**Research Design 2: Assignment Ismael Ben Daoud**

**Task 1**

**Section 1**

Title: Applying Machine Learning Technique to Forecast Stock Exchange Markets.

Chosen Research:

1. Description of Theme and Topic Rationale.
   1. The main topic of this study is machine learning. ML is a branch of artificial intelligence which focuses on the use of data and algorithms to imitate the way that a user can learn and improve accuracy. The Theme will concentrate on the return-on-investment ROI. In this study supervised learning is used to train algorithms to properly categorize datasets and predict outcomes.
2. Positioning and Research Onion.
   1. Research philosophy: In this study the guided by a positivist philosophy, which emphasizes objectivity and the use of quantitative methods to determine if markets can be forecasted and collect numerical data and select a survey based on the research strategy to gather data from a representative sample of different people with different knowledgebase base weather it’s on artificial intelligence or business oriented which can be proven to be true or false and thus this knowledge is seen as meaningful.
   2. Research approach: The research approach is reasoned as deductive, as the study aims to test a specific hypothesis related to the use of machine learning techniques to stock exchange markets.
   3. Research Strategy: The research strategy involves conducing a survey and a collection of datasets from existing sources such as yahoo finance APIs in predicting currency exchange markets.
   4. Choices: The data analysis phase will make use of a quantitative approach which will involves the use of historical data through the use of yahoo finance API to extract the required historical data from different stock markets and a statistical method, such as T-Test, Chi-Square, and ANOVA testing through the use of IBM SPSS to analyse the data and assess the accuracy of the machine learning prototype in predicting rates.
   5. Time Horizon: The study will use a cross-sectional study to collect data on the historical prices of the different stock markets. The study would also involve a survey that is administered in relation to the prototype.
   6. Techniques and Procedures: In this research a quantitative approach is being undertaken as the purpose is to forecast the stock exchange market by creating a python script and get historical data and create a survey based on that prototype. Therefore, the use of interviews will not be considered as the purpose is to gain a high-volume numerical data response which survey is far better suited to.
3. Background to this research theme
   1. The researcher has a great interest in both software and business aspects. As a business analytics student, it is a great opportunity to develop a project-based on both software sector and business sector combined together while also using different techniques of analysing the effects of forecasting on the markets through the use of machine learning.
4. Hypothesis
   1. As for the hypothesis it will focus on maximising profits by implementing machine learning to improve decision making from predictions using past data and current exchange rates.
5. Research Aim and Purpose Statement
   1. Research Aim: The research aim is to forecast the forex exchange market by creating a python script and get historical data. This would help better understand the forex market and provide a more refined experience for future trading. The research method will be quantitative, as the data will be collected from existing forex data to monitor and analyse how the market is changing and developing creating a detailed outlook and unique perspective on how the market is moving over a period. The experiment will make use of a past datasets to analyse the factors and types of markets that can be used for predictions. The experiment will be developed using Python to deploy a bot and use it to scrape data for market insight which can be made to establish different datasets per each market.

**Section 2**

**Introduction**

Due to the volatility of the stock market, researchers find it challenging to find a solution to forecast price movement. Before the age of computers, people traded stocks and commodities primarily on intuition. As the level of trading and investing grew, people searched for methods and tools that would increase their gains whilst minimizing their risks. Therefore, traditional stock price prediction methods including a statistical approach has been implemented such as statistics, technical analysis, fundamental analysis, and linear regression, which are all used to attempt to predict and benefit from the market’s direction. However, none of these techniques has proven to be consistently correct prediction tool that is desired, and many analysts argue about the usefulness of many of the approaches. However, machine learning approaches are proven to be more accurate and efficient in predicting price movement than traditional methods.

1. In recent years, there has been a growing interest in the use of artificial intelligence and machine learning techniques for stock market prediction. In the literature review reveals a method have been employed in previous studies, including neural networks and deep learning. Additionally, some studies have incorporated natural language processing to analyse historical data for predicting stock prices.
2. Distinguish between academic and non-academic materials:
   1. Distinguishing between academic and non-academic material is crucial while conducting a comprehensive literature review. While non-academic material such as blog posts and online articles may provide useful insights, academic material such as peer-reviewed journals are more reliable sources of information.
3. Recommend 5 articles from peer reviewed journals.
   1. [1]. Stock price prediction using machine learning: A survey. This survey paper provides a comprehensive overview of the different machine learning techniques used for stock price predication, including their strengths and limitations.
   2. According to the research [2] this book offers an in-depth look at the use of machine learning techniques in finance, with a particular focus on the challenges and opportunities in applying these techniques to the foreign exchange market.
   3. In this study proposed a deep learning framework from predicting financial time series, including foreign exchange rates, using a combination of stacked auto coders and long-short term memory networks [3].
   4. A deep learning-based approach for international currency exchange rate forecasting. Neural computing and applications, [4]. This article describes a deep learning-based approach for forecasting international currency exchange rates, using a convolutional neural network.
   5. A hybrid forecasting model of financial time series based on deep learning and optimization algorithm [5]. Applied soft computing, this study proposes a hybrid forecasting model of financial time series, including foreign exchange rates, based on a combination of deep learning and optimization algorithms.
4. Contextualized literature and research material.
   1. Contextualizing literature and research material involves summarizing and synthesizing the findings from various studies to provide a comprehensive understanding of the research topic. In analysing the existing literature on machine learning for stock market prediction, it can be concluded that while machine learning techniques have shown promise in improving the accuracy of stock price prediction, the effectiveness of these models is influenced by various factors, such as the quality and quantity of data, the choice of algorithm, and the market conditions.
   2. Critical literature arguments involve comparing and contrasting the findings from different studies and identifying knowledge gaps and challenges that need to be addressed in future research. One major challenge in using machine learning for stock market prediction is the tendency of these models to overfit to the training data, resulting in poor performance on unseen data. Additionally, the impact of external factors such as economic, political, and social events on stock prices is often unpredictable and difficult to incorporate into machine learning models. Therefore, the research focuses on developing models that can handle uncertainty and incorporate external factors into their predictions. Moreover, there is a need for more comprehensive and standardized datasets to evaluate the performance of different machine learning models for stock price prediction.
5. By using critical literature arguments into the research, it will provide a balanced perspective and contribute to the identifications of areas where further research and improvements are needed. As stated by [6] this paper critically examines the strengths and weaknesses of artificial neural networks for forecasting, including stock market prediction. It discusses issues such as overfitting, the need for appropriate training data, and the interpretability of ANN models. Another study by [7], compares the performance of various forecasting models, including machine learning techniques for predicting volatility and option prices of the S&P 500 index. It highlights the strengths and weaknesses of different models and identifies knowledge gaps in terms of model selection and performance evaluation. These papers serve as examples of research that critically evaluates the existing methodologies and identifies any limitations and gaps in the field of stock market forecasting through machine learning. The key objective of this research is how to maximises profits by implementing machine learning to improve decision making from predictions using past data and current exchange rates.
6. Construct a Literature Map:

A picture containing text, diagram, font, plan

Description automatically generated

**Section 3**

**Reflection on the chosen Methodology**

1. Define your research questions and define your objectives.
   1. Research Question 1: “*Are movements in rates of stock exchange markets predictable when taking into consideration past datasets with exchange rates?*”

The objective of this research question is to investigate the predictability of stock exchange movements by analysing past datasets. The research aims to access whether historical data can provide insight into future stock movements. Therefore, determining the degree of predictability in stock exchange rates based on past information.

* 1. Research Question 2: “Can machine learning contribute to an optimal decision making when dealing with stock exchange?”

The objective is to explore the potential contribution of machine learning techniques in making the right decisions related to the markets chosen. This research aims to evaluate the effectiveness of machine learning algorithms, models, and methodologies used in analysing the stock market data, by identifying patters and generating predictions to make better decisions. The objective is to determine whether machine learning can enhance decision-making processes in the context of stock exchange and to potentially improve decision-making for investors or traders.

1. For the research philosophy positivism will be selected as the knowledge will be obtained through objective observation and measurements. The positivist research aims to uncover general laws and casual relationships by implementing scientific method which in this case will be machine learning. As for the data collection method quantitative data will be collected and statistical analysed and establish empirical evidence of the prototype. As for the research approach as stated before a quantitative approach will be taken into consideration. This approach will focus on collecting and analysing numerical data to identify forecasting, patters, and statistical significance. The research employees a structured data from yahoo finance API, and a survey based on the prototype to draw objective conclusion. Lastly, as mentioned before this research aligns as deductive approach, aiming to identify patters and draw conclusions based on the data collected.
2. Choose a suitable methodology from those reviewed.
   1. Based on the research objectives mentioned earlier, a suitable methodology would be a combination of quantitative and machine learning approaches. Firstly, to assess the predictability of stock movements, a quantitative approach will be employed. Historical datasets of exchange rates and corresponding stock movements are collected ana analysed through statistical methods. Time series analysis techniques, such as classifications models and exponential smoothing method to identify patters, trends, and correlations in the data. This approach will provide quantitative insights into the relationships between past and present stock market movements. As for machine learning techniques to explore in decision-making, various algorithms can be applied. Supervised learning algorithms, such as neural networks, support vector machines or random forests, can be trained on historical data to predict stock movement. The models performance than is evaluated using appropriate evaluation metrics such as accuracy, mean squared error, root mean squared error and mean absolute percentage error.
3. Initial description of your chosen research methodology, experiment design and method of analysis and why this fits your research objectives.
   1. The chosen research methodology for this study is a combination of quantitative analysis and machine learning techniques. This approach aims to investigate the predictability of stock exchange movements by analysing past datasets of exchange rates and explore the contribution of machine learning in decision-making related to the stock market. The experiment design will involve collecting historical datasets of exchange rated and corresponding market movements over a specific time period. The dataset will contain information on various factors that influence exchange rates and stock behaviour, such as market trends. The datasets will be divided into two subsets: a training set and a testing set. The training set will be used to train the machine learning algorithms, while the test set will be used to evaluate the performance of the trained models in forecasting stock movements. Lastly, for method of analysis the quantitative analysis will involve conducing statistical analysis on the historical datasets. Time series analysis techniques, such as autoregression integrated moving average models or exponential smoothing method, will be applied to identify patters, trends, and correlations in the data. As for machine learning techniques, various machine learning algorithms will be employed to predict stock market movements based on past exchange rates. Such as neural network supervised learning to train the dataset. The combination of quantitative analysis and machine learning techniques allows for a comprehensive investigation of the research objectives. The quantitative analysis provides a foundational understanding of the relationship between exchange rates and stock market movements, while the machine learning algorithms enable the exploration of non-linear patterns and complex interactions. The experiment design, with separate training and test sets, ensures the evaluation of the models' performance on unseen data, assessing their ability to generalize to real-world scenarios. By employing this methodology, the research aims to provide insights into the predictability of stock exchange movements using past exchange rate data and assess the potential contribution of machine learning in decision-making related to the stock market. The detailed analysis of both quantitative and machine learning approaches will enhance the understanding of the research problem and address the research objectives effectively.
4. Any ethical considerations that you may envisage relating to your research endeavour.
   1. When conducting research, there are several ethical considerations that need to be addressed. Some of the key ethical considerations include are:
      1. Privacy and Confidentiality: Ensuring the privacy and confidentiality of the data used in the research is essential. Researchers should obtain appropriate consent and handle personal or sensitive information in a secure manner to protect the confidentiality of the organizations involved.
      2. bias and fairness: The machine learning algorithm can be prone to biases if the training data is unrepresentative or contains inherent biases. It is important to consider and mitigate biases to ensure fairness and prevent discrimination. The research should carefully select and pre-process data to minimize bias and evaluate the model for fairness throughout the research process.
      3. Transparency and Explainability: The machine learning models often operate as black boxes, making it difficult to understand the reasons behind their predictions. Ensuring transparency and Explainability of the models is important to build trust and be able to understand the decision-making process. Researchers should strive to interpret and explain the models’ outcomes and provide insights into the underlying mechanisms.
      4. Responsible use of prediction: the prediction or recommendations generated by machine learning model should be used responsibly. Researchers should emphasize the limitations and uncertainties associated with the predictions and ensure that they are not used to manipulate or exploit financial markets.
      5. research integrity and accuracy: Researchers should maintain high standards of research integrity, ensuring the accuracy, reliability, and reproducibility of the research findings. Any conflicts of interest or potential biases should be disclosed to maintain transparency and credibility.
      6. Informed consent and participant protection: If human participants are involved in this research, informed consent should be obtained, and their rights and welfare should be protected. Participants should be fully informed about the nature of the research, its purpose, potential risks and benefits and their right to withdraw from the study at any time.
   2. Addressing these ethical considerations requires researchers to establish ethical review processes, follow ethical guidelines and codes of conduct and prioritize the well-being and rights of individuals and organizations involvement in the research. By addressing these considerations in a comprehensive and thorough manner, the research can ensure ethical conduct and contribute to responsible and trustworthy outcomes.

**Section 4: Results, analysis, and discussion.**

1. Present your results using established metrics and aggregations relevant to your chosen methodology, in the form of tables, figures, graphs and other appropriate vials where applicable.

**Introduction:**

* 1. In this chapter, the LSTM (Long Short-Term Memory) results for different companies is discussed and then the forecasting of stock markets. The aim will be to improve a better insight into whether implying LSTM will improve the prototype performance. The result of this research study will be displayed by the average performance of each evaluation technique, along with an accompanying visualization highlighting of both actual and predicted close price.

**Metrics:**

In the field of stock prediction, studies are typically categorized into two main approaches: regression and classification. Regression models aim to predict the actual values, often using the future price as the target variable. On the other hand, classification models are employed to determine the direction of movement in stock prices, whether they will increase or decrease.

Text

Description automatically generated

Figure 4.1: Root-Mean-Square Error

The root mean square error (RMSE) is a metric used to measure the standard deviation of prediction errors in regression analysis. It provides a measure of how far the predicted values deviate from the actual regression line. To calculate the RMSE, the predictions are subtracted from the actual data, squared, and then averaged by dividing the result by the total number of data points (sample size, N). Finally, the square root of the average is taken to obtain the RMSE value.

A picture containing font, text, white, number

Description automatically generated

Figure 4.2: Mean Squared Error

The mean squared error (MSE) or mean squared deviation (MSD) is a metric used to measure the average squared difference between estimated values and true values. It provides a measure of the average error or deviation of an estimator. The MSE is a risk function that represents the expected value of the squared error loss. It is always non-negative, and lower values indicate better performance, with values close to zero being desirable. The MSE takes into account both the variance (spread) of the estimator and its bias (deviation from the true value), as it is the second moment of the error related to the origin.

Text

Description automatically generated

Figure 4.3: Mean Absolute Percentage Error

Lastly, the mean absolute percent error measures the predicted and actual data error into percentage error. This metric is calculated from ground truth data that are being compared when deciding the performance of the model.

**Prototype Results:**

**LSTM Network Performance**

the results visualised below are that of LSTM, utilising the close prices of AAPL, META, and TSLA. The primary set for the model analysation is that the data got split 70% for training and 30% for testing. Its results were obtained by running the program four times each with different epoch and batch size values. Afterwards, these results got utilised to find 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**

A picture containing text, screenshot, plot, line

Description automatically generated

Figure 4.3 APPL LSTM Network Performance with 100 Epochs and Batch Size 10

**Graphical user interface, chart, line chart

Description automatically generated**

Figure 4.4 APPL LSTM Network Performance with 100 Epochs and Batch Size 15

Graphical user interface, chart, line chart

Description automatically generated

Figure 4.5 APPL LSTM Network Performance with 200 Epochs and Batch Size 10

Graphical user interface, chart, line chart

Description automatically generated

Figure 4.6 APPL LSTM Network Performance with 200 Epochs and Batch Size 15

The model exhibited excellent performance and accuracy across different test configurations, with results ranging between 97% and 98%. Based on Table 4.1, configuration 2 achieved the highest accuracy and lowest error values, using 100 epochs and a batch size of 15. This configuration achieved an accuracy of 98.17%, RMSE of 3.661, MSE of 13.833, and MAPE of 1.83%. Notably, this configuration required fewer epochs and a smaller batch size to achieve better results. The graph in Figure 4.6 depicts configuration 3, which demonstrates the model's ability to accurately forecast fluctuations in the highest close prices. While configuration 3 achieved the second-best results, with an accuracy of 97.92%, it had slightly higher error rates with an RMSE of 4.078, MSE of 16.628, and MAPE of 2.08%. Configurations 1 and 4, with 100 and 200 epochs and batch sizes of 10 and 15 respectively, also produced promising results. Configuration 1 achieved an accuracy of 97.77%, while configuration 4 scored 97.60%. Figures 4.3 and 4.6 illustrate the successful prediction of the highest fluctuations in close prices by the LSTM model in configurations 1 and 4. It is noteworthy that all configurations achieved accuracy percentages above 90% when using AAPL's historical data.

**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

A picture containing text, plot, screenshot, line

Description automatically generated

Figure 4.11 META LSTM Network Performance with 100 Epochs and Batch Size 10

A picture containing text, line, plot, screenshot

Description automatically generated

Figure 4.12 META LSTM Network Performance with 100 Epochs and Batch Size 15

A picture containing text, line, plot, screenshot

Description automatically generated

Figure 4.14 META LSTM Network Performance with 200 Epochs and Batch Size 10

A picture containing text, line, plot, screenshot

Description automatically generated

Figure 4.15 META LSTM Network Performance with 200 Epochs and Batch Size 15

LSTM also performed well with META data, achieving an average accuracy of 97.15%. Configuration 4, with 200 epochs and a batch size of 15, obtained the highest accuracy and the lowest error values according to Table 4.3 and Figure 4.15. The graph indicates that the model encountered some difficulties in predicting the closing prices in certain instances. Configuration 4 achieved an accuracy of 97.55%, with an RMSE of 7.771, MSE of 60.395, and MAPE of 2.45%. For META prediction, LSTM required more epochs and a larger batch size. However, the other three configurations still produced good predictions, with accuracies ranging from 96% to 97%. Configuration 2, shown in Figure 4.12, was the runner-up in terms of accuracy, achieving 97.49%, with slightly higher error rates of 7.803 RMSE, 72.655 MSE, and 2.51% MAPE. Configuration 1 had 100 epochs and a batch size of 10, while configuration 3 had 200 epochs and a batch size of 10, and both configurations yielded excellent results. Figure 4.15 illustrates that configuration 4 had difficulty accurately predicting the lowest points of closing price fluctuations. On the other hand, configurations 1, 2, and 3 shown in figures 4.11, 4.12, and 4.13 demonstrated that LSTM was able to accurately predict the largest variations in close prices. When comparing the three configurations, it can be concluded that configuration 2 outperformed the others with a 97.49% accuracy.

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

A picture containing text, screenshot, plot, line

Description automatically generated

Figure 4.20 TSLA LSTM Network Performance with 100 Epochs and Batch Size 10

A picture containing screenshot, text, plot, line

Description automatically generated

Figure 4.21 TSLA LSTM Network Performance with 100 Epochs and Batch Size 15

A picture containing text, screenshot, diagram, plot

Description automatically generated

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

A picture containing screenshot, text, plot, line

Description automatically generated

Figure 4.23 TSLA LSTM Network Performance with 200 Epochs and Batch Size 15

LSTM also achieved excellent performance with TSLA's data, with an average accuracy of 96.84%. All four configurations obtained accuracies ranging from 95% to 97% and showed slight fluctuations in their respective fractional numbers. Configuration 2, using 100 epochs and a batch size of 15, achieved the highest accuracy of 97% and the lowest error rates, including 11.069 RMSE, 122.520 MSE, and 3% MAPE. Figure 4.21 provides a graphical visualization of this configuration, showing that LSTM successfully predicted the highest point of price fluctuations throughout the prediction period. In contrast to AMZN and GOOG, for TSLA's data, LSTM required fewer epochs and a smaller batch size to achieve better accuracy and lower error rates. Configuration 1 resulted in an accuracy of 96.84%, which is relatively close to that of configuration 2. However, it produced slightly higher error rates, including 11.087 RMSE, 122.928 MSE, and 3.16% MAPE. Configuration 3 and 4 yielded accuracies of 95.53% and 96.35%, respectively, along with higher RMSE, MSE, and MAPE values, as listed in Table 4.5. When comparing configurations 3 and 4, configuration 4 performed better with a difference of 1.22% in accuracy. It is noteworthy that only configuration 1 (Figure 4.20) and configuration 2 (Figure 4.21) managed to anticipate the lowest close prices of TSLA. On the other hand, configuration 3 (Figure 4.22) and configuration 4 (Figure 4.23) successfully predicted the highest close prices of TSLA.

**Compare and contrast:**

**Overview**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LSTM** | | | | |
|  | **RMSE** | **MSE** | **MAPE** | **Accuracy** |
| **APPL** | **4.241** | **18.254** | **2.14%** | **97.87%** |
| **META** | **8.247** | **68.264** | **2.85%** | **97.15%** |
| **TSLA** | **13.131** | **178.104** | **3.57%** | **96.43%** |

**Table 4.6 LSTM Average Performance**

Table 4.6 provides the average values of the evaluation techniques used and the accuracy of each configuration test for the different financial markets analysed. In general, LSTM configurations demonstrated efficient performance across all five financial markets, as evident from the graphical figures mentioned above. None of the figures displayed significant difficulties in forecasting the testing data. However, it is worth noting that LSTM performed most efficiently with APPL's dataset, as it achieved the highest accuracy percentage out of the five investments. This could be attributed to the fact that APPL's dataset had the largest pricing difference among the analysed markets.

**Discussion of Results in relation to the hypothesis and other studies**

For the hypothesis to be tested the researcher needed to conduct the analysis and discussion of the results. This was done by using a prototype solution with the research question also being answered. As stated earlier, in this study, the hypothesis of this research was “*The possibility of maximising profitability by implementing machine learning to improve decision making from predictions applying past data and current exchange rates*”. This hypothesis has the maximising profitability by implementing machine learning as a direct variable and the prediction applying past datasets as indirect variables. The implemented solution covered the direct. This emphasised whether an LSTM RNN program could truly improve decision-making for prediction. It was noted that the model performed rather well. It was noted previously that the majority of the markets outputted high results. This along with the performance noted proves that the hypothesis can be true.

A graph forecasting representation was conducted to assess the accuracy of the data models. Specifically for the LSTM model, it was observed that the model was able to forecast with nearly identical accuracy predictions when using the original dataset for each stock market, with accuracy ranging between 96% and 97%.

Compared with other studies that investigated their model’s performance with LSTM, the implementation model was able to forecast as accurately as the results found in this study, and also proven by [26] study was able to deliver a similar accurate ensemble model.

**Section 5**

**Conclusion**

**Main Conclusion of your research:**

Stock market prediction is a very sough after field in the modern world. Many researchers have built many models to try to achieve very high accuracy on stock price predictions. Since the stock market is affected by various factors, most researchers focus on different features to make a prediction, but only a few try to incorporate more than one factor. A more commonly used approach for the stock predictions is using a trend-based approach, which uses company’s historical stock prices data to predict its future value. This is a tried and trusted approach, mainly because investors or traders as a whole have the most impact on the value of a company, and if everyone uses the approach, which gives the same result, then people follow this approach. Thereby giving the expected result.

This study focused on utilizing AI algorithm to predict the closing prices of financial markets. The chosen model for this prediction task was a neural network with LSTM architecture, which was identified as the most prominent model for price forecasting based on the literature reviewed in Chapter 2. The neural network model, specifically LSTM, was trained using various configurations, and an evaluation was conducted to determine the optimal technique for this study. In Chapter 3, Section 4.4.1, the architecture of LSTM was defined, and during the training phase, the model achieved excellent evaluation metrics. Notably, LSTM demonstrated outstanding performance and accuracy across all the financial markets analysed. Its accuracy ranged from 97% and above for each market, showcasing its effectiveness in predicting closing prices.

**Addressing the Research Questions and Hypothesis:**

The hypothesis of this research stated that the possibility of maximizing profitability by implementing machine learning to improve decision making is composed of a number of factors which have to be taken into account before implementing any algorithms. The factors which are effecting this action have resulted to be the dataset used for testing and the model itself.

The target feature used in the LSTM model was the close price of the stock market. Figure 4 shows different sets of stock markets being predicted, indicating that the investigation did not find a significant difference when utilizing other features from the stock market dataset. When all features were applied, the LSTM performance for each stock market resulted in high accuracies ranging between 96% and 97%.

**Identify any shortcoming in methodology:**

Due to time constraints, this study focused on analysing three financial markets (APPL, META, TSLA) using a single AI model Long Short-Term Memory (LSTM). The choice of test configurations was limited in order to complete the research within the allocated time as the quality and availability of historical data can significantly impact the accuracy and generalizability to the findings. The inaccurate or incomplete data, limited time periods, or data from specific market conditions may introduce biases or limit the ability to capture complex patterns accurately.

**Ideas for further research:**

The original methodology aimed to focus on using the "Close" attribute, which represents the last transacted price of the day, for stock market prediction. However, the data obtained from Yahoo Finance's API includes five additional features: "Open," "High," "Low," "Adjusted Close," and "Volume." These features provide valuable information related to the stock market and incorporating them into the model could potentially improve its accuracy.

Also, for further research investigating real-time stock market forecasting can be taken into account. This research can focus on developing models that can adapt and update predictions in real-time as new market data becomes available. Which will have the ability to provide timely and accurate forecasts results. Lastly, further research on exploring the application of machine learning techniques in cross-market analysis can be taken into consideration. This can include analysing the interconnections and effects between different stock markets or examining the relationship between stock markets and other financial markets, such as currency exchange or commodity markets.