedback, whereas [31] conventional neural networks are believed to have inputs that are independent of one another. A recurrent neural network (RNN) is a type of neural network designed to learn from sequential and time-varying patterns. Due to the nonlinear character of the stock market, RNN is an excellent method for preserving historical data and making accurate forecasts. In a variety of practical applications, such as music or text analysis, the employed data is frequently not independent. This is especially true for sequential data, where the feedback mechanism enables the network to comprehend the meaning of sentences or paragraphs in text or a sequence of tones that sounds pleasing to humans. The feedback loop in a recurrent neural network (RNN) functions as a form of "memory," allowing the network to accumulate information about previously performed calculations. [31].

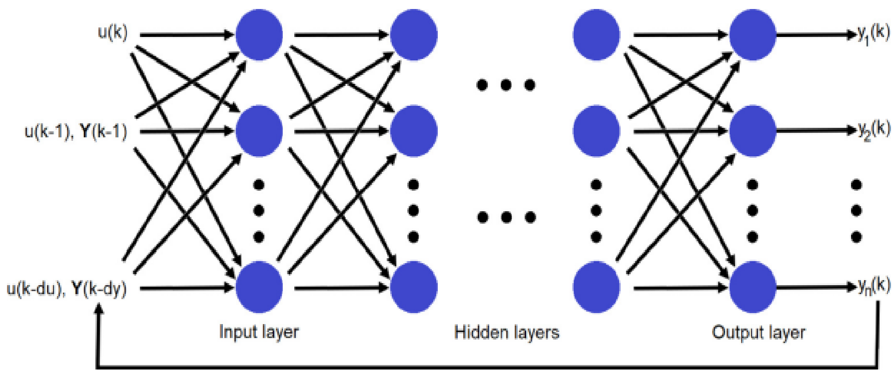


Figure 2.5: *Artificial Neural Network Architecture [3]*

Figure 2.5 depicts the input layer's function of normalising the meaning RNN by adjusting the input data to a specified range. The layer in question is accountable for both the processing and output of individual units, given that it comprises a minimum number of neurons equivalent to the number of outputs from the network [28]. The nomenclature depicted in Figure 5 comprises the following designations:

Y(x) = the vector of feedback outputs at the discrete-time instant specified by (x),

Y1(k) = the outputs at the discrete-time instant specified by (.), I = 1, n,

Du = maximal level of input delay,

Dy = maximal level of feedback outputs delay

2.8.1 Long Short-Term Memory Network

LSTM units are a type of RNN units. An RNN that incorporates LSTM units is commonly referred to as an LSTM network. The present thesis employs the acronym LSTM to denote Long Short-Term Memory. Hochreiter and Schmidhuber (1997) proposed LSTM as a solution to a particular problem.

An LSTM model can effectively learn and analyse past data. The stock market's complex data structure necessitates the use of a unique unit, referred to as a memory cell, to generate an appropriate response. LSTM outperforms RNN due to its utilisation of specialised units in conjunction with conventional units, which facilitates the retention of information over longer periods [32].

[40] identifies the key distinction between a conventional RNN and an LSTM lies in the latter's ability to selectively erase and retrieve information from its memory. Compared to a Recurrent Neural Network (RNN), the Long Short-Term Memory (LSTM) model is more effective in managing prolonged dependencies.

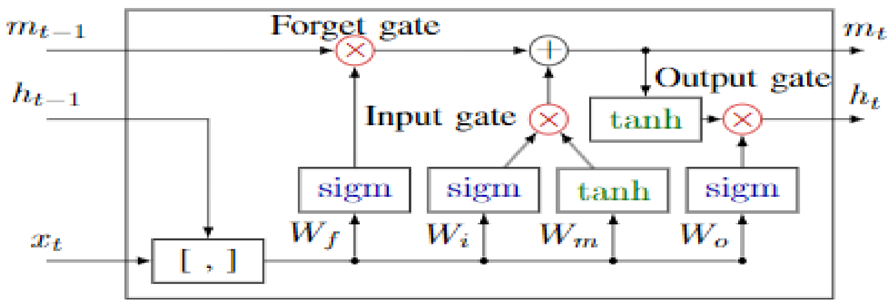


Figure 2.6: *Architecture structure of Long Short-Term Memory Unit [4]*

The LSTM unit depicted in Figure 6 utilises gating, which is a form of component-wise multiplication. Furthermore, it illustrates the disregard gate (f), the entrance gate (i), and the egress gate (o). The stock's historical data is stored in memory location (x). The input, output, and erase gates respectively regulate the writing of input to the cell, reading of the cell's output, and resetting of the previous cell value. Accurate and timely stock price forecasts are crucial for maintaining national economic stability. According to Wang (2015), stock price forecasting presents challenging characteristics due to market volatility. This network's efficacy lies in its utilisation of the gating method, which enables the network to regulate the activation and deactivation of its switches. As a result, it will ascertain the necessary and discretionary pieces of information.

2.9 Conclusion

This chapter provides a review of the literature on the stock exchange market and explores the potential of machine learning techniques for predicting stock exchange prices. Various market conditions that impact financial markets were analysed.

Forecasting stock prices is challenging due to the constant fluctuations in market prices, as evidenced by research findings. The optimal approach for researchers involves amalgamating historical stock data with financial information. Recent research indicates that LSTM exhibits superior efficacy compared to other machine learning techniques.

This paper presents a detailed analysis of the research methodology and model implementation for predicting stock movement. This chapter presents a literature review relevant to the chosen research topic. This chapter will cover the Research Methodology, encompassing the research strategy, data collection methods and tools, prototype, pilot testing, errors, and ethical considerations.

Chapter 3: Research Methodology

This section outlines the methodology employed in the present investigation. It offers a thorough depiction of the evaluated information. Furthermore, the methodology for data cleansing and partitioning into training and testing sets is described. To enhance the understanding of the variables involved in the study.

This chapter presents the utilisation of Long Short-Term Memory (LSTM) and Neural Networking processing language (NNL) analysis for stock prediction through machine learning techniques. The obtained results are discussed.

3.1 Research Strategy

Deductive reasoning involves deriving logical conclusions from one or more given statements. The process, as explained in reference [33], is marked by decreased uncertainty as the conclusion is obtained by utilising universally applicable principles within a confined domain of discussion. In the end, a conclusion is reached by narrowing down the potential options. The aforementioned conclusion was deduced through deductive reasoning.

Applied research was employed to fulfil the hypothesis requirements. This is achieved through the utilisation of machine learning algorithms on past datasets. The second phase of this study involved the extraction of data from Yahoo Finance through web scraping techniques. The research employed a quantitative approach for data analysis and presentation of results. The study utilises an independent and dependent variable to train the model, as depicted in the diagram.

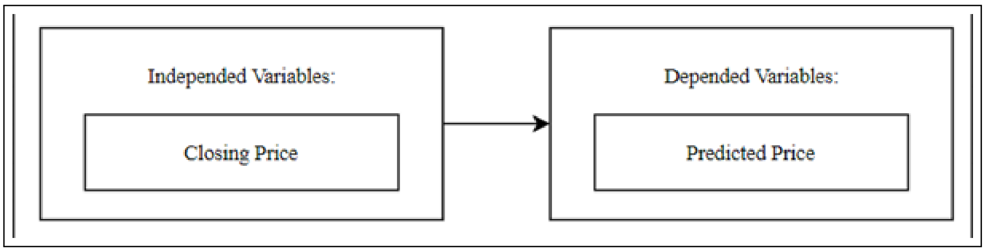


Figure 3.1: *Variable Framework.*

In this study, the relationship between the closing price of historical datasets and the predicted price values is investigated. In this context, the independent variable is the closing price of historical datasets. Predicted price values constitute the dependent variable. The purpose of this research is to develop a model or methodology that can forecast or predict future price values using historical closing prices. Using the available historical data, the predicted price values are the outcomes or results that researchers wish to estimate or forecast.

3.1.1 Research Method

This study focuses on the application of long-short-term memory analysis of neural networks in machine learning techniques for stock price prediction. Market prediction is a complex task, as outlined in Section 2.2.3 of Chapter 2. Various methodologies are employed to develop models that account for the impact of external factors on stock fluctuations. Neural networks, as previously discussed, consist of interconnected neurons that work in parallel to solve problems. To facilitate this analysis, a five-year historical dataset is employed. Historical data-based stock prediction studies employed Open, High, Low, Volume, and Close variables [34]. This research aims to evaluate the precision and efficacy of adjusted close, a metric that incorporates dividends and closing price, in comparison to a machine-based approach.

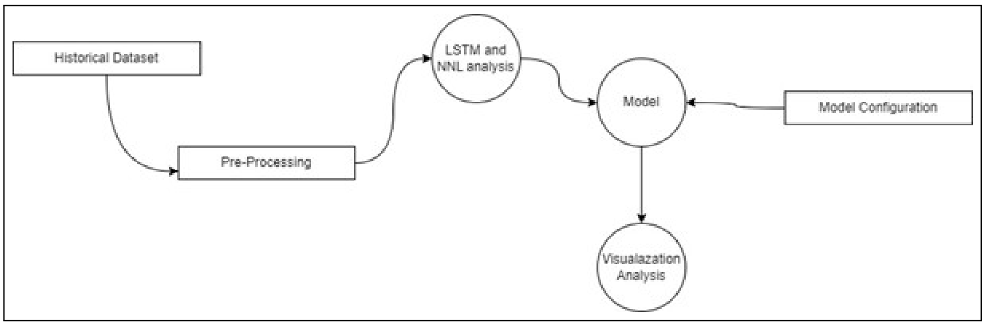


Figure 3.2: *Data Flow Diagram: Overview.*

Figure 3.2 presents a comprehensive overview of the prototype. The initial segment focuses on the acquisition of historical data and the preliminary preparation of an LSTM model on the obtained data. LSTM is chosen as the analytical tool for the research presented in section 2.7 due to its superiority over other tools in handling larger data sets. Subsequently, the model shall be configured and executed to facilitate the visualisation of outcomes.

3.1.2 Research Method Justification

Quantitative research methodology was selected for this investigation due to its accuracy and reliability in data collection. The numerical presentation of collected data enhances its reliability. The utilisation of quantitative methods enables a rapid and uncomplicated analysis of outcomes. Furthermore, it yields precise outcomes of the investigation by relying on quantitative data, thereby circumventing subjective viewpoints and predispositions.

Qualitative research diverges from quantitative research by not relying on numerical data, but rather by prioritising the collection of respondents' subjective opinions, perspectives, and apprehensions. Qualitative research differs from quantitative research in its approach. In qualitative research, researchers have the ability to adjust their questions in order to obtain more detailed information from participants on a particular research topic. Qualitative research allows for flexible data collection procedures that are not bound by a fixed framework. Apart from the responses, significant insights can be obtained by analysing the behaviour, tone, and emotions of the respondents. Qualitative data can offer valuable context and insights that may not be captured by numerical data, even if the responses collected do not align with the researcher's expectations.

3.2 Data Collection Methods and Tools

Data is a crucial component in machine learning applications, particularly those related to foreign stocks. This study involved generating long-term projections, which necessitated the compilation of particular data categories. The aggregated dataset comprised the Highs, Lows, Open, Close, Adjusted Close, and Volume for each stock market, alongside other relevant metrics sourced from Yahoo Finance.

The ”High” and ”Low” values of a stock represent its maximum and minimum prices during a trading day, respectively. The ”Open” and ”Close” values in stock trading respectively denote the initial and final prices of a stock for a given day. The ”Adjusted Close” denotes the final trading price of a stock during the preceding trading session. Moreover, the ”Volume” metric denotes the quantity of shares exchanged for a particular stock, index, or investment within a designated timeframe.

Historical data for the equities was obtained using the Yahoo Finance API, covering the period from 2018 to 2022. The historical data for the chosen equities was obtained and organised into a CSV file. The aim of this study was to enhance the precision of stock price forecasting by utilising long-short term memory (LSTM) methodologies with the collected data.

Markets	Start Date	End Date
Apple – (AAPL) Google – (GOOG) Meta – (META) Amazon – (AMZN) Tesla – (TSLA)	02-01-2018	01-12-2022

3.2.1 Data

The importance of data in the effectiveness of machine learning algorithms is especially notable in the context of stock markets, as previously discussed. Choosing suitable data can pose a difficulty, given the varied datasets employed in research studies and the absence of consensus on the most pertinent variables. This study focused on five corporations, namely Apple, Amazon, Tesla, Google, and Meta, to address this issue. Yahoo Finance furnished the historical stock prices of these companies for the past five years.

Overfitting poses a notable challenge in the field of machine learning, as it may lead to the identification of spurious correlations that lack causal significance. Furthermore, as elaborated in the second chapter, evaluating the adequacy of available stock price data is challenging. The accurate assessment of data suitability is ambiguous. The Open, Close, High, and Low values were employed as input variables for each stock market and compared with the Adjusted Close value. The aim was to forecast the closing prices of a specified stock for the following day.

The next step entailed calculating the required sample size for the study. Both probability and non-probability sampling techniques were evaluated. Probability sampling is a method of selecting individuals from a population through the application of probability theory. It involves a random selection process. Non-probability sampling is based on the researcher's subjective evaluation and decision-making in selecting the sample. Non-probability sampling was utilised in this study as participants with specialised expertise in machine learning, artificial intelligence, and business were involved. The sample size was determined using a sample calculator, as shown in the figure provided. The calculator determined that a sample size of 11 is necessary to achieve a 95% confidence level and a 10% margin of error for a population of 12. After sample selection, the qualitative technique was applied to participants and the data obtained was analysed.

3.2.2 Advantages and Limitations of the chosen Data Collection Tools

Google Forms' user-friendly interface enables easy survey creation. The study benefited from the flexibility provided by the survey platform in terms of structuring and customising survey sections and questions, which enabled the separation of distinct objectives. In the context of question-answering, individuals are able to promptly retrieve information or succinct explanations, which proves to be advantageous when encountering unfamiliar jargon. Disseminating the survey to participants can be easily accomplished through the use of links or email.

Google Forms allows for quantitative research data to be analysed and visualised as an additional advantage. However, certain limitations must be taken into account. Internet access is a prerequisite for survey completion, and file size

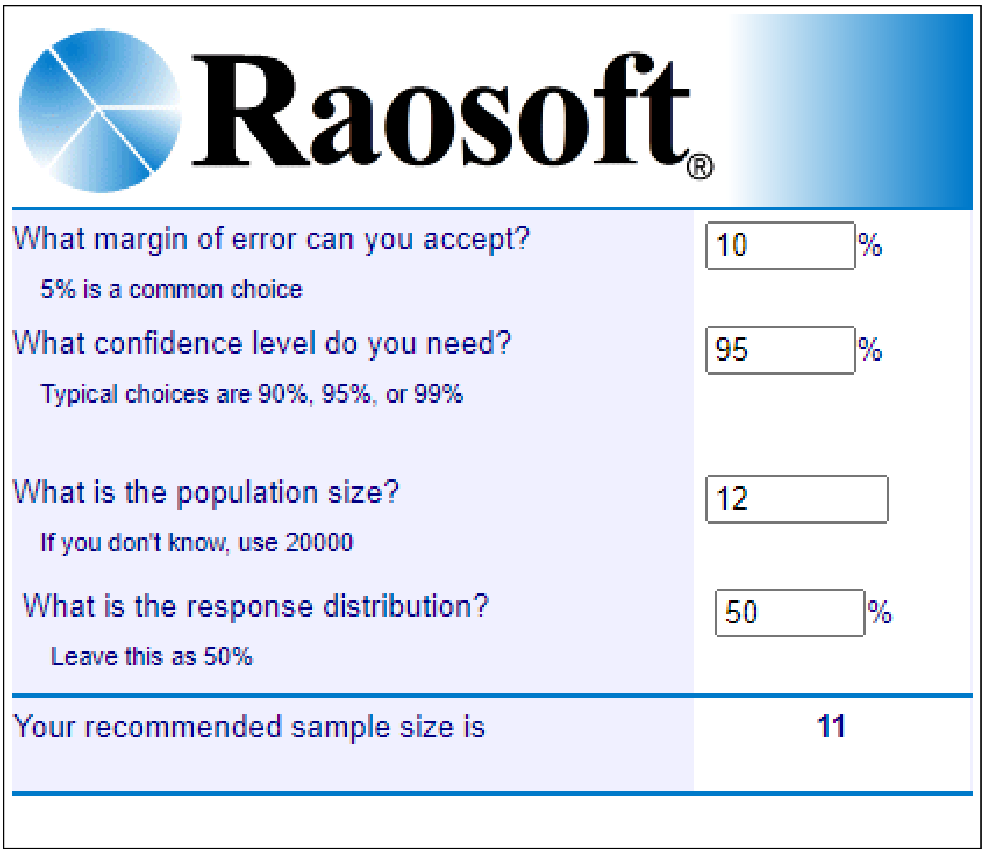


Figure 3.3: *Sample Size Calculator.*

limitations may affect select responses.

The prototype's code, which is based on Visual Studio Code, can be easily obtained by learners with minimal storage space and internet access. This eliminates the necessity of hardware provision for the prototype's execution and mitigates the likelihood of computer component damage. The experimental design aims to be simple and easy to follow, mirroring the abilities of participants with limited familiarity with machine learning. Despite its advantages, this method is limited in terms of efficacy and outcomes. This thesis considers Apple, Amazon, Tesla, Google, and Meta as relevant companies for stock price prediction, despite the availability of additional data, such as macroeconomic indicators and national economic data, that could potentially improve predictions.

3.2.3 Data Collection Tools Structure and Information

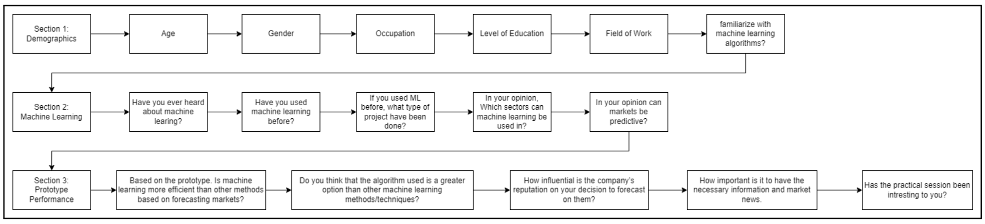


Figure 3.4: *Survey Structure.*

The survey structure utilised in this study, following prototype testing by participants, is illustrated in Figure 3.4. The survey is structured into three distinct sections, each with a specific focus.

Section 1 of the study primarily collects demographic data from the participants and provides contextual information. Section 2 is devoted to assessing the participants' familiarity with machine learning. The primary investigation assesses the participants' acquaintance with machine learning and solicits elaboration on its pragmatic application. The objective of the following inquiries is to examine how participants' familiarity with algorithms may affect their performance during practical sessions. Section three is focused on the participants' feedback regarding the prototype. The first two inquiries assess the prototype's readability, navigability, and comprehensibility. This inquiry seeks to determine whether the prototype provided a comparable experience to the participants' genuine work environments. The following questions relate to the participants' feedback on the recently completed session. The final set of inquiries pertains to the significance of the collected stock market data and the level of involvement exhibited by the participants throughout the session. A structured survey design enables the acquisition of pertinent data regarding participant demographics, machine learning familiarity, and prototype interaction. This methodology provides significant insights for research.

3.2.4 Prototype

The LSTM Neural Networks artificial model was used to develop and predict stock prices for this prototype. It was important to consider previous research regarding the foreign exchange markets. In previous research, an analytical instrument with machine learning algorithm capabilities was utilized.

3.2.5 Development Tools and Libraries

Visual Studio Code is a powerful tool for implementation; however, certain models and algorithms may not be readily available. Thus, multiple software packages were employed and integrated in this study. Multiple libraries were used for this project. The Python script's mathematical functions were made accessible through the utilisation of a mathematical Python model. This study employs the math library model to ascertain the optimal number of training rows. Specifically, a function is developed using 70% of the training data for this purpose. The survey utilised pandas-data reader to retrieve public financial data from Yahoo Finance and import it into a Python data frame. NumPy is a Python library utilised for converting data frames from the panda library into NumPy arrays. The NumPy array data was scaled using Sklearn.preprocessing and MinMaxScaler. Pre-processing techniques such as scaling, or normalisation of input data are beneficial for improving the performance of neural network algorithms. Keras models and layers were employed to ascertain the anticipated input shape. To build and train an LSTM model, the initial layer of a sequential model must be informed of the input geometry.

3.2.6 Implementation of Models

For the purpose of this study, an implementation was designed to estimate stock prices. LSTM architecture was utilized in the neural network model. The implementation was subdivided into numerous aspects, including: models, Conversion of data, visualization.

Phase 1: The Models

In the Anaconda environment, the model was developed. A model was constructed based on the discovery of a model for each phase of the LSTM. Prior to implementation, the models were properly installed to make it simpler for participants to test effectively and reduce the possibility of them being confused by unexpected errors.

Phase 2: Conversion of Data

Subsequently, the data conversion process was executed. The aim of this phase was to transform Yahoo Finance data into a data frame containing the close column. Subsequently, the data frame was converted into a NumPy array to determine the number of rows for training the model. Specifically, 70% of the 1258 records, or 881 records, were used for this purpose. Subsequently, the data should be scaled and partitioned into independent and dependent variables for the purpose of generating a training data set. Specifically, the training data set should include all values of index 0, with the independent variables assigned to the x train variable and the dependent variables assigned to the y train variable. Subsequently, transform the x train and y train into a NumPy array to enable the restructuring of the data into a three-dimensional output configuration, as mandated by LSTM. Subsequently, the testing data set can be generated alongside a novel array that encompasses normalised values from the index. This procedure serves to restructure both the data and the model's projected price values.

Phase 3: Visualization

Subsequently, matplotlib was employed to generate the visualisation. The data is graphically depicted in the code as a bar chart, which displays the daily closing prices alongside their corresponding dates. Upon successful establishment of the neural network, we plotted the data incorporating the near price values, valid result, and predicted price.

3.3 Pilot Testing

Before the final version was made available to the intended audience, pilot testing was conducted to identify and address any flaws or limitations. The following summary details the tests conducted to verify the prototype's core functionality.

Action	Expected	Actual	Pass
Install all libraries	All libraries installed correctly providing the required dependencies.	The libraries were all installed properly	Yes
Get the stock quote for the required stock markets via csv file.	Show the stock markets with the required dates from beginning of 2018 to the end of 2022.	Show the stock markets with the required dates from beginning of 2018 to the end of 2022.	Yes
Data cleaning	Clean the data as to change all attributes in the column to be type float64 and get the number of rows and columns in the dataset.	Clean the data as to change all attributes in the column to be type float64 and get the number of rows and columns in the dataset.	Yes

Visualization of closing price history	Display a working stock chart showing the closed price history.	Display a working stock chart showing the closed price history.	Yes
Create data frame for 'Close' column	Create data frame for 'Close' column and convert the data frame to a NumPy array to get the number of rows to train the model on.	For 'Close' column a data frame was created to convert NumPy array to get the number of rows to train the model	Yes
Create the training data sets.	Split the data into x_training and y_training.	Data split into two different sets.	Yes
Build LSTM model and train the model.	Model to be passed through a neural network and an epoch of 1.	Model passed successfully with epoch test phase of all 822 datasets.	Yes

Get the root mean squared error.	Get the result of the root mean squared error with an accuracy as below as possible.	The result from the root mean squared resulted as 9.	Yes
Plot the data.	Display an output of the data with the training models.	Output the final results of the trained dataset with the validation followed by the predicted value.	Yes

3.4 Errors

Multiple errors were recorded during the prototype's execution. The errors that may occur in programming encompass minor mistakes such as typographical errors, absence of necessary libraries or incompatible versions that could result in logical errors. Additionally, incorrect utilisation of Visual Studio libraries may adversely affect the output. This section presents supplementary details pertaining to errors.

3.4.1 Error 1: Missing Libraries

This error resulted from lacking library files. At the beginning of the project, ”Openyxl” was not discovered for scraping historical data. The cause of the issue was that the kernel version did not support the required API. In order to install it, the logic had to be modified to select a kernel that supported this API.

3.4.2 Error 2: Confident mean error

The observed root mean squared error exceeds the expected value by a substantial margin, reaching 16%. Insufficient training of the machine learning algorithm was addressed by increasing the number of epoch iterations for improved learning and training. The duration of time has transitioned from a single epoch to a span of five epochs. The problem was solved with a root mean squared error below 5%, indicating a higher level of accuracy compared to the previous results.

3.5 Ethical Considerations

This research aims to protect confidential and delicate information. The data will be securely stored on the researcher's hardware with exclusive access granted only to the researcher. The study will ensure participant anonymity by refraining from collecting or storing personal information such as names and addresses, as they are not relevant to the research. The data will be stored for up to one year and then securely disposed of. The data will be used solely for the purposes of the study and participant identities will be kept confidential.

The study guarantees the protection of participants' physical and psychological well-being. The study materials are considered safe and do not carry any risk of harm or damage to the participants. The study environment will be designed with participant safety as a top priority. The study's practical tasks were performed by the participants using a laptop as instructed. The research will be conducted in a manner that prioritises the avoidance of harm to financial institutions. The collected data will only be accessible to the superintendent and examiners with restricted access. Participants have the right to withdraw from the study voluntarily and without justification. Upon withdrawal, all data and records pertaining to the individual will be erased.

The participants have complete control over the data they provide and only allow its use for the specific purposes of this study. The transparency of the data collection process will be ensured, and all transcriptions will be expunged upon completion. The adherence to confidentiality, anonymity, and data protection protocols will be strictly maintained. Upon solicitation, the subjects will receive an electronic version of the investigation. The researcher will provide a clear explanation of the study and respond to any questions participants may have regarding their participation. Participants' privacy and sensitivity will be respected, and all data collected will be used solely for research purposes. Preferential treatment will be provided to individuals with distinct needs or those who associate themselves with specific social groups. Participants are afforded the opportunity to withdraw from the study without providing any rationale.

3.6 Conclusion

This chapter presents the methodologies utilised to achieve the objectives of the thesis. The initial phase entailed the collection of past financial data, as depicted in a data flow diagram that delineates the key procedures. The following step involved the development and arrangement of the experimental models. The final step entailed clarifying the methodology of enhancing forecasts. This chapter will provide an analysis of the experimental findings. The following chapter will provide a comprehensive evaluation of the metrics utilised and the outcomes achieved.

Chapter 4: Analysis of Results and Discussion

4.1 Introduction

This chapter outlines the results of LSTM analysis for individual companies, followed by an evaluation of stock market forecasting. The objective is to assess the potential enhancement of the prototype's performance through the implementation of LSTM. The thesis will provide an overview of the performance of each evaluation technique, accompanied by a visual depiction of the observed and forecasted prices.

This chapter consists of six sections. The first section provides a comprehensive description of the metrics utilised to produce the results. The following sections present the outcomes of the implemented methodology and prototype. The conclusion of the study draws comparisons between the findings and prior literature and related investigations.

4.2 Metrics

The stock prediction field is typically divided into two primary domains: regression and classification. Regression models are utilised to predict observed values, with the future price often being used as the response variable. Classification models are utilised to predict the direction of stock prices, specifically, whether they will rise or fall.

The root mean square error (RMSE) is a statistical measure utilised in regres-

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

Figure 4.1: *Root-Mean-Square Error*

sion analysis to assess the extent of deviation of prediction errors. It measures the amount by which predicted values deviate from the actual regression line. The RMSE is computed by squaring the difference between the predicted and actual values, dividing the sum of these squared differences by the sample size (N), and taking the square root of the resulting value. The RMSE is computed by taking the square root of the mean.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Figure 4.2: *Mean Squared Error*

Figure 4.2: Mean Squared Error The Mean Squared Error (MSE) or Mean Squared Deviation (MSD) is a statistical metric that calculates the average of the squared differences between predicted and observed values. The metric offers an assessment of the typical discrepancy or deviation of an estimator. The Mean Squared Error (MSE) is a risk function that quantifies the anticipated value of the squared error loss. The efficacy is deemed superior when the values are lower,

with values approaching zero being the most favourable. The Mean Squared Error (MSE) evaluates the accuracy of an estimator by accounting for both its variance and bias, as it measures the second moment of the error with respect to the origin.

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{E_t - A_t}{A_t} \right|$$

Figure 4.3: *Mean Absolute Percentage Error*

The mean absolute percentage error (MAPE) quantifies the relative deviation between predicted and actual data in percentage terms. The model's efficacy is assessed by measuring it against actual data using a computed metric.

4.3 Data analysis method for online survey

This research utilised various analytical methods to investigate survey data and achieve a comprehensive understanding of the forecasting process in multiple markets, with participation from a diverse range of individuals. The research utilised IBM SPSS software for statistical analysis, which provides various statistical tools. Statistical analyses, such as Chi-Squared tests, T-tests, and One-Way ANOVA tests, were performed using the IBM SPSS software. The collected online survey data underwent descriptive analysis through the utilisation of Google Forms.

4.4 Analysis and Discussion of the online survey

Upon completion of the prototype, an online survey was conducted to collect data on participant demographics, familiarity with machine learning, usage of machine learning, and feedback regarding the prototype. The data collection process involved the use of Google Forms and its subsequent importation into IBM SPSS software, as described in section 4.2.

4.4.1 Demographics

A sample size of 12 participants was deemed sufficient to achieve a representative sample with a 10% margin of error. This sample was subsequently subjected to both descriptive and statistical analyses. This study's demographic data collection includes age, gender, occupation, level of education, and familiarity with machine learning. The pie chart depicted in Figure 1 indicates that the age variable was comprehensively represented, as responses were obtained from all age cohorts. The majority of participants fell within the age range of 21 to 25, accounting for 43.8% (7 participants). The age bracket of 16-20 years old accounted for the second largest cohort of participants, representing 18.8% (3 individuals). The third cohort comprising individuals aged 26-30 years and those aged 40 years and above exhibit similar outcomes, with a combined frequency of 12.5% and a sample size of 2 participants. The age categories of 31-35 years and 36-40 years had the lowest representation among the respondents, with 6.3% and 1 response, respectively, out of the entire sample. The inclusion of respondents from diverse age groups enhances the representativeness of the sample and provides a range of age-related viewpoints to the study. 75% (12 participants) of the total number

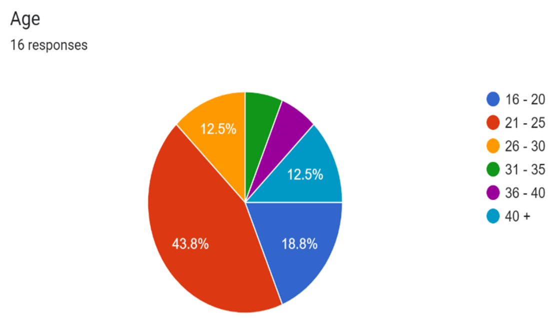


Figure 4.4: *Descriptive statistics Analysis on Age*

of participants were male, as shown in Figure 4.4. 25% (4 participants) of the total responses were provided by females.

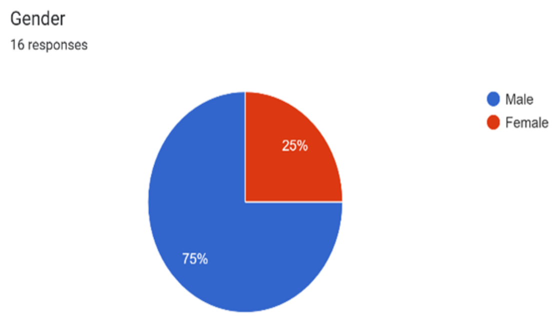


Figure 4.5: *Descriptive Statistics on Gender*

Figure 4.5 illustrates the occupations of the participants. 50% of the participants (8 individuals) are engaged in full-time employment, as indicated by the preponderance of responses. The cohort of full-time students constitutes 37.5% (n=6) of the overall sample. Subsequently, both part-time employment and part-time student categories exhibit similar response rates, amounting to 6.3% (1 response) each.

Figure 4.6 illustrates the educational attainment of the participants. The majority of participants, specifically 56.3% (9 individuals), have completed a bachelor's

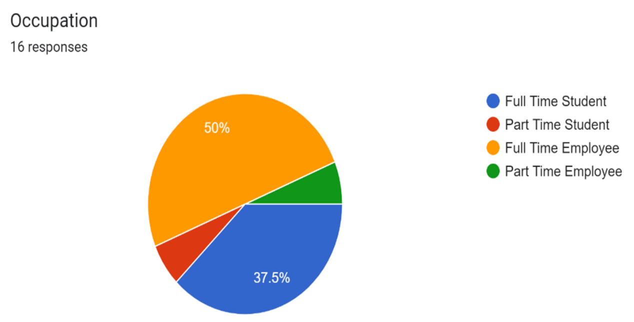


Figure 4.6: *Descriptive Statistic on occupation*

degree as their highest educational attainment. The cohort of respondents who possess postsecondary education as their highest level of attainment comprises the runner-up group, which accounts for 25% (4 responses) of the entire respondent pool. The highest level of education for 18.8% (3 responses) of participants was a master's degree. The data suggests that participants were recruited from various educational backgrounds, excluding individuals with a doctoral degree.

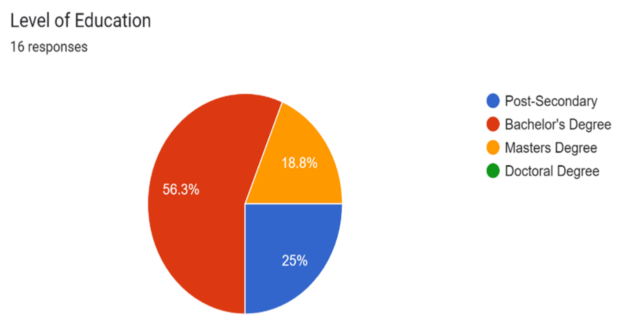


Figure 4.7: *Descriptive statistics Education*

Figure 4.7's demographics section concludes with a description of participants' familiarity with machine learning. The majority of participants, comprising 81.3% (n=13), possess knowledge of machine learning, whereas a minority of 18.8% (n=3) lack familiarity with the subject.

4.4.2 T-Tests

The online survey data underwent analysis through IBM SPSS. T-tests were employed to conduct a statistical analysis of the relationship between the variable under consideration and gender.

Table 1 presents the outcomes of an independent samples t-test examining the association between education level and the conviction that markets can be anticipated. The p-value for the two-tailed tests presented in the table is 0.115. This signifies the acceptance of the null hypothesis. The influence of education level on the belief in the predictability of markets is not significant.

Group Statistics

	Level of Education	N	Mean	Std. Deviation	Std. Error Mean
In your opinion can markets be predictive?	1	4	2.75	.957	.479
	2	9	3.67	.866	.289

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
In your opinion can markets be predictive?	Equal variances assumed	.085	.777	-1.710	11	.058	.115	-.917	.536	-2.096	.263
	Equal variances not assumed			-1.640	5.315	.079	.159	-.917	.559	-2.328	.495

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
In your opinion can markets be predictive?	Cohen's d	.892	-1.028	-2.259	.245
	Hedges' correction	.959	-.956	-2.101	.227
	Glass's delta	.866	-1.058	-2.315	.252

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 4.8: *T-test analysing level of education and markets predictiveness*

Figure 4.8 presents the results of an independent sample t-test conducted to examine the relationship between participants' age and the effectiveness of the prototype in predicting market outcomes. The statistical analysis conducted at a 90% level of significance yielded a two-tailed significance value of 0.69, indicating a significant impact of the prototype's age on its performance. The acceptance of the null hypothesis is based on the statistical significance of the result.

Group Statistics

	Age	N	Mean	Std. Deviation	Std. Error Mean
Based on the prototype, is machine learning more efficient than other methods based on forecasting markets?	1	3	3.67	.577	.333
	2	7	3.86	.690	.261

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Based on the prototype, is machine learning more efficient than other methods based on forecasting markets?	Equal variances assumed	.028	.872	-.416	8	.344	.688	-.190	.458	-1.247	.866
	Equal variances not assumed			-.450	4.621	.337	.673	-.190	.423	-1.306	.925

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Based on the prototype, is machine learning more efficient than other methods based on forecasting markets?	Cohen's d	.664	-.287	-1.638	1.081
	Hedges' correction	.735	-.259	-1.478	.976
	Glass's delta	.690	-.276	-1.626	1.096

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 4.9: *T-test analysing age and the efficiency of the prototypes machine learning method*

Figure 4.9 presents the results of a survey that inquired about the participants' perception of the effectiveness of the utilised algorithm in comparison to other machine learning methods. Given that this is an independent samples t-test, the demographic characteristics of each gender were taken into account. The obtained two-tailed significance value of 0.417 indicates a statistically significant difference. Thus, the null hypothesis can be accepted. A correlation has been observed between gender and the perception of the optimal machine learning approach.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Do you think that the algorithm used is a greater option than other machine learning methods/techniques?	1	12	3.83	.577	.167
	2	4	3.50	1.000	.500

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Do you think that the algorithm used is a greater option than other machine learning methods/techniques?	Equal variances assumed	2.000	.179	.837	14	.208	.417	.333	.398	-.521	1.188
	Equal variances not assumed			.632	3.691	.282	.564	.333	.527	-1.180	1.846

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Do you think that the algorithm used is a greater option than other machine learning methods/techniques?	Cohen's d	.690	.483	-.671	1.620
	Hedges' correction	.730	.457	-.634	1.531
	Glass's delta	1.000	.333	-.851	1.468

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 4.10: *T-test between gender and their opinion on the machine learning method used.*

In Table 4, participants were asked to rate the impact of the company's reputation on forecasting decisions. At a substantial value of 0.629, gender does not appear to affect the reputation of the company. As a result, the null hypothesis can be adopted, as the obtained result does not indicate a statistically significant difference.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
How influential is the companies repuation on your decision to forecast?	1	12	4.00	.739	.213
	2	4	3.75	1.258	.629

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
How influential is the companies repuation on your decision to forecast?	Equal variances assumed	1.939	.185	.494	14	.314	.629	.250	.506	-.835	1.335
	Equal variances not assumed			.376	3.715	.364	.727	.250	.664	-1.652	2.152

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
How influential is the companies repuation on your decision to forecast?	Cohen's d	.876	.285	-.856	1.417
	Hedges' correction	.927	.270	-.809	1.339
	Glass's delta	1.258	.199	-.958	1.325

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 4.11: *T-test between gender and the company's reputation to forecast.*

The statistical significance of stock market information in relation to the gender of participants was assessed through a t-test on an independent sample, as shown in Table 5. The insignificance of the obtained two-tailed significance value of 0.475 indicates the acceptance of the null hypothesis, as there is no significant difference. This implies that gender does not affect the worth of information.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
How important is it to have the necessary information and market news.	1	12	4.33	.492	.142
	2	4	4.00	1.414	.707

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p			Lower	Upper
How important is it to have the necessary information and market news.	Equal variances assumed	5.645	.032	.734	14	.238	.475	.333	.454	-.641	1.308
	Equal variances not assumed			.462	3.246	.337	.673	.333	.721	-1.867	2.533

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
How important is it to have the necessary information and market news.	Cohen's d	.787	.424	-.726	1.559
	Hedges' correction	.832	.400	-.686	1.473
	Glass's delta	1.414	.236	-.928	1.363

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 4.12: *T-test between gender and the importance to have the necessary information on stock markets.*

The present study poses a question regarding the participants' level of enthusiasm towards the practical session, as reflected in the ultimate table. This is analysed in relation to the participants' gender. The insignificance of the two-tailed significance value (p = 0.54) indicates that the null hypothesis cannot be rejected, thereby implying that there is no significant difference between the two groups.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Has the practial session been intresting to you?	1	12	4.33	.985	.284
	2	4	4.00	.816	.408

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Has the practial session been intresting to you?	Equal variances assumed	.735	.406	.607	14	.277	.554	.333	.549	-.845	1.511
	Equal variances not assumed			.670	6.216	.263	.527	.333	.497	-.874	1.540

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
Has the practial session been intresting to you?	Cohen's d	.951	.350	-.795	1.483
	Hedges' correction	1.006	.331	-.751	1.402
	Glass's delta	.816	.408	-.795	1.552

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 4.13: *T-test between gender and their interest during the practical session*

4.4.3 Chi-Squared Tests

The data collected from the online survey was analysed using IBM SPSS to compare variables and perform cross tabulation. Table 7 displays the statistical analysis of the cross-tabulation between gender and prior usage of machine learning. The study's results indicate that most respondents have prior experience with machine learning.

Gender * Have you used machine learning before? Crosstabulation

			Have you used machine learning before?		
			No	Yes	Total
Gender	1	Count	4	8	12
		% within Gender	33.3%	66.7%	100.0%
		% within Have you used machine learning before?	66.7%	80.0%	75.0%
		% of Total	25.0%	50.0%	75.0%
	2	Count	2	2	4
		% within Gender	50.0%	50.0%	100.0%
		% within Have you used machine learning before?	33.3%	20.0%	25.0%
		% of Total	12.5%	12.5%	25.0%
Total		Count	6	10	16
		% within Gender	37.5%	62.5%	100.0%
		% within Have you used machine learning before?	100.0%	100.0%	100.0%
		% of Total	37.5%	62.5%	100.0%

Figure 4.14: *Cross tabulation between gender and experience with machine learning*

After conducting cross-tabulation, the researcher performed statistical analysis between gender and machine learning experience. The Pearson Chi-Squared displayed in Table 8 has a p-value of 0.551. This indicates that the null hypothesis indicating no significant difference between gender and prior machine learning experience can be adopted.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.356[a]	1	.551		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.349	1	.555		
Fisher's Exact Test				.604	.489
N of Valid Cases	16				

a. 3 cells (75.0%) have expected count less than 5. The minimum expected count is 1.50.

b. Computed only for a 2x2 table

Figure 4.15: *Pearson Chi-Squared test between gender and experience with machine learning*

The cross-tabulation analysis between education level and machine learning usage is displayed in Table 9 below. Almost every level seems to have utilised machine learning in the past, with a few exceptions.

Level of Education * Have you used machine learning before? Crosstabulation

			Have you used machine learning before? No	Have you used machine learning before? Yes	Total
Level of Education	1	Count	2	2	4
		% within Level of Education	50.0%	50.0%	100.0%
		% within Have you used machine learning before?	33.3%	20.0%	25.0%
		% of Total	12.5%	12.5%	25.0%
	2	Count	3	6	9
		% within Level of Education	33.3%	66.7%	100.0%
		% within Have you used machine learning before?	50.0%	60.0%	56.3%
		% of Total	18.8%	37.5%	56.3%
	3	Count	1	2	3
		% within Level of Education	33.3%	66.7%	100.0%
		% within Have you used machine learning before?	16.7%	20.0%	18.8%
		% of Total	6.3%	12.5%	18.8%
Total		Count	6	10	16
		% within Level of Education	37.5%	62.5%	100.0%
		% within Have you used machine learning before?	100.0%	100.0%	100.0%
		% of Total	37.5%	62.5%	100.0%

Figure 4.16: *cross tabulation between level of education and experience with machine learning*

Table 9 involved cross-tabulation analysis, which was subsequently followed by a chi-squared test as presented in Table 10. This study examines the correlation between educational attainment and proficiency in machine learning. The statistical analysis reveals a p-value of 0.84, suggesting a lack of association between the utilisation of machine learning and the level of education. Thus, the null hypothesis may be accepted, suggesting that there is no significant impact of machine learning experience on educational attainment.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.356[a]	2	.837
Likelihood Ratio	.349	2	.840
N of Valid Cases	16		

a. 5 cells (83.3%) have expected count less than 5. The minimum expected count is 1.13.

Figure 4.17: *Person Chi-Squared test between level of education and experience with ML*

Table 11 displays the cross-tabulation of respondents' familiarity with machine learning and the relative effectiveness of the employed machine learning method compared to another method. Consensus among those knowledgeable in machine learning affirms the accuracy of the approach employed for stock market prediction.

Based on the prototype. Is machine learning more efficient than other methods based on forecasting markets? * Are you familiar with machine learning? Crosstabulation

			Are you familiar with machine learning?		
			No	Yes	Total
Based on the prototype. Is machine learning more efficient than other methods based on forecasting markets?	3	Count	2	3	5
		% within Based on the prototype. Is machine learning more efficient than other methods based on forecasting markets?	40.0%	60.0%	100.0%
		% within Are you familiar with machine learning?	66.7%	23.1%	31.3%
		% of Total	12.5%	18.8%	31.3%
	4	Count	1	8	9
		% within Based on the prototype. Is machine learning more efficient than other methods based on forecasting markets?	11.1%	88.9%	100.0%
		% within Are you familiar with machine learning?	33.3%	61.5%	56.3%
		% of Total	6.3%	50.0%	56.3%
	5	Count	0	2	2
		% within Based on the prototype. Is machine learning more efficient than other methods based on forecasting markets?	0.0%	100.0%	100.0%
		% within Are you familiar with machine learning?	0.0%	15.4%	12.5%
		% of Total	0.0%	12.5%	12.5%
Total		Count	3	13	16
		% within Based on the prototype. Is machine learning more efficient than other methods based on forecasting markets?	18.8%	81.3%	100.0%
		% within Are you familiar with machine learning?	100.0%	100.0%	100.0%
		% of Total	18.8%	81.3%	100.0%

Figure 4.18: *Cross tabulation between familiarities with machine learning and the chosen model*

The null hypothesis can be refuted because the chi-square test in table 12 produced a significant value of 0.32, indicating that there is a significant difference. This indicates that the method has an effect on machine learning-savvy users.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.288[a]	2	.318
Likelihood Ratio	2.433	2	.296
N of Valid Cases	16		

a. 5 cells (83.3%) have expected count less than 5. The minimum expected count is .38.

Figure 4.19: *Pearson Chi-Squared test*

4.4.4 Anova Test

Table 13 employs One-way ANOVA testing to analyse the relationship between participants' ages and their perspectives on the predictability of the market. The obtained result of 0.88 is statistically significant, suggesting that there is no association between age and the perception of market predictability. Thus, the null hypothesis is accepted.

ANOVA

In your opinion can markets be predictive?

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.723	5	.345	.337	.879
Within Groups	10.214	10	1.021		
Total	11.938	15			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
In your opinion can markets be predictive?	Eta-squared	.144	.000	.217
	Epsilon-squared	-.283	-.500	-.175
	Omega-squared Fixed-effect	-.261	-.455	-.162
	Omega-squared Random-effect	-.043	-.067	-.029

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Figure 4.20: *One-Way anova testing between age and market predictive opinion*

Table 14 presents the outcomes of the one-way ANOVA analysis. This study incorporated education level as a demographic variable and investigated participants' perspectives on the predictability of stock markets through statistical analysis. The obtained significance value of 0.21 suggests the absence of a significant association between education level and opinions. Thus, the adoption of the null hypothesis is warranted.

ANOVA

In your opinion can markets be predictive?

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.521	2	1.260	1.740	.214
Within Groups	9.417	13	.724		
Total	11.937	15			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	Upper
In your opinion can markets be predictive?	Eta-squared	.211	.000	.466
	Epsilon-squared	.090	-.154	.384
	Omega-squared Fixed-effect	.085	-.143	.369
	Omega-squared Random-effect	.044	-.067	.226

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Figure 4.21: *one-way anova testing between level of education and market predictive opinion*

Subsequently, the participants were queried about the importance of obtaining access to essential information and market updates. Table 15 presents the results of a one-way ANOVA test examining the relationship between participants' education level and their responses. The null hypothesis is accepted at a significance level of 0.72, suggesting that education level remains unaffected.

ANOVA

How important is it to have the necessary information and market news.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.444	2	.222	.338	.720
Within Groups	8.556	13	.658		
Total	9.000	15			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
How important is it to have the necessary information and market news.	Eta-squared	.049	.000	.266
	Epsilon-squared	-.097	-.154	.153
	Omega-squared Fixed-effect	-.090	-.143	.145
	Omega-squared Random-effect	-.043	-.067	.078

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Figure 4.22: *One-way anova between level of education and importance of market information*

4.5 Prototype Results

4.5.1 LSTM Network Performance

The LSTM model employed the closing prices of AAPL, AMZN, META, GOOG, and TSLA to generate the subsequent results. The prevalent method for model analysis entails segregating the data into training and testing subsets, with a proportion of 70% and 30%, correspondingly. The study generated results through four programme runs, each with unique epoch and batch size settings. The results were used to calculate the average performance of the model.

AAPL

Configuration	Epochs	Batch	RMSE	MSE	MAPE	Accuracy
1	100	10	4.593	21.099	2.23%	97.77%
2	100	15	3.661	13.833	1.83%	98.17%
3	200	10	4.078	16.628	2.08%	97.92%
4	200	15	4.632	21.455	2.40%	97.60%
Average performance:			4.241	18.254	2.14%	97.87%

Table 4.1: *APPL LSTM Results*

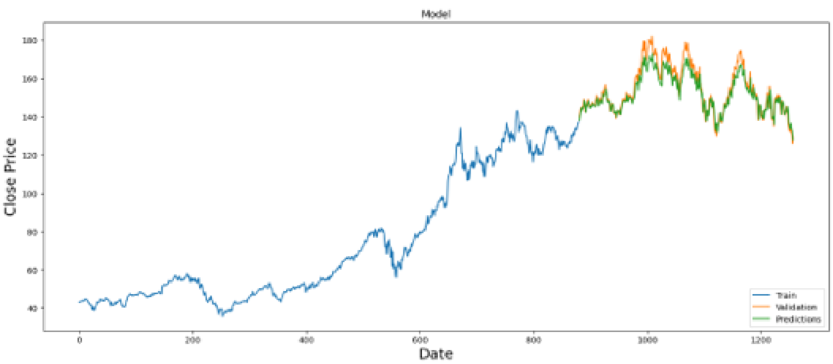


Figure 4.23: *AAPL LSTM 100 Epochs 10 Batch Size*

newpage

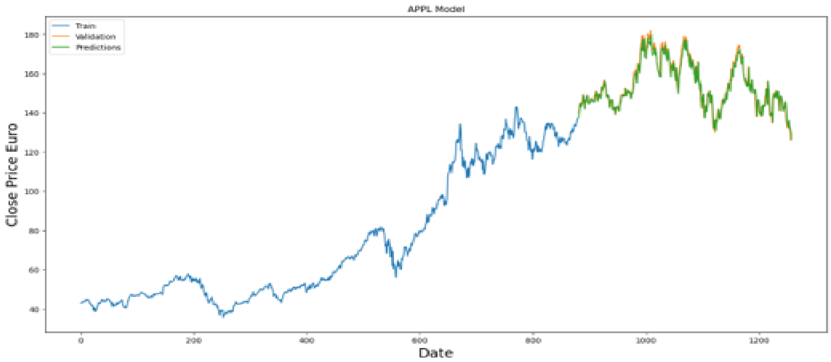


Figure 4.24: *AAPL LSTM 100 Epochs 15 Batch Size*

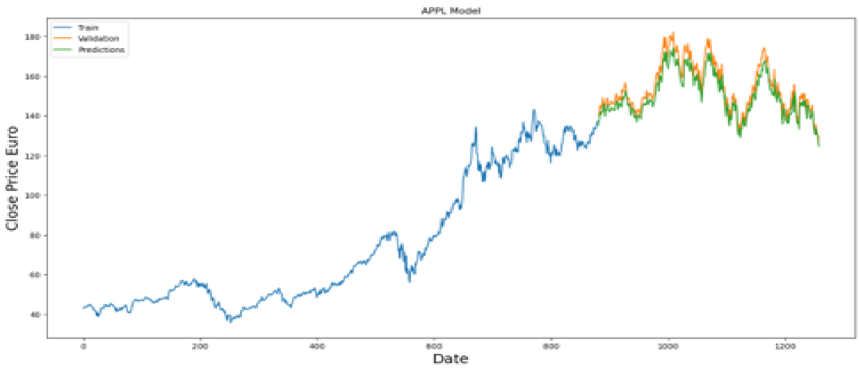


Figure 4.25: *AAPL LSTM 200 Epochs 10 Batch Size*

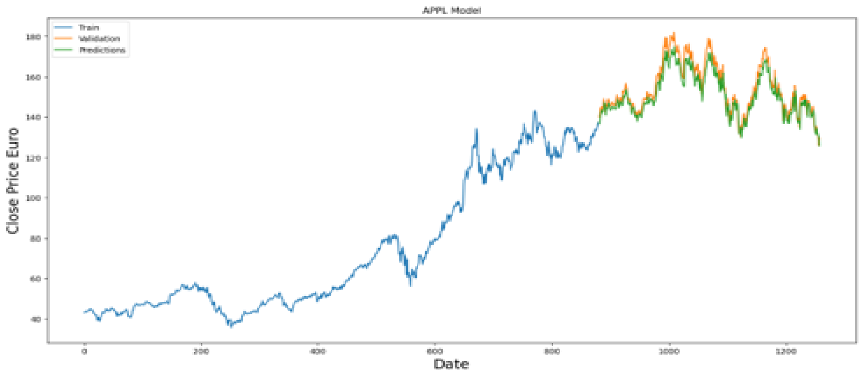


Figure 4.26: *AAPL LSTM 200 Epochs 15 Batch Size*

The model exhibited outstanding precision and efficacy across all test setups, attaining outcomes ranging from 97% to 98%. According to Table 4.1, the optimal configuration for achieving the highest accuracy and lowest error values involved a batch size of 15 and 100 epochs. The setup described above attained a precision level of 98.17%. The corresponding error metrics were: RMSE of 3.661, MSE of 13.833, and MAPE of 1.83%. Surprisingly, this configuration produced superior outcomes despite a decreased number of epochs and a smaller sample size. Configuration 3, as depicted in Figure 4.6, demonstrates the model's proficiency in precisely predicting fluctuations in the maximum closing prices. Although Configuration 3 achieved a high accuracy score of 97.92%, it displayed a slightly increased error rate. The RMSE, MSE, and MAPE error metrics were measured at 4.078, 16.628, and 2.08%, correspondingly. Encouraging results were noted in configurations 1 and 4, where 100 and 200 epochs were utilised with batch sizes of 10 and 15, respectively. Configuration 1 demonstrated a 97.77% accuracy, while Configuration 4 achieved a marginally lower accuracy of 97.60%. Figures 4.3 and 4.6 exhibit the LSTM model's accurate prediction of the major fluctuations in closing prices for configurations 1 and 4. The configurations obtained from the historical data of AAPL demonstrated accuracy rates surpassing 90%, which is a remarkable result.

AMZN

Configuration	Epochs	Batch	RMSE	MSE	MAPE	Accuracy
1	100	10	4.3.355	18.970	2.42%	97.58%
2	100	15	3.845	14,787	2.15%	97.85%
3	200	10	3.863	17.721	2.12%	97.88%
4	200	15	3.918	15.215	2.20%	97.80%
Average performance:			3.989	16.673	2.22%	97.78%

Table 4.2: *AMZN LSTM Results*

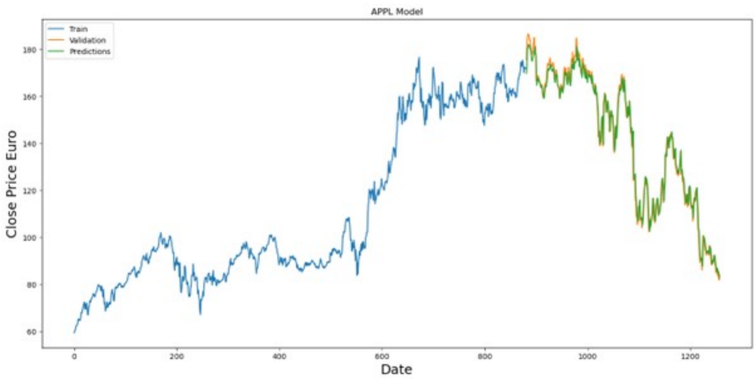


Figure 4.27: *AMZN LSTM Network Performance with 100 Epochs and Batch Size 10*

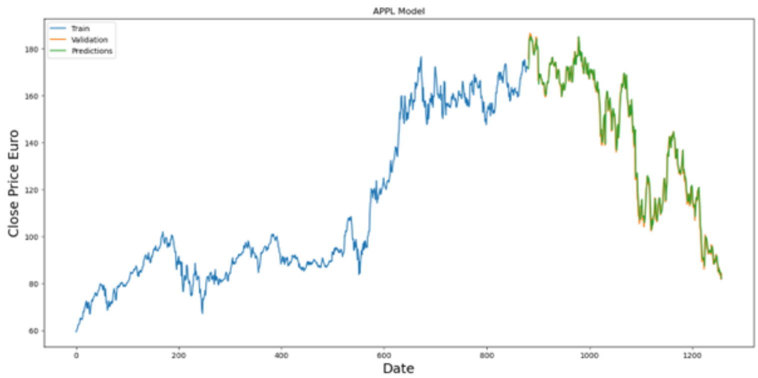


Figure 4.28: *AMZN LSTM Network Performance with 100 epochs and Batch size 15*

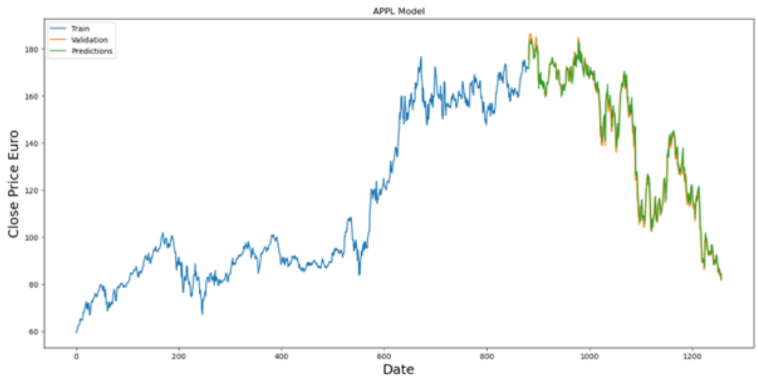


Figure 4.29: *AMZN LSTM Network Performance with 200 Epochs and Batch Size 10*

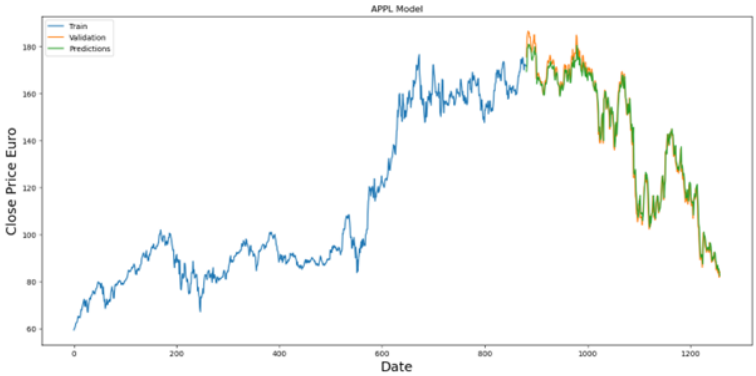


Figure 4.30: *AMZN LSTM Network Performance with 200 Epochs and Batch Size 15*

The model demonstrated a noteworthy level of performance when tested with AMZN's dataset, producing an average accuracy of 97.78%. The findings in Table 4.2 and Figure 4.9 demonstrate that Configuration 3, which employed 200 epochs and a batch size of 10, demonstrated higher accuracy and reduced error values. The graph illustrates the model's accurate forecast of the most notable price fluctuations. Configuration 3 attained a high level of accuracy, registering a 97.88% accuracy rate, and exhibited low error metrics, such as an RMSE of 3.836, MSE of 17.721, and MAPE of 2.10%. Similar to the case of AAPL, forecasting AMZN using an LSTM model required a larger number of epochs and a smaller sample size. It is worth mentioning that the three remaining configurations also produced accurate predictions, with a 97% precision rate across all configurations. As per the data presented in Figure 4.8, Configuration 2 attained a score of 97.85% and demonstrated error rates of 3.845 RMSE, 14.787 MSE, and 2.15% MAPE, thereby positioning it as the second most precise alternative. Configurations 4 and 1, utilising 200 and 100 epochs respectively, with batch sizes of 15 and 10 respectively, produced outstanding outcomes. The LSTM model exhibited precise forecasting of significant fluctuations in the closing price, as depicted in Figures 4.7-4.10. Configuration 3 demonstrated the highest level of accuracy among all configurations, achieving 97.88%.

META

Configuration	Epochs	Batch	RMSE	MSE	MAPE	Accuracy
1	100	10	8.894	79.115	2.93%	97.07%
2	100	15	7.803	60.890	2.51%	97.49%
3	200	10	8.523	72.655	3.50%	96.50%
4	200	15	7.771	60.395	2.45%	97.55%
Average performance:			8.247	68.264	2.85%	97.15%

Table 4.3: *META LSTM Table*

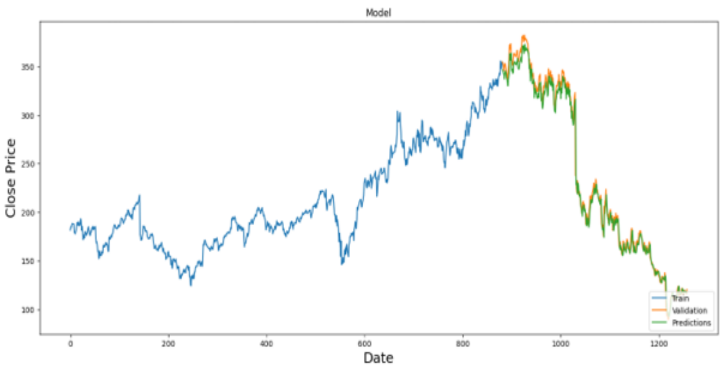


Figure 4.31: *META Network Performance with 100 Epochs and Batch Size 10*

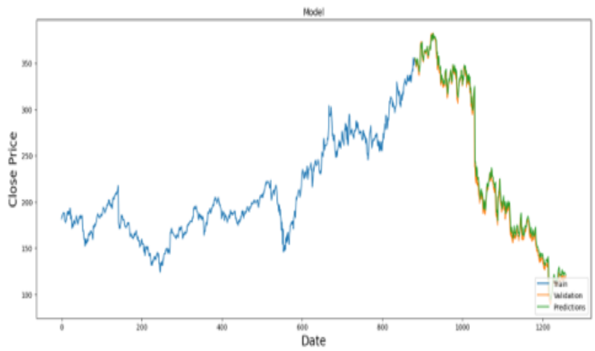


Figure 4.32: META Network Performance with 100 Epochs and Batch Size 15

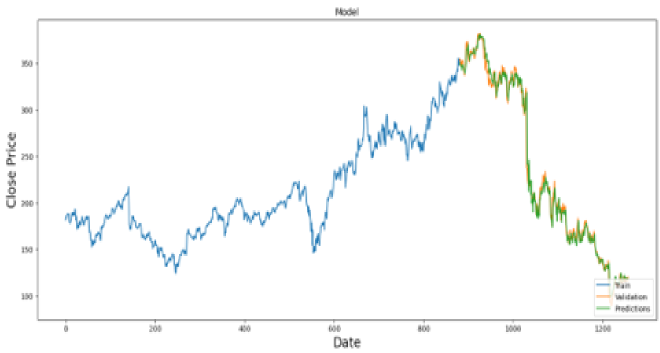


Figure 4.33: META Network Performance with 200 Epochs and Batch Size 10

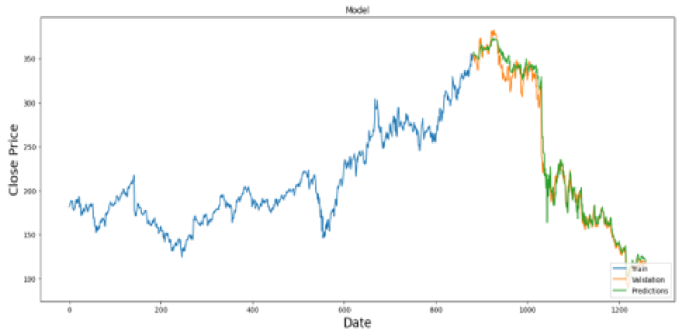


Figure 4.34: *META Network Performance with 200 Epochs and Batch Size 15*

The LSTM model demonstrated impressive performance when applied to META data, achieving an average accuracy of 97.15%. The results presented in Table 4.3 and Figure 4.15 indicate that configuration 4, utilising 200 epochs and a batch size of 15, yielded the most favourable outcomes in terms of accuracy and error metrics. The model exhibited prediction challenges in forecasting closing values, as depicted in the graph. Configuration 4 achieved an accuracy of 97.55%, as evidenced by its RMSE of 7.771, MSE of 60.395, and MAPE of 2.45%. The LSTM model necessitated a greater number of epochs and a larger sample size to achieve accurate META prediction. Nonetheless, the remaining three arrangements also produced precise forecasts, with a range of 96% to 97%. The second-best accuracy configuration achieved 97.49%, with slightly higher error rates of 7.803 RMSE, 72.655 MSE, and 2.51% MAPE, as shown in Figure 4.12. Both setups yielded excellent outcomes. Configuration 1 and Configuration 3 differed in their training parameters, with Configuration 1 utilising 100 epochs and a batch size of 10, and Configuration 3 utilising 200 epochs and a batch size of 10. The findings depicted in Figure 4.15 indicate that Configuration 4 encountered challenges in accurately forecasting the minimum values of closing price fluctuations. Figures 4.11, 4.12, and 4.13 illustrate configurations 1, 2, and 3, respectively, which demonstrate the efficacy of LSTM in predicting significant fluctuations in close prices. The comparison of the three configurations reveals that configuration 2 exhibited superior performance, achieving an accuracy rate of 97.49%.

Configuration	**Epochs**	**Batch**	**RMSE**	**MSE**	**MAPE**	**Accuracy**
1	100	10	3.210	10.304	2.18%	97.82%
2	100	15	3.623	13.129	2.48%	97.52%
3	200	10	3.482	12.124	2.24%	97.76%
4	200	15	2.563	6.572	1.66%	98.34%
Average performance:			3.2195	10.532	7.32%	97.86%

Table 4.4: *GOOG LSTM Results*

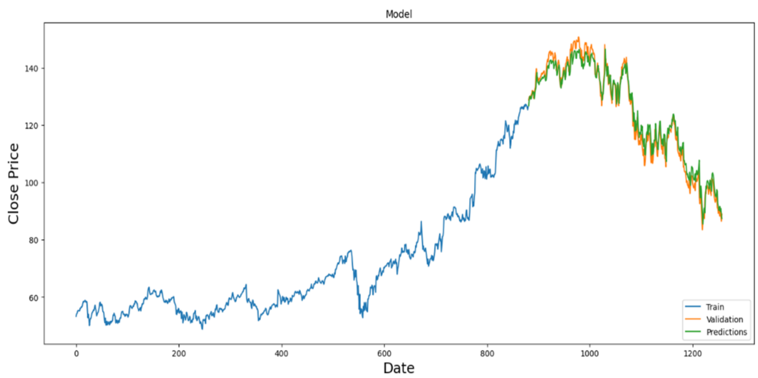


Figure 4.35: *GOOG LSTM Network Performance with 100 Epochs and Batch Size 10*

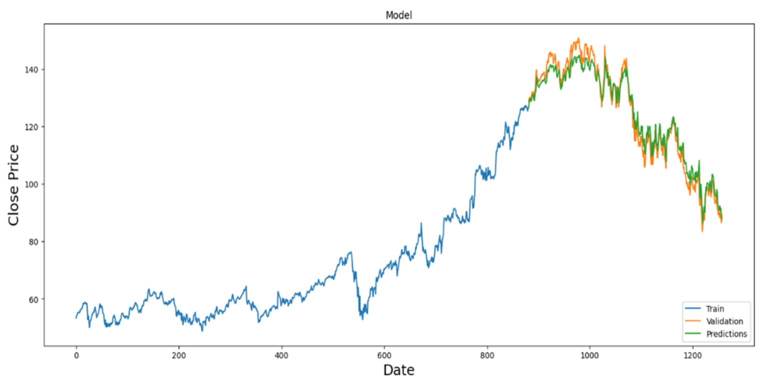


Figure 4.36: *GOOG LSTM Network Performance with 100 Epochs and Batch Size 15*

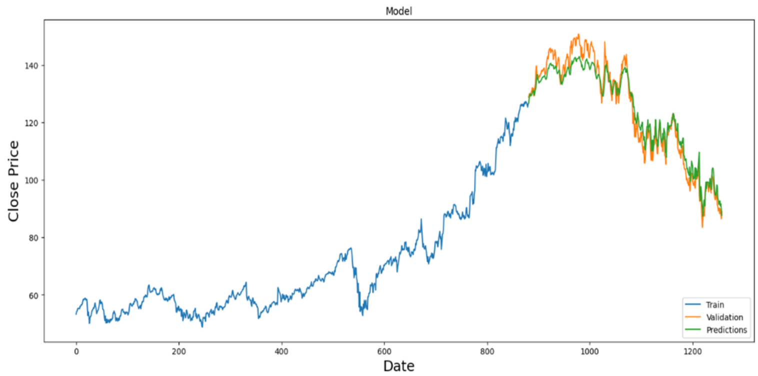


Figure 4.37: *GOOG LSTM Network Performance with 200 Epochs and Batch Size 10*

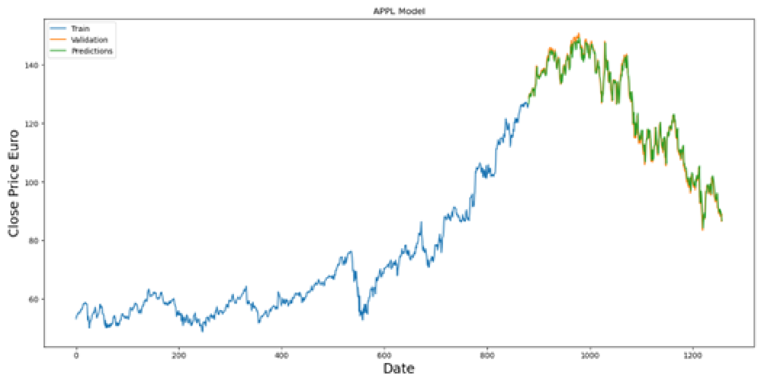


Figure 4.38: *GOOG LSTM Network Performance with 200 Epochs and Batch Size 15*

The GOOG dataset was subjected to LSTM modelling, which yielded exceptional results with an average precision of 97.86%. Table 4.4 and Figure 4.19 demonstrate that Configuration 4 outperformed other configurations in accuracy and error values, utilising a batch size of 15 and 200 epochs. The graph illustrates the accurate predictions of the model regarding the major price fluctuations. Configuration 4 attained a high level of accuracy, registering a 98.34% accuracy rate, and exhibited low error metrics, such as an RMSE of 2.563, MSE of 6.563, and MAPE of 1.66%. The results suggest that, similar to the results observed in the case of AMZN, predicting GOOG using LSTM required a higher number of epochs and a larger sample size. However, the three remaining configurations demonstrated accurate predictions, with accuracy rates ranging from 97% to 98%. The configuration with the second-highest accuracy, depicted in Figure 4.16, attained an accuracy of 97.82%. Nevertheless, it exhibited marginally elevated error rates, namely 3.210 RMSE, 10.304 MSE, and 2.18% MAPE. Both configurations produced favourable results. Configuration 2 underwent 100 epochs of training with a batch size of 15, while Configuration 3 was trained for 200 epochs with a batch size of 10. Figures 4.17 and 4.18 demonstrate that configurations 2 and

3 encountered challenges in precisely predicting the minimum points of the close price oscillations. Configuration 3 outperformed the other configuration by a significant margin of 97.76%.

TSLA

Configuration	Epochs	Batch	RMSE	MSE	MAPE	Accuracy
1	100	10	11.087	122.928	3.16%	96.84%
2	100	15	11.069	122.520	3.00%	97%
3	200	10	16.889	285.247	4.47%	95.53%
4	200	15	13.480	181.720	3.65%	96.35%
Average performance:			13.131	178.104	3.57%	96.43%

Table 4.5: *TSLA LSTM Results*

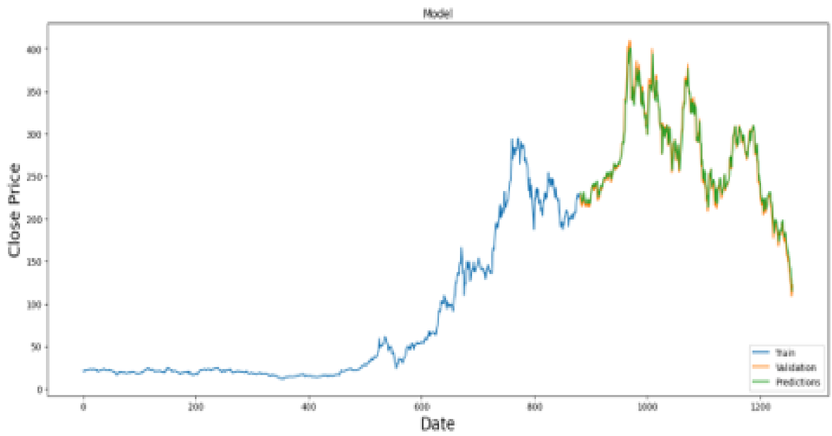


Figure 4.39: *TSLA LSTM Network Performance with 100 Epochs and Batch Size 10*

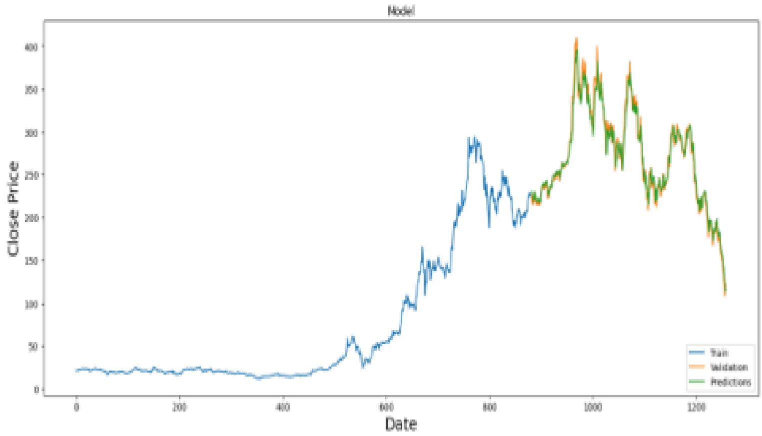


Figure 4.40: *TSLA LSTM Network Performance with 100 Epochs and Batch Size 15*

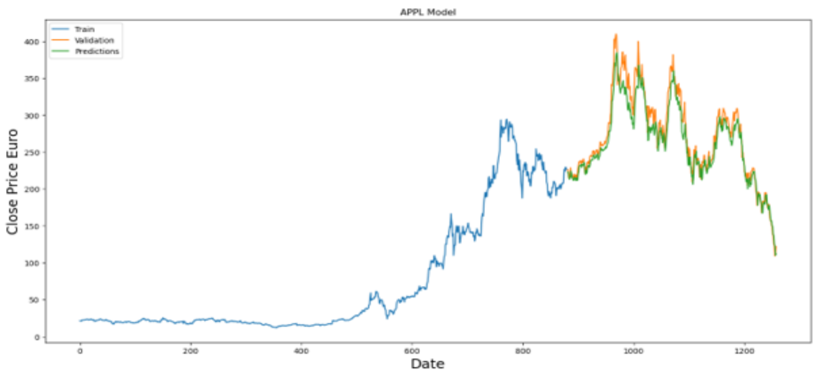


Figure 4.22 TSLA LSTM Network Performance with 200 Epochs and Batch Size 10

Figure 4.41: *TSLA LSTM Network Performance with 200 Epochs and Batch Size 10*

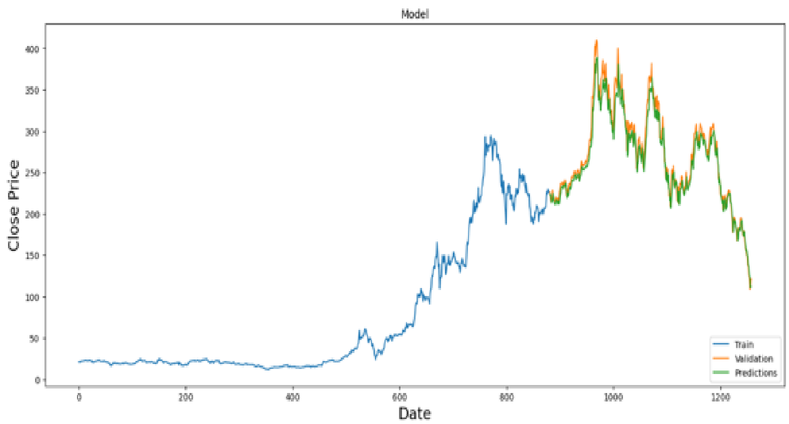


Figure 4.42: *TSLA LSTM Network Performance with 200 Epochs and Batch Size 15*

The LSTM model demonstrated exceptional performance when applied to TSLA's data, achieving an average accuracy of 96.84%. All four configurations achieved accuracies ranging from 95% to 97%, with minor variations in decimal values. Configuration 2 attained superior performance with a 97% accuracy and minimal error rates, specifically 11.069 RMSE, 122.520 MSE, and 3% MAPE, after undergoing 100 epochs of training with a batch size of 15. Figure 4.21 showcases the efficacy of LSTM in accurately predicting the peak of price fluctuations over the entire prediction duration. LSTM outperformed AMZN and GOOG in terms of precision and error rates when applied to TSLA's data, while requiring fewer epochs and a smaller batch size. Configuration 1 attained a 96.84% accuracy rate, comparable to that of configuration 2. However, the results showed a slightly higher error rate, with an RMSE of 11.087, MSE of 122.928, and MAPE of 3.16%. Table 4.5 indicates that Configurations 3 and 4 achieved accuracies of 95.53% and 96.35%, respectively. However, it should be noted that these configurations also resulted in higher RMSE, MSE, and MAPE values. Configuration 4 outperformed configuration 3 with a 1.22% difference in accuracy. Only Configurations 1 and 2, illustrated in Figures 4.20 and 4.21, respectively, accurately predicted the minimum TSLA closing prices. Figures 4.22 and 4.23 illustrate Configurations 3 and 4, respectively, as the most accurate in predicting TSLA closing prices.

Overview Table 4.6 displays the mean values of the assessment methods em-

LSTM				
	RMSE	**MSE**	**MAPE**	**Accuracy**
APPL	4.241	18.254	2.14%	97.87%
AMZN	3.988	16.673	2.22%	97.78%
META	8.247	68.264	2.85%	97.15%
GOOG	3.219	10.532	7.32%	97.86%
TSLA	13.131	178.104	3.57%	96.43%

Table 4.6: *LSTM Average Performance*

ployed and the precision of each configuration test for the financial markets under investigation. The aforementioned graphs demonstrate that LSTM configurations demonstrated efficacy across all five financial markets. All figures exhibited satisfactory predictive performance on the test data. The LSTM model demonstrated superior performance when evaluated with APPL's dataset, exhibiting the highest error values and the lowest accuracy percentage compared to the other four investments. The observed phenomenon can be explained by the dataset of APPL displaying a greater degree of price variability in comparison to the other markets under investigation.

4.6 Analysis and discussion in relation to the literature

This study utilised a methodology similar to two previous studies [32], [35] cited in the references, as described in Chapter 2. In a previous study, the authors examined the use of Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM) to forecast daily Nike stock prices and future stock market values. The findings of this investigation are reported in reference []. The study utilised a dataset covering Nike pricing and transaction volume from December 2, 1980, to December 19, 2019. The research utilised LSTM RNN to train data and evaluated the model's effectiveness by conducting experiments with varying epoch values (12, 25, 50, and 100). The findings indicate that the model effectively tracked fluctuations in the initial asset prices. In a comparative study, the efficacy of three distinct time series models, namely LSTM, GRU, and LSTM NN models, was evaluated for predicting stock prices. The study is cited as [35]. The research utilised the S&P 500 Index, the Dow Jones Index, and the Hang Seng Index as empirical data. The effectiveness of LSTM models in analysing time series data was assessed using techniques akin to those utilised in reference [32]. The evaluation of the models was conducted by measuring their capacity to minimise Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The results of the study suggest that both models exhibited superior outcomes, displaying enhanced fitting and prediction accuracy when compared to raw and predicted values. Reference [32] reported that LSTM achieved optimal performance in the proposed research. The NLSTM model exhibited notable performance and ranked as the second-best performer. The GRU model exhibited a high level of accuracy, albeit with a slightly lower level of efficiency, which is consistent with the findings reported in reference [32]. The prediction models were comparatively evaluated using performance metrics, including MSE, MAE, MAPE, and R-Squared, to determine the optimal model. The LSTM model demonstrated superior predictive performance, achieving accuracy rates above 90% across all financial markets analysed. The GRU algorithm demonstrated a 90% accuracy rate across various financial markets. The efficacy of the WLSTM model was exceptional, as evidenced by accuracy scores exceeding 90% for all datasets.

4.7 Analysis and discussion in relation to the hypothesis and research questions

To confirm the hypothesis, the investigator had to scrutinise and deliberate on the findings. The research query was addressed through the use of a prototype solution. The study's hypothesis is that the use of machine learning to enhance prediction-based decision making through the analysis of historical data and current exchange rates can lead to increased profitability. This hypothesis proposes the use of machine learning as a direct variable to optimise profit, while utilising historical datasets as an indirect variable for prediction. The implemented solution addressed the direct variable, whereas the online survey focused on the indirect variables. This study assessed the efficacy of an LSTM RNN algorithm in improving predictive decision-making using survey responses. The performance of the pupils was found to be satisfactory while using the prototype, and the participants expressed overall satisfaction with their experience. Chapter 4 reported that most students found this strategy fascinating. The hypothesis is supported by the participant responses in chapter 4 and the practical session performance. The Figure provides evidence supporting the hypothesis.

4.7.1 Can machine learning contribute to be optimal when dealing with stock exchange?

In addition to the evaluation method outlined in Section [4], a graphical forecasting approach was employed to assess the accuracy of the data models. The LSTM model achieved a prediction accuracy of 96-97% when trained on the respective original datasets for each equity market. The present study's implementation model demonstrated comparable forecasting accuracy to the Literature Review study that evaluated their model's performance using LSTM. Additionally, a previous study [35] showed that the implementation model was capable of producing an ensemble model with similar accuracy.

4.7.2 Are movements in rates of stock markets predictable when taking into consideration past datasets required for this research?

The LSTM model employed in this study necessitated a sole dataset spanning from 2018 to 2022. The LSTM model was designed to predict the closing price of the stock market. Figure 4 displays several sets of projected stock markets, revealing the absence of a statistically significant distinction upon utilising additional characteristics from the stock market dataset in the analysis. Upon application of all features, the LSTM model exhibited a high degree of accuracy ranging from 96% to 97% across all stock markets.

Previous discussions did not incorporate the proximity attribute as the primary focus in their research. However, the researchers incorporated all the available features in their datasets during the investigations. Given that the focus of the study was on utilising the close feature as the dependent variable, it is evident that relying solely on the close feature yields comparable accuracy to utilising all features.

It is worth mentioning that certain studies included datasets with varying sizes, either small or large, in terms of aggregate quantity. The dataset used in our study was similar in size to those used in studies [32] and [35].

4.8 Conclusion

This chapter provides a comprehensive analysis of the research findings. The paper comprises a research strategy, data analysis method, analysis and discussion of the online survey, prototype, literature, and hypothesis and research question. The final chapter will contain the conclusion and recommendations, as well as limitations, suggestions for future research, and concluding remarks. The sections aim to offer a brief overview of the study's importance, a summary of the results, recommendations for enhancement, and potential avenues for future research.

Chapter 5: Conclusions and Recommendations

5.1 Introduction

Currently, the prediction of stock market trends is a highly coveted area of study. Several scholars have created various models to accurately predict stock prices. Most researchers concentrate on distinct characteristics to forecast the stock market, while only a few endeavours to integrate multiple factors, given the market's susceptibility to various influences. The utilisation of a trend-based methodology, which relies on a company's past stock price data to anticipate its future value, has become a more widespread technique for forecasting stock prices. This method has been widely adopted due to the significant impact that investors and merchants have on a company's valuation. The uniformity of this approach ensures its widespread adoption and adherence. As a result, achieving the intended outcome. The focus of this study was to utilise AI algorithms for the purpose of predicting financial market closing prices. The chosen model for price prediction was an LSTM neural network, which was identified as the most prominent model for this task based on the literature review presented in Chapter 2. The LSTM neural network model was trained with diverse configurations and its optimal technique was determined through an evaluation for this study. Chapter 3, Section 4.4.1 explicated the LSTM architecture, which yielded favourable evaluation metrics during training. LSTM exhibited remarkable performance and accuracy in all financial markets examined. The efficacy of predicting closing prices was demonstrated by the accuracy range of 90 to 97% for each market.

5.2 Limitations

5.2.1 Test Configurations and Models

This study analysed five financial markets (AAPL, AMZN, META, GOOG, TSLA) using a singular AI model, LSTM, due to time constraints. In order to conclude the research within the allotted time frame, the available test configurations were limited.

5.2.2 Hardware

The investigation was conducted on a Windows 11 Home 64-bit-capable Asus Vivo Book laptop. The laptop was outfitted with a 1.19GHz Intel Core i5 10th generation processor and 8GB of RAM. During the LSTM model's training, the execution duration for each configuration occasionally exceeded 40 minutes.

5.3 Future Work

5.3.1 Dataset Ascribe

The methodology aimed to forecast the stock market by utilising the ”Close” attribute, denoting the final traded price of the day. However, the data obtained through Yahoo Finance’s API comprises of five supplementary attributes, namely ”Open,” ”High,” ”Low,” ”Adjusted Close,” and ”Volume.” Incorporating these characteristics into the model could enhance its precision, as they offer valuable insights into the stock market.

5.3.2 Frontend of Prototype

Designing an accessible frontend interface for users to assess the prototype could serve as a viable experiment for the proposed investigation. This would particularly benefit individuals lacking a robust IT or technical proficiency. One possible strategy is to establish a website for deploying the developed prediction models. The implementation of predictive models on the website enables users to track their desired financial markets and gain insights into potential price variations. The models’ predictions would enable them to make informed decisions about investment timing, retention, or divestment. A qualitative approach may be employed for the experiment’s design. The proposed approach involves gathering user feedback and insights on their experience with the frontend interface and the provided predictions. Users may be asked to provide feedback on the usability, clarity, and efficacy of the interface in aiding their investment decision-making. The qualitative data could offer significant insights into the user experience and the practical usefulness of the prediction models in real-life situations.

5.4 Concluding Remarks

The concluding chapter presents the researcher's final thoughts and suggestions for future research. The research process has enhanced personal and educational skills, such as research proficiency and effective problem-solving. The researcher made significant efforts to overcome obstacles and present the research in a refined manner. The study's noteworthy discovery is that the LSTM Neural Network model effectively projected stock market values for all five datasets examined. The statement suggests that sophisticated AI technology has the ability to predict financial market fluctuations with precision. Furthermore, these models are capable of being customised for pragmatic purposes, thereby aiding users in making investment choices. To optimise the project, it is recommended to explore a diverse range of test configurations and model parameters. Moreover, the utilisation of high-performance hardware can expedite programme execution, facilitating increased experimentation. Increasing the size of the dataset utilised for both training and testing would lead to a more thorough evaluation of the model's efficacy. Furthermore, the deployment of the completed programme on a website for public consumption is a critical measure towards enhancing the accessibility of the predictions to a broader demographic. Enabling users to monitor multiple financial markets facilitates the acquisition of valuable information regarding potential price fluctuations, thereby simplifying investment decision-making.

List of References

[1] Rakhi Batra and Sher Muhammad Daudpota. Integrating stocktwits with sentiment analysis for better prediction of stock price movement. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 03 2018.

[2] Fatma Jemai, Mohamed Hayouni, and Sahbi Baccar. Sentiment analysis using machine learning algorithms, 06 2021.

[3] Facundo Bre, Juan M. Gimenez, and Víctor D. Fachinotti. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158:1429–1441, 01 2018.

[4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 11 1997.

[5] IBM Cloud Education. What is machine learning?, 07 2020.

[6] Bleed ai - custom ai solutions, 01 2020.

[7] JAY WILPON, DAVID THOMSON, SRINIVAS BANGALORE, PATRICK HAFFNER, and MICHAEL JOHNSTON. The fundamentals of machine learning.

[8] Hassan Raza, M. Faizan, Ahsan Hamza, Ahmed Mushtaq, and Naeem Akhtar. Scientific text sentiment analysis using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 10, 2019.

[9] Dase R.K, Pawar D. D, and Daspute D.S. Methodologies for prediction of stock market: An artificial neural network. *International Journal of Statistika and Mathematika*, 1:8–15, 2011.

[10] Reddy K. Stock market prediction using machine learning techniques. *International Research Journal of Engineering and Technology (IRJET)*, 7:6780–6785, 2020.

[11] Investor.gov. Stocks — investor.gov, 2022.

[12] Adam Hayes. Stock, 07 2022.

[13] Darren Duxbury and Songyao Yao. Are investors consistent in their trading strategies? an examination of individual investor-level data. *International Review of Financial Analysis*, 52:77–87, 07 2017.

[14] Giovanni Di Crescenzo. Privacy for the stock market. pages 269–288, 02 2002.

[15] Zineb Lanbouri and Said Achchab. A new approach for trading based on long-short term memory ensemble technique. *IJCSI International Journal of Computer Science Issues*, 16, 05 2019.

[16] Halit Yanikkaya. Trade openness and economic growth: a cross-country empirical investigation. *Journal of Development Economics*, 72:57–89, 10 2003.

[17] GIOVANNI CESPA and XAVIER VIVES. The beauty contest and short-term trading. *The Journal of Finance*, 70:2099–2154, 09 2015.

[18] Rupesh A. Kamble. Short and long term stock trend prediction using decision tree, 06 2017.

[19] Hans R. Stoll and Robert E. Whaley. Stock market structure and volatility. *Review of Financial Studies*, 3:37–71, 01 1990.

[20] Mohd Sabri Ismail, Mohd Salmi Md Noorani, Munira Ismail, Fatimah Abdul Razak, and Mohd Almie Alias. Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from kuala lumpur stock exchange. *Applied Soft Computing*, 93:106422, 08 2020.

[21] Laurence Saes. Unit test generation using machine learning. *Faculteit Der Natuurwetenschappen, Wiskunde En Informatica Master Software Engineering*, page 56, 08 2018.

[22] Basav Roychoudhury, Sharad Bhattacharya, and Mousumi Bhattacharya. Applicability of machine learning tools in foreign exchange market. *International Journal of Digital Content Technology and Its Applications(JDCTA)*, 9, 08 2015.

[23] FIRUZ KAMALO and IKHLAAS GURRIB. *MACHINE LEARNING BASED FORECASTING OF SIGNIFICANT DAILY RETURNS IN FOREIGN EXCHANGE MARKETS*. KMG, 2022.

[24] Monica Gupta. A comparative study on supervised machine learning algorithm. *International Journal for Research in Applied Science and Engineering Technology*, 10:1023–1028, 01 2022.

[25] Vladimir Nasteski. An overview of the supervised machine learning methods. *HORIZONS.B*, 4:51–62, 12 2017.

[26] Muhammad Usama, Junaid Qadir, Aunn Raza, Hunain Arif, Kok-lim Alvin Yau, Yehia Elkhatib, Amir Hussain, and Ala Al-Fuqaha. Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access*, 7:65579–65615, 2019.

[27] Pieter-Jan Kindermans, Kristof T Schütt, Maximilian Alber, Klaus-Robert Müller, Dumitru Erhan, Been Kim, and Sven Dähne. Learning how to explain neural networks: Patternnet and patternattribution. 05 2017.

[28] Krzysztof Laddach, Rafał Łangowski, and Bartosz Puchalski. An automatic selection of optimal recurrent neural network architecture for processes dynamics modelling purposes. 116:108375–108375, 12 2021.

[29] Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, and Lamblinp@iro Ca. Exploring strategies for training deep neural networks pascal lamblin. *Journal of Machine Learning Research*, 1:1–40, 2009.

[30] Risto Miikkulainen, Jason Liang, Elliot Meyerson, Aditya Rawal, Daniel Fink, Olivier Francon, Bala Raju, Hormoz Shahrzad, Arshak Navruzyan, Nigel Duffy, and Babak Hodjat. Evolving deep neural networks. *Artificial*

Intelligence in the Age of Neural Networks and Brain Computing, pages 293–312, 2019.

[31] Akira Yoshihara, Kazuki Fujikawa, Kazuhiro Seki, and Kuniaki Uehara. Predicting stock market trends by recurrent deep neural networks. *Lecture Notes in Computer Science*, pages 759–769, 2014.

[32] Shibboleth authentication request.

[33] Jiayu Qiu, Bin Wang, and Changjun Zhou. Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLOS ONE*, 15:e0227222, 01 2020.

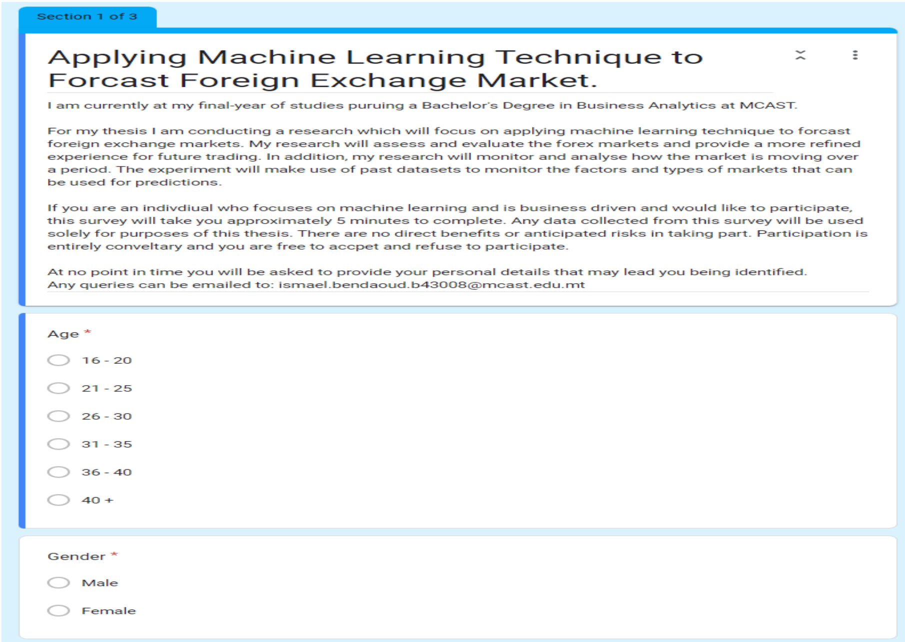
[34] Shibboleth authentication request.

[35] Adil Moghar and Mhamed Hamiche. Stock market prediction using lstm recurrent neural network. *Procedia Computer Science*, 170:1168–1173, 2020.

Chapter A: Ethics Consent Form

Ethics Consent Form

Chapter B: Online Survey



Section 1 of 3

Applying Machine Learning Technique to Forcast Foreign Exchange Market.

I am currently at my final-year of studies puruing a Bachelor's Degree in Business Analytics at MCAST.

For my thesis I am conducting a research which will focus on applying machine learning technique to forcast foreign exchange markets. My research will assess and evaluate the forex markets and provide a more refined experience for future trading. In addition, my research will monitor and analyse how the market is moving over a period. The experiment will make use of past datasets to monitor the factors and types of markets that can be used for predictions.

If you are an indivdiual who focuses on machine learning and is business driven and would like to participate, this survey will take you approximately 5 minutes to complete. Any data collected from this survey will be used solely for purposes of this thesis. There are no direct benefits or anticipated risks in taking part. Participation is entirely conveltary and you are free to accpet and refuse to participate.

At no point in time you will be asked to provide your personal details that may lead you being identified.
Any queries can be emailed to: ismael.bendaoud.b43008@mcast.edu.mt

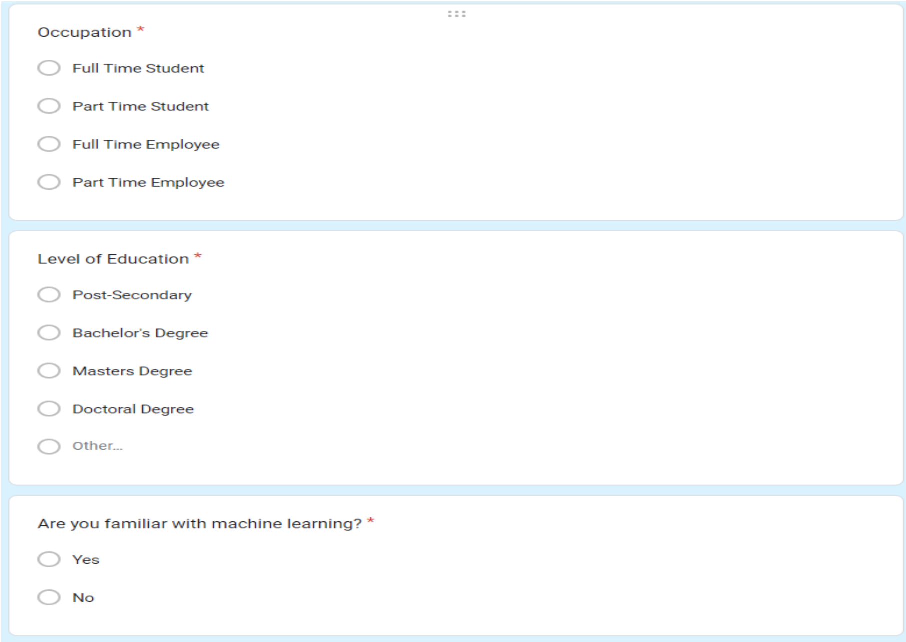
Age *

- 16 - 20
- 21 - 25
- 26 - 30
- 31 - 35
- 36 - 40
- 40 +

Gender *

- Male
- Female

Figure B.1: *Online Survey Part 1*



Occupation *

- Full Time Student
- Part Time Student
- Full Time Employee
- Part Time Employee

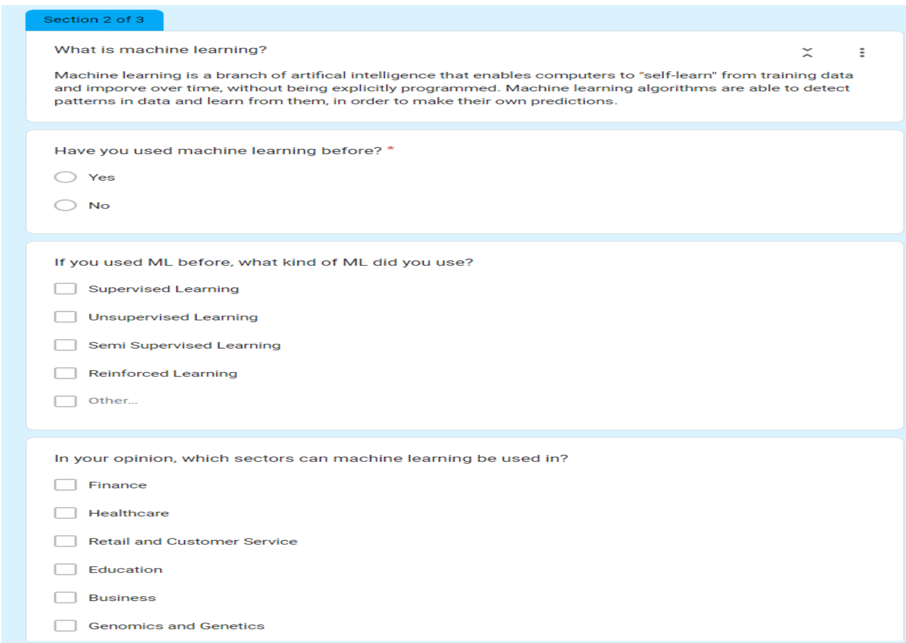
Level of Education *

- Post-Secondary
- Bachelor's Degree
- Masters Degree
- Doctoral Degree
- Other...

Are you familiar with machine learning? *

- Yes
- No

Figure B.2: *Online Survey Part 2*



Section 2 of 3

What is machine learning?

Machine learning is a branch of artifical intelligence that enables computers to "self-learn" from training data and imporve over time, without being explicitly programmed. Machine learning algorithms are able to detect patterns in data and learn from them, in order to make their own predictions.

Have you used machine learning before? *

- Yes
- No

If you used ML before, what kind of ML did you use?

- Supervised Learning
- Unsupervised Learning
- Semi Supervised Learning
- Reinforced Learning
- Other...

In your opinion, which sectors can machine learning be used in?

- Finance
- Healthcare
- Retail and Customer Service
- Education
- Business
- Genomics and Genetics

Figure B.3: *Online Survey Part 3*

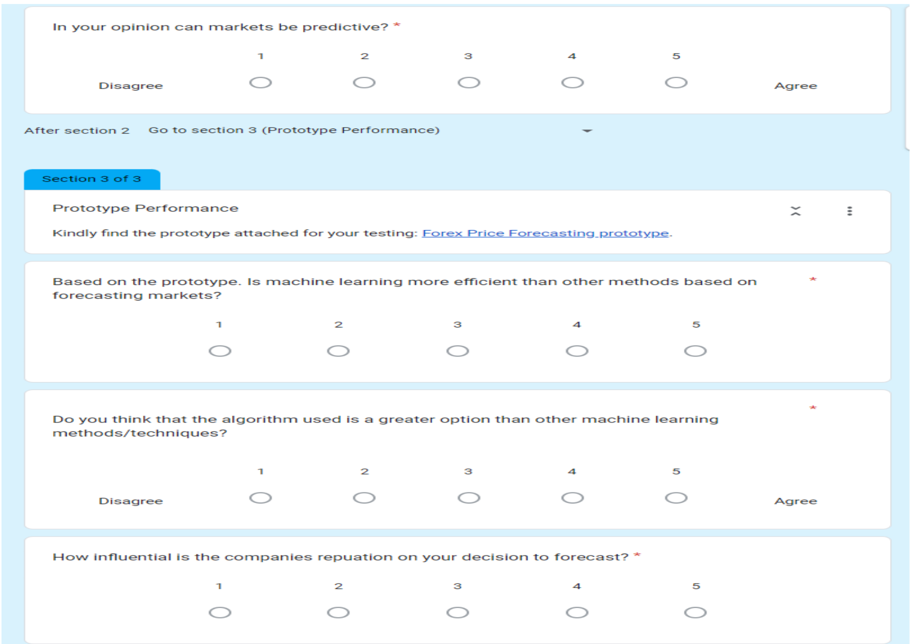
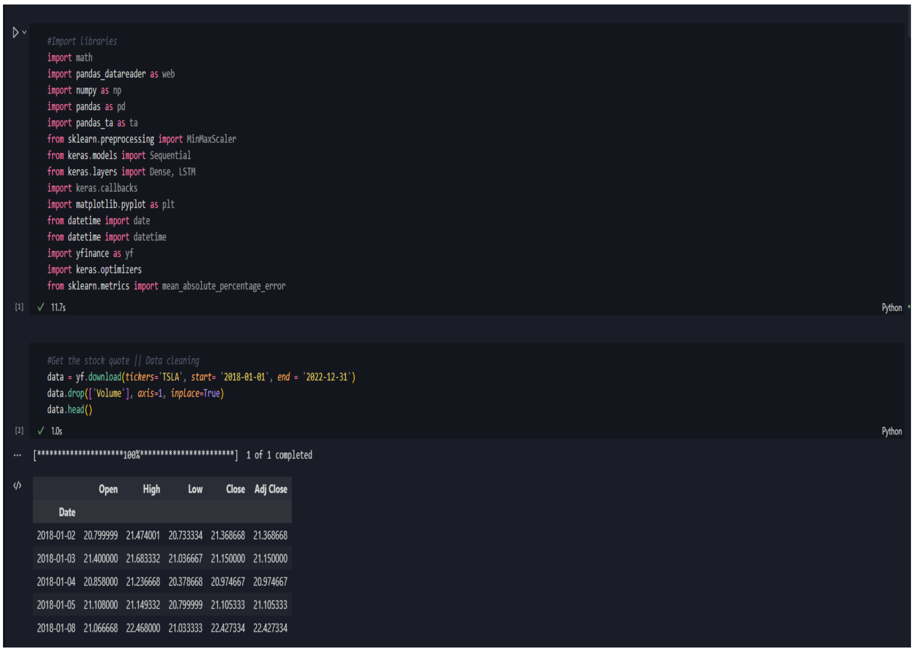


Figure B.4: *Online Survey Part 4*

Chapter C: Code Snippets

Code Snippets:



```
#Import libraries
import math
import pandas_datareader as web
import numpy as np
import pandas as pd
import pandas_ta as ta
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense, LSTM
import keras.callbacks
import matplotlib.pyplot as plt
from datetime import date
from datetime import datetime
import yfinance as yf
import keras.optimizers
from sklearn.metrics import mean_absolute_percentage_error
```

[1] ✓ 11.7s Python

```
#Get the stock quote || Data cleaning
data = yf.download(tickers='TSLA', start= '2018-01-01', end = '2022-12-31')
data.drop(['Volume'], axis=1, inplace=True)
data.head()
```

[2] ✓ 1.0s Python

```
[*********************100%***********************]  1 of 1 completed
```

Date	Open	High	Low	Close	Adj Close
2018-01-02	20.799999	21.474001	20.733334	21.368668	21.368668
2018-01-03	21.400000	21.683332	21.036667	21.150000	21.150000
2018-01-04	20.858000	21.236668	20.378668	20.974667	20.974667
2018-01-05	21.108000	21.149332	20.799999	21.105333	21.105333
2018-01-08	21.066668	22.468000	21.033333	22.427334	22.427334

Figure C.1: *Code Snippet 1*

```
#Adding Indicators
data['Target'] = data['Close']-data.Open
data['Target'] = data['Target'].shift(-1)

data['TargetNextClose'] = data['Adj Close'].shift(-1)

data.dropna(inplace=True)
data.reset_index(inplace = True)
```
[4] ✓ 0.0s

```
data_set = data.iloc[:, 1:11]#.values
pd.set_option('display.max_columns', None)
data_set.head(20)
```
[5] ✓ 0.0s

	Open	High	Low	Close	Adj Close	Target	TargetNextClose
0	20.799999	21.474001	20.733334	21.368668	21.368668	-0.250000	21.150000
1	21.400000	21.683332	21.036667	21.150000	21.150000	0.116667	20.974667
2	20.858000	21.236668	20.378668	20.974667	20.974667	-0.002666	21.105333
3	21.108000	21.149332	20.799999	21.105333	21.105333	1.360666	22.427334
4	21.066668	22.468000	21.033333	22.427334	22.427334	-0.098000	22.246000
5	22.344000	22.586666	21.826668	22.246000	22.246000	0.173332	22.320000
6	22.146667	22.466667	22.000000	22.320000	22.320000	0.180668	22.530001
7	22.349333	22.987333	22.217333	22.530001	22.530001	-0.160666	22.414667
8	22.575333	22.694000	22.244667	22.414667	22.414667	0.168001	22.670668
9	22.502666	23.000000	22.320000	22.670668	22.670668	0.445999	23.143999
10	22.698000	23.266666	22.650000	23.143999	23.143999	-0.073334	22.971333
11	23.044666	23.486668	22.916000	22.971333	22.971333	0.334667	23.334667
12	23.000000	23.372667	22.840000	23.334667	23.334667	0.143999	23.437332
13	23.293333	23.855333	23.280001	23.437332	23.437332	-0.480667	23.519333
14	24.000000	24.033333	23.400000	23.519333	23.519333	-0.579332	23.059334
15	23.638666	23.650000	22.901333	23.059334	23.059334	-0.708668	22.509333
16	23.218000	23.280001	22.426666	22.509333	22.509333	0.090000	22.856667
17	22.766666	22.933332	22.380667	22.856667	22.856667	0.645332	23.302000
18	22.656668	23.389999	22.552000	23.302000	23.302000	0.045334	23.054667

Figure C.2: *Code Snippet 2*

```
valid.head(20) #close price vs the forcasting closing price
```

	Close	Predictions
881	226.300003	228.014114
882	219.860001	228.778702
883	214.883331	223.475708
884	217.603333	218.677643
885	218.983337	219.437485
886	228.566666	220.091690
887	222.846664	228.891068
888	217.793335	224.573669
889	216.866669	220.449707
890	214.740005	217.696396
891	215.406662	215.361481
892	220.166672	216.662613
893	218.429993	221.807526
894	216.419998	221.289780
895	214.460007	219.441437
896	219.206665	216.129608
897	214.926666	219.136841
898	215.660004	215.768646
899	225.783340	217.546997
900	229.066666	226.584244

Figure C.3: *Code Snippet 3*

```python
#Create the training data set
#1. Create the scaled training data set
train_data = scaled_data[0:training_data_len , :] #contining all the values of index 0 of training data len

#split the data into x_train and y_train data sets
x_train = [] #indipented training variables
y_train = [] #dependent variable / target variable

for i in range(60, len(train_data)):
    x_train.append(train_data[i-60:i, 0])
    y_train.append(train_data[i, 0])
    if i<= 61:
        print(x_train)
        print(y_train)
        print()
```

```
[10]  ✓ 0.0s

[array([0.02370959, 0.02316023, 0.02271974, 0.02304801, 0.0263693 ,
       0.02591373, 0.02609964, 0.02662723, 0.02633748, 0.02698063,
       0.02816979, 0.027736  , 0.02864881, 0.02890674, 0.02911275,
       0.02795709, 0.02657531, 0.02744792, 0.02856674, 0.02794536,
       0.02936733, 0.02851984, 0.02759866, 0.02581994, 0.02596063,
       0.02780802, 0.0228219 , 0.02201629, 0.02290565, 0.02423383,
       0.02400772, 0.02597738, 0.02621521, 0.02609462, 0.02584841,
       0.02800398, 0.02898881, 0.02988822, 0.02881127, 0.02748309,
       0.02545147, 0.02615324, 0.02585679, 0.02499422, 0.02568092,
       0.02514496, 0.02482171, 0.02789344, 0.02727876, 0.02473127,
       0.02455876, 0.02384693, 0.0225422 , 0.02203806, 0.02303964,
       0.0217952 , 0.020529  , 0.02097117, 0.01678396, 0.01319972])]
[0.014598249993150625]

[array([0.02370959, 0.02316023, 0.02271974, 0.02304801, 0.0263693 ,
       0.02591373, 0.02609964, 0.02662723, 0.02633748, 0.02698063,
       0.02816979, 0.027736  , 0.02864881, 0.02890674, 0.02911275,
       0.02795709, 0.02657531, 0.02744792, 0.02856674, 0.02794536,
       0.02936733, 0.02851984, 0.02759866, 0.02581994, 0.02596063,
       0.02780802, 0.0228219 , 0.02201629, 0.02290565, 0.02423383,
       0.02400772, 0.02597738, 0.02621521, 0.02609462, 0.02584841,
       0.02800398, 0.02898881, 0.02988822, 0.02881127, 0.02748309,
       0.02545147, 0.02615324, 0.02585679, 0.02499422, 0.02568092,
       0.02514496, 0.02482171, 0.02789344, 0.02727876, 0.02473127,
       0.02455876, 0.02384693, 0.0225422 , 0.02203806, 0.02303964,
...
       0.02384693, 0.0225422 , 0.02203806, 0.02303964, 0.0217952 ,
```

Figure C.4: *Code Snippet 4*