IMPROVING STOCK PRICE PREDICTION USING MACHINE LEARNING: A COMPARATIVE STUDY OF LSTM, CNN AND TRADITIONAL METHODS.

BY

BOGERE MARK

2020-B291-10364

A Project Report Submitted to the Faculty of Science in Partial Fulfillment of the Requirements for the Award of the Degree of Bachelor of Science in Computer Science of Uganda Martyrs University

Supervisors

Dr Sanya Rahman and Mr Kasozi Brian

Department of Computer Science and Information Systems

Faculty of Science, Uganda Martyrs University

DECLARATION

I, Bogere Mark, hereby declare that the project titled "Improving Stock Price Prediction using Machine Learning: A Comparative Study of LSTM, CNN, and Traditional Methods" is the result of my own original research work conducted under the guidance and supervision of Dr Sanya Rahman and Mr Kasozi Brian.

Furthermore, I declare that this project has not been submitted in part or in whole, for any other degree, diploma, or academic recognition, nor has it been published or presented in any other forum without proper acknowledgement.

I affirm that this project is a genuine and authentic piece of work carried out with integrity, adhering to the principles of academic honesty and ethical research practices. I have duly acknowledged all the sources of information, data, and literature that have been utilized in this project, giving appropriate credit to the respective authors and researchers.

I acknowledge that the datasets and tools used in this project have been obtained and utilized in accordance with the relevant legal and ethical guidelines. I have taken the necessary measures to ensure data privacy, confidentiality, and security.

I am aware that this project may be subject to scrutiny and evaluation by the academic community. Therefore, I am open to sharing the methodologies, results, and insights of this research with other researchers, faculty members, or interested parties, in the pursuit of knowledge and collaboration.

In conclusion, I affirm that I have read and understood the guidelines and regulations of the institution regarding research projects, and I have complied with all the requirements stated therein while conducting this study.

Signature: ……………………………….

Date: …………………………………….

Name: ………………………………………

APPROVAL

We, Dr Sanya Rahman and Mr Kasozi Brian, wholeheartedly approve and commend Bogere Mark's exceptional research project. Mark has demonstrated an outstanding level of dedication, creativity, and analytical prowess throughout his work. His meticulous attention to detail and innovative approach have resulted in groundbreaking findings that contribute significantly to the field. We applaud Mark's comprehensive methodology, rigorous data analysis, and insightful conclusions. His research project is a testament to his intellectual rigor and academic excellence. We proudly endorse Bogere Mark's work and anticipate that his contributions will have a lasting impact on the scientific community.

Dr Sanya Rahman

Department of Computer Science and Information Systems

Faculty of Science

Uganda Martyrs University

Signature: ……………………………….

Date: …………………………………….

Mr Kasozi Brian

Head Department of Computer Science and Information Systems

Faculty of Science

Uganda Martyrs University

Signature: ……………………………….

Date: …………………………………….

DEDICATION

I dedicate this research project to the pursuit of knowledge and the endless quest for understanding. May it serve as a testament to the relentless curiosity, passion, and dedication that I embody in my academic journey. This work is dedicated to all the mentors, peers, and loved ones who have supported and inspired me along the way. May their unwavering belief in my abilities continue to fuel my intellectual growth and pave the way for groundbreaking discoveries.

ACKNOWLEDGMENT

I would like to express my deepest gratitude and appreciation to all those who have contributed to the successful completion of my research project titled "Improving Stock Price Prediction Using Machine Learning: A Comparative Study of LSTM, CNN, and Traditional Methods." This project has been a significant endeavour, and I could not have accomplished it without the support and assistance of several individuals.

First and foremost, I would like to thank my late Dad, Kakooza John Brown, whose unwavering support and encouragement played a pivotal role in my academic journey from primary to secondary education. His belief in the power of education and his constant motivation pushed me to excel and pursue higher studies. Though he is no longer with me, his memory and influence continue to inspire me.

I am also indebted to my research supervisor, Dr Sanya Rahman and Mr Kasozi Brian, for their invaluable guidance, expertise, and patience throughout the research process. Their insightful suggestions and constructive feedback were instrumental in shaping the direction and scope of this study.

Furthermore, I would like to express my gratitude to the faculty and staff of Science, Uganda Martyrs University for providing me with a conducive academic environment and access to the necessary resources. Their commitment to excellence and dedication to fostering a culture of learning has played a significant role in shaping my research abilities.

I would also like to extend my appreciation to my friends and colleagues especially Nalwanga Hafusa who provided support, encouragement, and valuable discussions during this research project. Their insights and diverse perspectives enriched my understanding of the subject matter and pushed me to explore new avenues.

Lastly, I am deeply grateful to all the participants who generously shared their time and insights for this study. Without their involvement, conducting the experiments and drawing meaningful conclusions would not have been possible.

In conclusion, I am humbled and honored to have had the opportunity to undertake this research project, and I extend my sincere thanks to everyone who played a part in its completion.

Table of Contents

LIST OF TABLES

ABSTRACT:

The stock market is known for its extreme complexity and volatility, and people are always looking for an accurate and effective way to guide stock trading. The accurate prediction of stock prices is of paramount importance in the financial market where returns and risks fluctuate wildly, and both financial institutions and regulatory authorities pay close attention to it. Stocks have always been favoured by investors as a method of asset allocation due to their higher returns. In recent years, researchers have been studying various methods to effectively predict stock market price and machine learning algorithms have emerged as one of the most promising techniques.

This paper proposes different methods for predicting stock market prices using machine learning architectures, with a focus on identifying latent dynamics in the data. Traditional methods, such as artificial neural networks, are also explored. The objective of the project is to improve the quality of the output of stock market predictions by using stock value as a predictor.

The paper presents a comparative study of machine learning architectures, including Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), and traditional methods like Arima, Random forest, logistic regression and K- nearest neighbors. The study analyzes historical stock data and compares the performance of each method based on various evaluation metrics. The results of the study demonstrate that LSTM and CNN outperform traditional methods in terms of accuracy, providing valuable insights for traders and investors.

Overall, this paper contributes to the existing literature on stock price prediction by providing a comprehensive analysis of machine learning methods for accurate stock price prediction. By identifying the best-performing machine learning architecture, this study can help traders and investors make more informed decisions and minimize financial risks in the volatile stock market.

ACRONYMS

ARIMA - Autoregressive Integrated Moving Average.

RNN - Recurrent Neural Network.

CNN - Convolutional Neural Networks.

LSTM - Long Short-Term Memory Networks.

GRU - Gated Recurrent Unit

MAE - Mean Absolute Error

MSE - Mean Squared Error

RMSE - Root Mean Squared Error

ML - Machine Learning

SVM - Support Vector Machine

EMH - Efficient Market Hypothesis

AI - artificial intelligence

NN - neural networks

RW - random walk

NSE - National Stock Exchange

CRNN - convolutional recurrent neural network

ANN - Artificial Neural Networks

EMH - efficient market hypothesis

PCA - Principal Component Analysis

GBM-wFE - Gradient Boosting Machine with Feature Engineering

PINSVR - parametric-insensitive nonparallel support vector regression

SDLC - Software Development Life Cycle

IDE - integrated development environment

ADF - Augmented Dickey-Fuller.

LB - Ljung-Box

UI - User Interface

RAM – Random Access Memory

CPU – Central Processing Unit

GB - Gigabyte

GHz – GigaHashes

OS – Operating System

SC - structure chart

UML - Unified Modeling Language

CHAPTER ONE: INTRODUCTION

Introduction

Financial market forecasting has traditionally been a focus of industry and academia. [1] The stock market, its volatility is complicated and nonlinear. [2] It is obviously unreliable and inefficient to rely solely on a trader's personal experience and intuition for analysis and judgment. In today's world, with the increasing level of trading and investing, individuals are searching for methods and tools to maximize their gains while minimizing risks. Predicting stock prices has long been a challenging task for researchers and investors are highly interested in this area of research. Effective prediction systems for the stock market help traders, investors, and analysts by providing supportive information such as the future direction of the stock market. [3]

Investors understand the importance of buying low and selling high, but this alone does not provide enough context to make informed investment decisions. Before investing in any stock, investors need to be aware of how the stock market behaves. Investing in a good stock at the wrong time can lead to disastrous results while investing in a mediocre stock at the right time can yield profits.

Time series prediction is a widely used technique in many real-world applications, including weather forecasting and financial market prediction. Recurrent Neural Networks (RNN) and their special types, such as Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU), are the most common algorithms used for time series prediction. In the past, several forecasting models have been developed for market prediction, including time series analysis, fundamental analysis, and technical analysis.

Time series analysis involves forecasting upcoming results based on past data. Fundamental analysis is based on the belief that companies that perform well should be rewarded with additional capital, leading to a surge in stock price. This analysis is commonly used by fund managers and is prepared from publicly available data like financial statement analysis. Technical analysis, on the other hand, is solely based on the trends of past values, and the long-run price of a stock is set based on these trends. However, these methods have limitations, and machine-learning approaches are becoming increasingly popular for predicting stock prices.

This study focuses on predicting the stock prices of Netflix using machine learning techniques, specifically comparing the effectiveness of LSTM, CNN, and traditional methods.

To achieve this goal, the study focuses on predicting the stock prices of Netflix using machine learning. Predicting the stock prices of companies like Netflix is a complex task, as stocks can fluctuate rapidly every hour based on world events. In this study, I aim to improve stock price prediction using machine learning. I will use a dataset covering 5 years from February 5, 2018, to February 5, 2022, to train and test my models. By analyzing and comparing the performance of these different methods, I hope to provide insights into the most effective approach for stock price prediction.

The study focuses on three main methods: Long-short Term Memory (LSTM), Convolutional Neural Networks (CNN), and traditional methods like ARIMA

LSTM is a type of RNN that has been proven to be effective in time-series prediction. CNNs are mainly used for image recognition tasks, but they can also be used for time-series prediction. Traditional methods, on the other hand, involve using statistical models to predict future prices based on historical data.

The study uses different evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to compare the effectiveness of the methods. By analyzing the results of the study, I hope to provide insights into the most effective approach for stock price prediction, which can be used by investors to make informed decisions about investing in the stock market.

Background

The stock market has always been closely watched by investors and stock forecasting has always been a research topic of great concern to researchers. In addition, the stock market is an important part of the country's financial market; it reflects the operation of the national economy, and the operation of the stock market has an important impact on the operation of the national economy [4]. Although the issue the of predictability of stocks has always been controversial, the study of stock forecasts still helps us understand the laws of market changes and development [5]. With the advancement of science and technology, large amounts of financial data have been retained giving provision for a solid data foundation for the analysis of the stock market. At times, the continuous development and updating of algorithms have provided a powerful tool for people to analyze the stock market [6].

The stock market provides a financing and investment environment for the country's companies and investors. Predicting the future performance of the stock market can not only provide investors with investment advice but also help companies formulate financing plans, thereby promoting the healthy development of the economy [7]. Establishing a stable investment portfolio based on the forecast results can help investors to further improve their investment returns [8].

The economy is developing rapidly, which makes the financial industry booming. Investors tend to pay more attention to the allocation of financial assets. In addition to savings and debt, relatively traditional investment and financial management methods such as securities and stocks have gradually become targets for investors. In human history, the first stock started in 1611 and was created in Amsterdam, Netherlands and the subject of the transaction is the East India Company in the Netherlands, which was established in 1602 [9].

The fact that investors adjust the allocation of investment assets to reduce their own decision-making risks, makes it important for them to predict the price of stocks. To accurately predict when and how to allocate asset budgets is difficult since there are many factors that can affect stock prices like the company’s allocation of assets [10]. Therefore, many investors have used technology and quantitative methods to try to predict the fluctuation of asset prices.

Applications

I. Business
II. Companies
III. Insurance company
IV. Government Agency

Project Motivation

Customer satisfaction and reviews are essential for businesses, and social media sentiment can affect stock markets. It is crucial for organizations to identify and address customer grievances to build trust and satisfaction. However, analyzing large volumes of customer feedback manually can be challenging and prone to errors and biases. To overcome this challenge, an unbiased automated system is necessary to classify customer reviews accurately. Sentiment analysis can help companies identify important issues from the customer's perspective and make data-driven decisions rather than relying solely on intuition.

Problem statement

The stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock. Investing in the stock market can be challenging for financial analysts who struggle to predict stock market behaviour accurately. They often face the problem of deciding which stocks to buy or sell to maximize profits due to the complexity and abundance of stock market data. Manual analysis of this data is almost impossible, and automation is essential. Machine-learning techniques have proven effective in predicting stock market behaviour by analyzing numerical time-series data. This project aims to address this issue by using Netflix data to predict stock price trends and notify investors of possible stock price movements. However, a gap exists between current practices and the need for automation, which this project aims to fill by implementing a graphical prediction model using Python. By automating the prediction process, our system will help financial analysts make informed decisions to maximize their chances of gaining a profit.

Objective

The main objective of this research is to develop an automated prediction model for stock market behaviour using machine learning techniques. The research aims to create a model that will help financial analysts make informed decisions on which stocks to buy or sell, based on predicted stock price trends.

Specific objectives

I. To collect and analyze data from Netflix to identify stock price trends.

II. To develop a prediction model using machine learning techniques to forecast the behaviour of the stock market.

III. To create a visualization of stock price predictions based on developed models.

IV. To evaluate the performance of the developed prediction model.

Significance

This research is significant because it will provide a solution to the problem faced by financial analysts who are not aware of the stock market behaviour, which affects their trading decisions. The development of an automated prediction model will provide a tool that can help financial analysts make informed decisions based on predicted stock price trends. The use of machine learning techniques will enable the prediction model to learn from historical data and improve its accuracy in predicting the behaviour of the stock market. The significance of this research lies in its potential to help investors maximize their chances of gaining a profit, which is important for the growth of the economy.

CHAPTER TWO-LITERATURE REVIEW

Introduction

The thinking of different people has always been an important piece of information during the decision-making process. The Internet and the Web made it possible to find out different opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics, people I have never heard of. Conversely, more and more people are making their opinions available to strangers via the Internet. The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is a driving force for this area of interest. There are many challenges involved in this process, which need to be walked all over in order to attain proper outcomes out of them. In this literature review, I analyzed the basic methodology that usually happens in this process and measures to be taken to overcome the challenges.

Existing/ related systems

Stock Market Prediction Using Machine Learning. [11] In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock or other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental time series analysis is used by most stockbrokers while making stock predictions. The programming language used to predict the stock market using machine learning is Python. In this paper, I propose a Machine Learning (ML) approach that will be trained from the available stock data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context, this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for large and small capitalizations in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

Forecasting the Stock Market Index Using Artificial Intelligence Techniques. [12] A dissertation was submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering. The weak form of the Efficient Market hypothesis (EMH) states that it is impossible to forecast the future price of an asset based on the information contained in the historical prices of an asset. This means that the market behaves as a random walk and as a result makes forecasting impossible. Furthermore, financial forecasting is a difficult task due to the intrinsic complexity of the financial system. The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the future price of a stock market index based on its historical price information. Artificial intelligence techniques have the ability to take into consideration financial system complexities and they are used as financial time series forecasting tools. Two techniques are used to benchmark the AI techniques, namely, Autoregressive Integrated Moving Average (ARIMA) which is a linear modelling technique and the random walk (RW) technique. The experimentation was performed on data obtained from the Johannesburg Stock Exchange. The data used was a series of past closing prices of the All Share Index. The results showed that the three techniques have the ability to predict the future price of the Index with acceptable accuracy. All three artificial intelligence techniques outperformed the linear model. However, the random walk method outperformed all the other techniques. These techniques show an ability to predict the future price however, because of the transaction costs of trading in the market, it is not possible to show that the three techniques can disprove the weak form of market efficiency. The results show that the ranking of performances support vector machines, neuro-fuzzy systems, and multilayer perceptron neural networks is dependent on the accuracy measure used.

Multi-Category Events Driven Stock Price Trends Prediction. [13] In this paper, multi-category news events are used as features to develop a stock price trend prediction model. The multi-category events are based on an already-defined feature word dictionary. And we have employed both neural networks and SVM models to analyze the relationship between stock price movements and specific multi-category news. Experimental results showed that the predefined multi-category news events are more improved than the baseline bag-of-words feature to predict stock price trends. As compared to long-term prediction, short-term prediction is better based on this study.

Indian stock market prediction using artificial neural networks on tick data. [14] A stock market is a platform for trading a company's stocks and derivatives at an agreed price. Supply and demand of shares drive the stock market. In any country stock market is one of the most emerging sectors. Nowadays, many people are indirectly or directly related to this sector. Therefore, it becomes essential to know about market trends. Thus, with the development of the stock market, people are interested in forecasting stock prices. But, due to the dynamic nature, a liable to quick changes in stock price, prediction of the stock price becomes a challenging task. Stock m Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lack common-sense knowledge, such as the intents and emotions of the event participants, which are useful for distinguishing event pairs when there are only subtle differences in their surface realizations. To address this issue, this paper proposes to leverage external common-sense knowledge about the intent and sentiment of the event. Experiments on three event-related tasks, i.e., event similarity, script event prediction and stock market prediction, show that our model obtains much better event embedding for the tasks, achieving 78% improvements on hard similarity tasks, yielding more precise inferences on subsequent events under given contexts, and better accuracies in predicting the volatilities of the stock market1. Markets are mostly nonparametric, non-linear, noisy and deterministic chaotic systems. As technology is increasing, stock traders are moving towards using Intelligent Trading Systems rather than fundamental analysis for predicting the prices of stocks, which helps them to take immediate investment decisions. One of the main aims of a trader is to predict the stock price such that he can sell it before its value decline, or buy the stock before the price rises. The efficient market hypothesis states that it is not possible to predict stock prices and that stock behaves in a random walk. It seems to be very difficult to replace the professionalism of an experienced trader in predicting the stock price. But because of the availability of a remarkable amount of data and technological advancements we can now formulate an appropriate algorithm for prediction whose results can increase the profits for traders or investment firms. Thus, the accuracy of an algorithm is directly proportional to gains made by using the algorithm.

Share Price Prediction using Machine Learning Technique. [15] This paper is mostly based on the approach of predicting the share price using Long Short Term Memory (LSTM) and Recurrent Neural Networks (RNN) to forecast the stock value on NSE data using various factors such as current market price, price-earnings ratio, base value and other anonymous events. The efficiency of the model is analyzed by comparing the true data and the predicted data using an RNN graph. Machine learning to predict stock price as seen in the model is able to predict the stock price very close to the actual price where this model captures the detailed feature and uses different strategies to make a prediction. The model trained for all the NSE data from the internet and recognize the input and group them and provide input according to the user configuration this RNN based architecture proved very efficient in forecasting the stock price by changing the configuration accordingly which also uses a backpropagation mechanism while gathering and grouping data to avoid mixing of data.

Automated Stock Price Prediction Using Machine Learning. [16] Traditionally and in order to predict market movement, investors used to analyze the stock prices and stock indicators in addition to the news related to these stocks. Hence, the Importance of news on the stock price movement. Most of the previous work in this industry focused on either classifying the released market news as positive, negative, or neutral and demonstrating their effect on the stock price or focused on the historical price movement and predicted their future movement. In this work, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted, the highest accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

Forecasting stock price in two ways based on LSTM neural network. [17] The LSTM neural network is used to predict Apple stocks by consuming single-feature input variables and multi-feature input variables to verify the forecast effect of the model on stock time series. The experimental results show that the model has a high accuracy of 0.033 for the multivariate input and is accurate, which is in line with the actual demand. For the univariate feature input, the predicted squared absolute error is 0.155, which is inferior to the multi-feature variable input.

Stock Price Correlation Coefficient Prediction with ARIMA LSTM Hybrid Model. Predicting the price correlation of two assets for future time periods is important in portfolio optimization [18]. We apply LSTM recurrent neural networks (RNN) in predicting the stock price correlation coefficient of two individual stocks. RNNs are competent in understanding temporal dependencies. The use of LSTM cells further enhances its long-term predictive properties. To encompass both linearity and nonlinearity in the model, we adopt the ARIMA model as well. The ARIMA model filters linear tendencies in the data and passes on the residual value to the LSTM model. The ARIMA-LSTM hybrid model is tested against other traditional predictive financial models such as the full historical model, constant correlation model, single-index model and multi-group model. In our empirical study, the predictive ability of the ARIMA-LSTM model turned out superior to all other financial models by a significant scale. Our work implies that it is worth considering the ARIMALSTM model to forecast the correlation coefficient for portfolio optimization.

Share Price Trend Prediction Using CRNN with LSTM Structure. The entire financial market majorly runs by the stock market and one of the most attractive research issues is predicting stock price volatility [19]. The information of historical stocks for assuming the future stock price as well deep learning method is applied to find the approximate trend value of stock prices which are mentioned in this paper. This paper not only stores the data of historical stock with the time scale but also estimates prices of the future stock by a designed neural network, this is due to the fact that the trend of stocks is usually connected to the previous information of stock price. In this paper, the design of the neural network proposed then with the memory performance of the convolutional recurrent neural network (CRNN) and for improving the long-term dependency of traditional RNN the Long Short-term memory (LSTM) are the major components. Also to enhance the accuracy as well as stability of prediction of the RNN LSTM architecture is put. This paper accumulated a total of ten stock historic data to test and accomplish an average error rate of 3.449 RMSE.

Stock Price Prediction Based on Information Entropy and Artificial Neural Network. The research done [20], shows that one of the most important components of the financial system is the stock market. Supporting the activity and evolvement, money is directed by the investors of the associated firm. Along with information theory and Artificial Neural Networks (ANN), the combination of machine learning framework is formed. Information entropy for non-linear causality and stock relevance also to facilitate ANN time series modelling are creatively used by this method. The feasibility of this machine learning framework is analyzed with Amazon, Apple, Google and Facebook prices. A time series analysis method based on information theory, as well as LSTM to model the stock price dynamics, are outlined in the paper. The transfer entropy between relevant variables to help LSTM time series prediction is merged in this modelling infrastructure, thus the accuracy of the assumption outcome is broadly granted. Modelled and real stock price is highly correlated while differing slightly in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) which are investigated by the outcomes.

The Stock Market and Investment. Investigating the relation between equity prices and aggregate investment in major European countries including France, Germany, Italy, the Netherlands and the United Kingdom [21]. Increasing integration of European financial markets is likely to result in an even stronger correlation between equity prices in different European countries. This process can also lead to convergence in economic development across European countries if developments in stock markets influence real economic components, such as investment and consumption. The vector autoregressive models suggest that the positive correlation between changes in equity prices and investment is in general significant. Hence monetary authorities should monitor the reactions of share prices to monetary policy and their effects on the business cycle.

An innovative neural network approach for stock market prediction. [22] To develop an innovative neural network approach to achieving better stock market predictions. Data were obtained from the livestock market for real-time and offline analysis and results of visualizations and analytics to demonstrate the Internet of Multimedia of Things for stock analysis. To study the influence of market characteristics on stock prices, traditional neural network algorithms may incorrectly predict the stock market since the initial weight of the random selection problem can be easily prone to incorrect predictions. Based on the development of word vectors in deep learning, demonstrate the concept of "stock vector." The input is no longer a single index or single stock index, but multi-stock high-dimensional historical data. Propose the deep long short-term memory neural network (LSTM) with embedded layer and the long short-term memory neural network with automatic encoder to predict the stock market. In these two models, we use the embedded layer and the automatic encoder, respectively, to vectorize the data, in a bid to forecast the stock via long short-term memory neural network. The experimental results show that the deep LSTM with an embedded layer is better. Specifically, the accuracy of the two models is 57.2 and 56.9%, respectively, for the Shanghai A-shares composite index. Furthermore, they are 52.4 and 52.5%, respectively, for individual stocks. Demonstrate research contributions in IMMT for neural network-based financial analysis.

Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models. According to this study [23], the Prediction of stock prices has been an important area of research for a long time. While supporters of the efficient market hypothesis believe that it is impossible to predict stock prices accurately, there are formal propositions demonstrating that accurate modelling and designing of appropriate variables may lead to models using which stock prices and stock price movement patterns can be very accurately predicted. It also states that researchers have worked on technical analysis of stocks with the goal of identifying patterns in stock price movements using advanced data mining techniques. The work of this study proposes an approach of hybrid modelling for stock price prediction building different machine learning and deep learning-based models using long-and short-term memory (LSTM) networks with a novel approach of walk-forward validation and exploits the power of LSTM regression models in forecasting the future NIFTY 50 open values.

A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models. The research done by [24], indicates that the Prediction of future movement of stock prices has always been a challenging task for researchers and while the advocates of the efficient market hypothesis (EMH) believe that it is impossible to design any predictive framework that can accurately predict the movement of stock prices, there is seminal work in the literature that have demonstrated that time series of a stock price can be predicted with a high level of accuracy. This work proposes that the agglomerative approach of model building that uses a combination of statistical, machine learning, and deep learning approaches can very effectively learn from the volatile and random movement patterns in stock price data.

A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing. According to [25], shows that Predicting the future movement of stock prices has been a subject matter of many research works and there is a gamut of literature on technical analysis of stock prices where the objective is to identify patterns in stock price movements and derive profit from it. It also indicates that improving the prediction accuracy remains the single most challenge in this area of research therefore proposing a hybrid approach for stock price movement prediction using machine learning, deep learning, and natural language processing where the authors select the NIFTY 50 index values of the National Stock Exchange (NSE) of India, and collect its daily price movement over a period of three years (2015-2017)

Cross-sectional Stock Price Prediction using Deep Learning for Actual Investment Management. The research work done by [26] shows that Stock price prediction is an important research theme both academically and practically where various methods to predict stock prices have been studied until now. The feature that explains the stock price by a cross-section analysis is called a "factor" in the field of finance and many empirical studies in finance have identified stocks having features in the cross-section relatively increase and decrease in terms of price hence presenting a cross-sectional daily stock price prediction framework using deep learning for actual investment management, which build a portfolio with information available at the time of market closing and invest at the time of market opening the next day and performs empirical analysis in the Japanese stock market to confirm the profitability of the framework.

A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-wFE). In this research [27], Malaysia propose a novel multiclass classification ensemble learning approach called Gradient Boosting Machine with Feature Engineering (GBM-wFE) and Principal Component Analysis (PCA) as the feature selection and finds out that GBM-wFE outperforms the previous studies and the overall prediction results are auspicious.

Stock price analysis using machine learning method (Non-sensory-parametric backup regression algorithm in linear and nonlinear mode). According to the author [28], The most common starting point for investors when buying a stock is to look at the trend of price changes. Different models have been used to predict stock prices by researchers, and since artificial intelligence techniques, including neural networks, genetic algorithms and fuzzy logic, have achieved successful results in solving complex problems, the research showed that the PINSVR algorithm in nonlinear mode has been able to predict the stock price over the years, rather than linear mode.

Analysis of Stock Market Manipulation using Generative Adversarial Nets and Denoising Auto-Encode Models. Market manipulation remains the biggest concern of investors in today's securities market and the development of technologies [29] and complex trading algorithms seems to facilitate stock market manipulation and make it inevitable for regulators to use Deep Learning models to prevent manipulation hence showing that GAN-DAE4 outperformed other deep learning models (with F2-measure 73.71%) such as Decision Tree, Random Forest, Neural Network, and Logistic Regression.

Conclusion

In conclusion, I have reviewed various approaches for Stock price prediction and all approaches have their own advantages and disadvantages. CNN, ARIMA & LSTM are the most popular algorithms used in the prediction of stock price but there are some challenges in these methods like the need for a lot of training data, High computational cost, without GPU data quite slow to train, depending on any previous information for prediction. A hybrid approach can be used to overcome these issues. While machine learning is able to provide highly accurate prediction results using standard tools and also outperforms all standard prediction methods.

CHAPTER THREE. METHODOLOGY

Introduction

Stock market prediction technology has great economic value for stock market investors and investment institutions, helping investors and investment institutions to make profits and avoid investment risks. From a social perspective, stock market forecasting technology can prevent systemic risks in the financial market, help rationally allocate social funds, and contribute to the harmonious and stable development of the economy. Stock data has its own characteristics, and the existing forecasting technology methods are not fully used. The multi-scale and multi-source heterogeneous prediction technology can not only be used for stock market prediction but also has broad application prospects in many fields such as personal health state prediction, energy demand prediction and website traffic prediction.

The multi-scale property of stock market data refers to the existence of data at different time intervals, and the data at different scales will reflect the stock movement state of different time periods. Large-scale stock market data can reflect the long-term movement state of the stock market, and small-scale stock market data can reflect the short-term movement state of the stock market. Data of different scales have associated information and their own unique information. In order to describe the current market state more accurately, it is necessary to comprehensively consider stock market data of multiple scales. However, most of the existing research only focuses on the single-scale data of the stock market. This can lead to less-than-expected forecast performance due to an inaccurate description of the state of the stock market.

Data source

The dataset utilized for analysis was selected from Netflix and it runs for 5 years i.e. from 5th Feb 2018 to 5th Feb 2022. It consists of 7 records of the required Stock price. The data reflected the stock price at some time intervals for every day of the year. It contains various data like date, Adj_Close, open, close, low, high and volume.

Close: The last price of the stock before the trading day ends.

Open: The price of the stock at the beginning of the trading day.

High: The highest price of the stock on the trading day.

Low: The last price of the stock before the trading day ends.

The data for only one company was considered. All the data was available in a file of CSV format which was first read and transformed into a data frame using the Pandas library in Python.

Methodology/ project work

The volatility of stock prices is controlled by the trend of the stock but is also sensitive to many other factors. Due to the relative stability and predictability of the intrinsic value of stocks, the factors that have impacts on the stock market price mainly include the following;

I. Macro factors
II. Industrial and regional factor
III. Company factors
IV. market factors.

This research project predicts the closing index of Netflix rather than specific company stock price forecasts. The more microscopic industry and company factors mainly focus on the influence of macroeconomic factors and market factors. Macroeconomic factors refer to the impact of the macroeconomic environment and its changes on stock prices, including regular factors such as cyclical fluctuations in macroeconomic operations and policy factors such as monetary policy implemented by the government. There are two types of stock price forecasting methods; qualitative analysis and quantitative analysis. The qualitative analysis method is the fundamental analysis method which is a subjective analysis method relying on the experience of financial practitioners. This thesis is a numerical prediction of the daily closing index of Netflix rather than a trend judgment of price fluctuations, hence focusing on the literature review of quantitative analysis methods.

Numerical data-based stock market forecasting research uses numerical data on certain time scales in the stock market, such as sky-level index prices and stock price volume data, to predict specific stocks or other investments in the stock market on the same scale to Predict the future price of the underlying. According to the focus of this research, these studies can be divided into research on the characteristics of numerical data stock market forecasting and research on the numerical data stock market forecasting model. In order to build this model, in addition to the traditional ARIMA model, the LSTM and CNN models will be used. The models will use 70%of the data for training, and the remaining 30% of the data is used for testing. For training, we use Root Mean Square Error and Adam algorithm to optimize the model.

Development methodology

In developing this project, the development cycle or strategy that will be used is the Waterfall Model Approach. This methodology is the soonest SDLC (Software Development Life Cycle) approach in programming or web advancement. The waterfall approach is one of two well-known strategies to handle programming projects. The other technique is known as Agile. It tends to be a more obvious cascade when you analyze it excessively Agile. Waterfall and Agile are two totally different tasks the management methodologies, however, both are similarly, legitimate and can be valuable relying upon the undertaking.

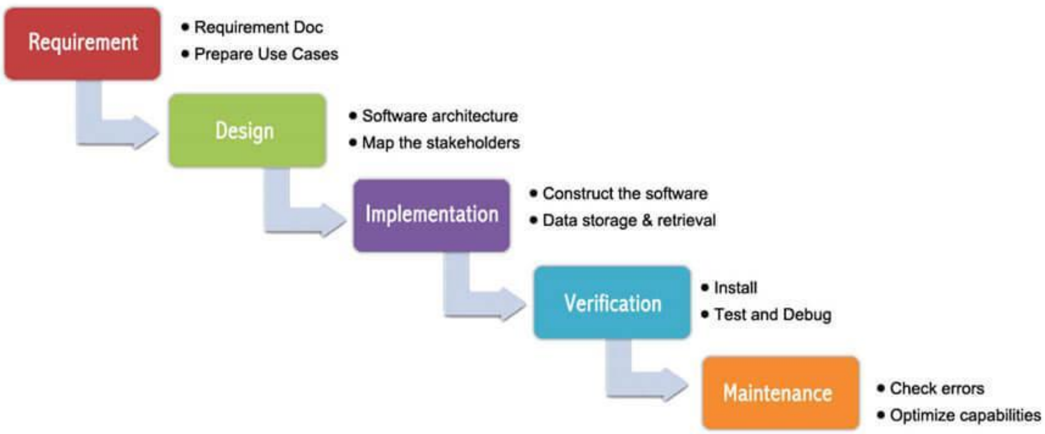


Figure 1: waterfall methodology

I. Requirements

The Waterfall methodology relies upon the conviction that all project necessities can be accumulated and perceived forthright. The project author gives a valiant effort to get a definite comprehension of the task support's prerequisites. Composed prerequisites, normally contained in a solitary archive, are utilized to depict each phase of the venture, including the expenses, presumptions, hazards, conditions, achievement measurements, and courses of events for consummation.

II. Design

Programming engineers plan a specialized answer for the issues set out by the item prerequisites, including situations, formats, and information models. A more elevated level or coherent plan is made that portrays the reason and extent of the task, the overall traffic stream of every part, and the reconciliation focuses. When this is finished, it is changed into an actual plan utilizing explicit equipment and programming advancements.

III. Implementation

When the plan is finished, specialized execution begins. This may be the briefest period of the Waterfall interaction on the grounds that careful examination and configuration have as of now been finished. In this stage, software engineers code applications dependent on project necessities and details, with some testing and execution occurring too. In the event that critical changes are needed during this stage, this might mean returning to the planning stage.

IV. Verification

Before a project can be delivered to clients, testing should be done to guarantee the item has no blunders and every one of the necessities has been finished, guaranteeing a decent client experience with the product. The testing group will go to the plan archives, personas, and client case situations provided by the item director to make their experiments.

V. Deploy and Maintenance

When the product has been conveyed on the lookout or delivered to clients, the upkeep stage starts. As deformities are found and change demands roll in from clients, a group will be allocated to deal with updates and discharge new forms of the product.

Development algorithms.

LSTM network. A long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

Working of LSTM.

LSTM is a special network structure with three "gate" structures. Three gates are placed in an LSTM unit, called the input gate, forgetting gate and output gate. While information enters the LSTM's network, it can be selected by rules. Only the information that conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate. The experimental data in this paper are the actual historical data downloaded from the Internet.

Illustration

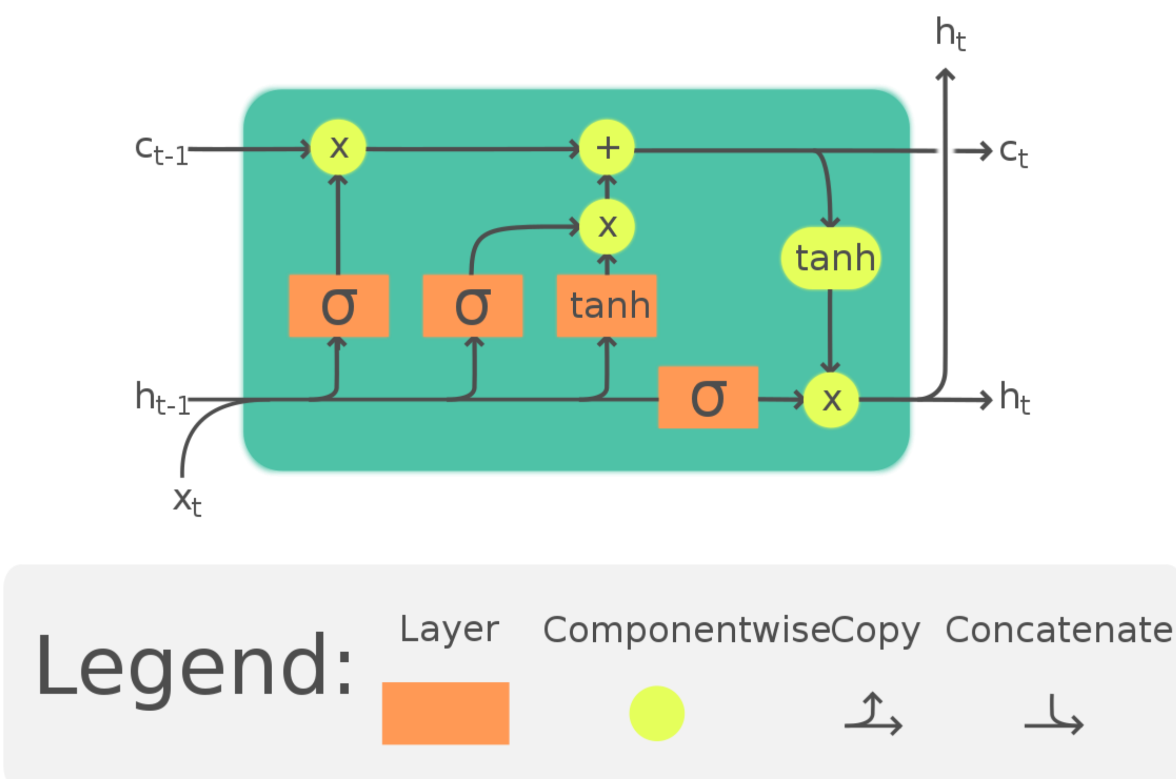


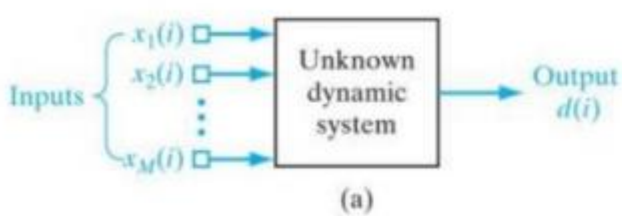
Figure 2: LSTM Architecture

Long Short-Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input within the present step. It tackled the matter of long-term dependencies of RNN within which the RNN will not predict the word hold on within the long-term memory however can offer additional accurate forecasts from the recent info. Because the gap length will increase RNN does not offer an economical performance. LSTM will by default retain the knowledge for a long period of time. It is used for processing, predicting and classifying on the basis of time-series data.

LMS filter**.** The LMS filter is a kind of adaptive filter that is used for solving

linear problems. The idea of the filter is to minimize a system by minimizing the least mean square of the error signal.

Illustration



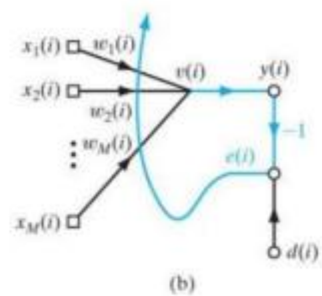


Figure 3: LMS Inputs and Outputs

Figure 4: LMS updating weights

Algorithm 1: LMS

Input:
x : input vector
d: desired vector
μ: learning rate
N: filter order
Output:
y: filter response
e: filter error
begin
  $M = size(x)$;
  $x_n(0) = w_n(0) = [0\ 0\ \dots\ 0]^T$;
  while $n < M$ **do**
    $x_{n+1} = [x(n);\ x_n(1:N)]$;
    $y(n) = w_n^H * x_n$;
    $e(n) = d(n) - y(n)$;
    $w_{n+1} = w_n + 2\mu e(n) x_n$;
  end
end

Figure 5: Lstm algorithm 1

Generally, I don’t know exactly if the problem can be solved very well with a linear approach, so I tested a linear and a non-linear algorithm. Since the internet always shows non-linear approaches, I used LMS to prove that stock market prediction can be done with linear algorithms with good precision. But this filter monetizes a system, so if I apply this filter in the data, will have the filter coefficients trained, and when input a new vector, the filter coefficients will output a response that the original system would.

The system. I will delay my input vector by l positions, where l would be the number of days we want to predict, this l new positions will be filled by zeros.

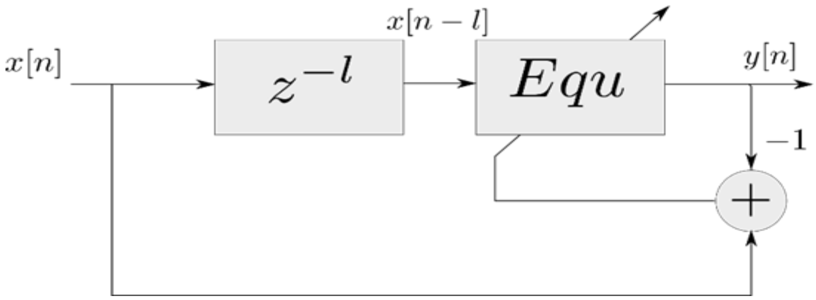


Figure 6: LMS updating weights

When I apply the LMS filter, I train the filter to the first 178 data. After that, I set the error as zero, so the system will start to output the answers as the original system to the last l values.

Algorithm 2: LMSPred

Input:
x: input vector
l: quantity of days to predict
μ: learning rate
N: filter order
Output:
y: filter response
begin
  $M = size(x_d)$;
  $x_n(0) = w_n(0) = [0\ 0\ \ldots\ 0]$;
  $x_d = [0\ 0\ ..\ 0\ x]$;
  while $n < M$ **do**
    $x_{n+1} = [x_d(n);\ x_n(1:N)]$;
    $y(n) = w_n^H * x_n$;
    if $n > M - l$ **then**
      $e = 0$;
    else
      $e(n) = d(n) - y(n)$;
    end
    $w_{n+1} = w_n + 2\mu e(n) x_n$;
  end
end

Figure 7: Lstm pred

Forget Gate. A forget gate is responsible for removing information from the cell state. The information that is no longer required for the LSTM to understand things or the information that is of less importance is removed via the multiplication of a filter. This is required for optimizing the performance of the LSTM network. This gate takes in two inputs; h_t-1 and x_t. h_t-1 is the hidden state from the previous cell or the output of the previous cell and x_t is the input at that particular time step.

Input Gate. Regulating what values need to be added to the cell state by involving a sigmoid function. This is basically very similar to the forget gate and acts as a filter for all the information from hi-1 and x_t. Creating a vector containing all possible values that can be added (as perceived from h_t-1 and x_t) to the cell state. This is done using the tanh function, which outputs values from -1 to +1. Multiplying the value of the regulatory filter (the sigmoid gate) to the created vector (the tanh function) and then adding this useful information to the cell state via addition operation.

Input Gate. The functioning of an output gate can again be broken down into three steps: Creating a vector after applying the tanh function to the cell state, thereby scaling the values to the range -1 to +1. Making a filter using the values of h_t-1 and x_t, such that it can regulate the values that need to be output from the vector created above. This filter again employs a sigmoid function. Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as output and also to the hidden state of the next cell.

Applications of LSTM include

I. Language Modelling
II. Machine Translation
III. Image Captioning
IV. Handwriting generation
V. Question Answering Chatbot

CNN Network.

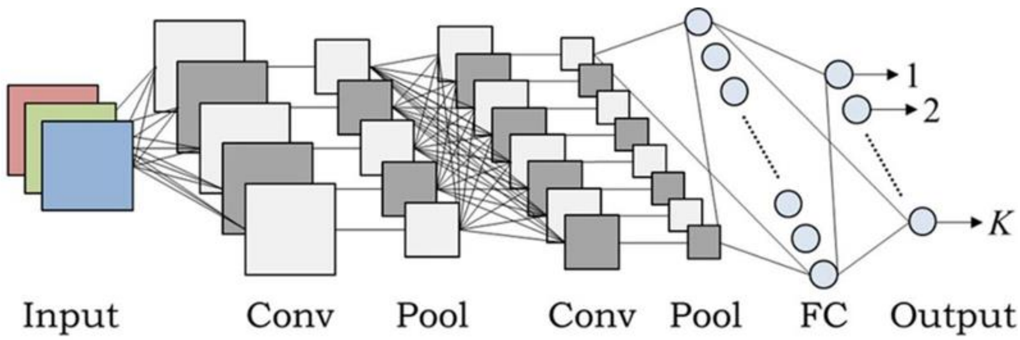


Figure 8: CNN Architecture

A layer of CNN model:

I. Convolution. This extracts the features from the input image given the output in matrix form.
II. MAX Pooling. It takes the largest element from a rectified feature map.
III. Dropout. Randomly selected neurons are ignored during training.
IV. Flatten. Feeds output into a fully connected layer. It gives data in list form.
V. Dense. A Linear operation in which every input is connected to every output by weight. Its followed by a nonlinear activation function.
VI. Activation. It's used a sigmoid function and predicts the probability of 0 and 1.

Applications of CNN include

I. Decoding Facial Recognition
II. Analyzing Documents

Root Mean Square Error (RMSE)

Root Mean Square Error (Root Mean Square Error) can be used to calculate the deviation between the observed value and the true value. Because the average index is non-robust, this makes the average error very sensitive to outliers.

Mean Square Error (MSE): Measures the average of the squares of the errors that is, the average squared difference between the estimated values and the actual value. MSE is a risk function, corresponding to the expected value of the squared error loss

Mean Absolute Error (MAE): A measure of errors between paired observations expressing the same phenomenon.

Hybrid (LSTM – CNN) model.

In the hybrid approach, the Convolutional Neural Networks (CNNs) offer benefits in choosing sensible options and Long Short-Term Memory (LSTM) networks have proven sensible skills to find out to learn sequential data. Each approach is reported to produce an improved result. CNNs possess convolute filters over every input layer so as to get the simple options and CNNs have shown enhancements in computer vision, natural language processing and different tasks. CNN may be a powerful tool to pick out features in order to improve prediction accuracy. The capabilities of LSTMs in learning data series by considering the previous outputs.

The multiple convolutional filters slide over the matrix to produce a new feature map and also the filters have numerous completely different sizes to generate different features. The Max- pooling layer is to calculate the most value as a corresponding feature to a particular filter. The output vectors of the Max-pooling layer become inputs to the LSTM networks to measure the long-run dependencies of feature sequences. One of the benefits of LSTMs is the ability to capture sequential data by considering the previous data. This layer takes the output vectors from the dropout layer as inputs. This layer includes a set number of units or cells and also the input of every cell is the output from the dropout layer. The final output of this layer has the same number of units within the network the outputs from LSTMs are merged and combined in one matrix and then passed to a fully connected layer. The array is converted into a single output in the range between 0 and 1 using the fully connected layer, in order to be finally classified using the sigmoid function.

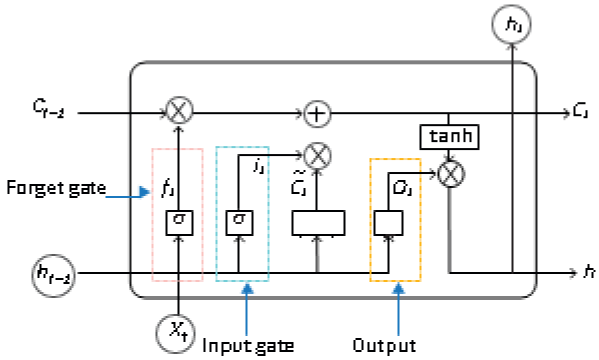


Figure 9: CNN-LSTM structure diagram.

CNN-LSTM Training and Prediction Process.

The CNN-LSTM process of training and prediction is shown in Figure 8.

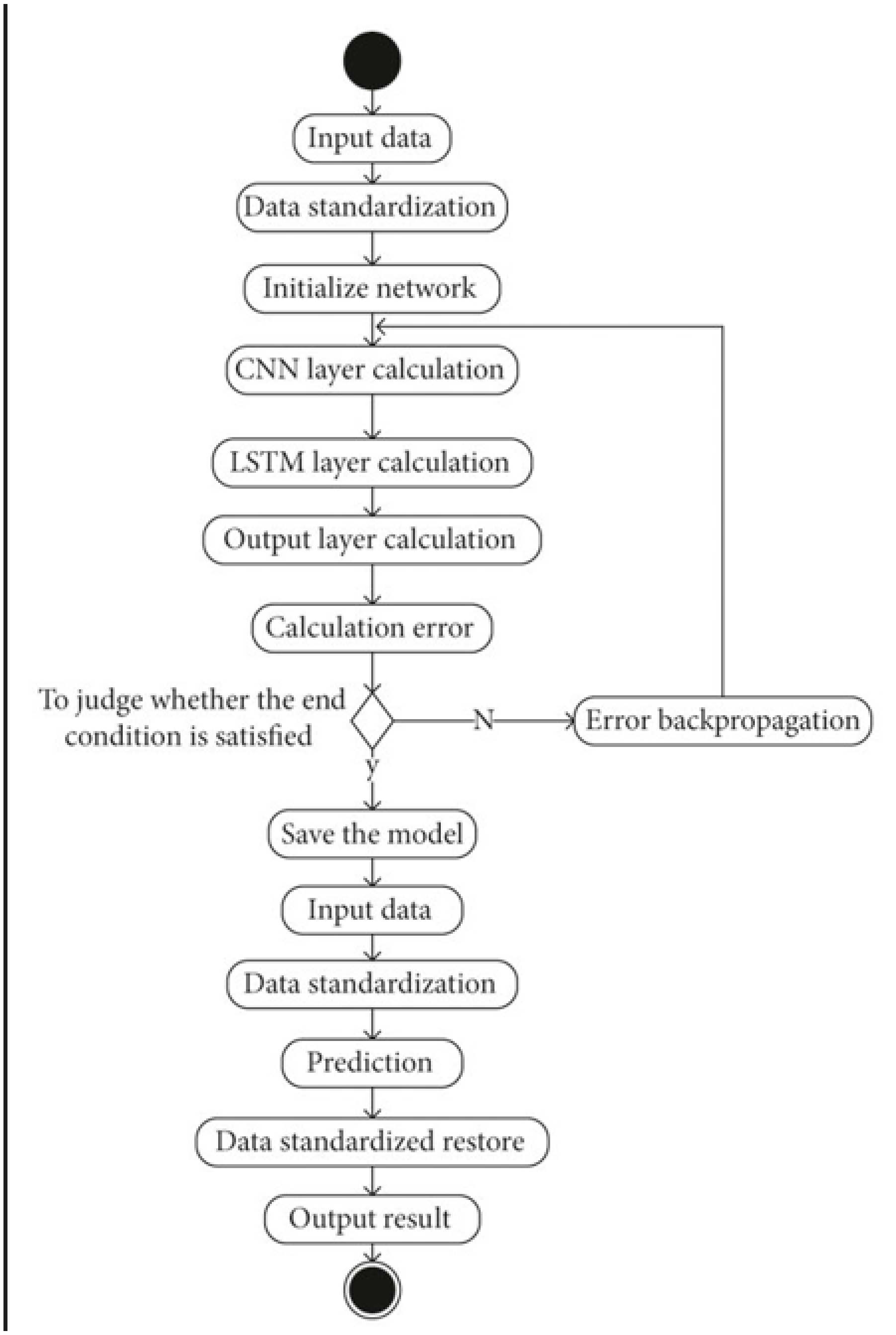


Figure 10: Activity diagram of CNN-LSTM training and prediction process.

I. Input data: input the data required for CNN-LSTM training.

II. Data standardization: as there is a large gap in the input data, in order to train the model better, the z- score standardization method is adopted to standardize the input data.

III. Initialize network: initialize the weights and biases of each layer of the CNN-LSTM.

IV. CNN layer calculation: the input data are successively passed through the convolution layer and pooling layer in the CNN layer, the feature extraction of the input data is carried out, and the output value is obtained.

V. LSTM layer calculation: the output data of the CNN layer are calculated through the LSTM layer, and the output value is obtained.

VI. Output layer calculation: the output value of the LSTM layer is input into the full connection layer to get the output value.

VII. Calculation error: the output value calculated by the output layer is compared with the real value of this group of data, and the corresponding error is obtained.

VIII. To judge whether the end condition is satisfied: the conditions for the end are to complete a predetermined number of cycles, the weight is lower than a certain threshold, and the error rate of the forecasting is lower than a certain threshold. If one of the conditions for the end is met, the training will be completed, update the entire CNN-LSTM network, and go to step 10; otherwise, go to step 9.

IX. Error backpropagation: propagate the calculated error in the opposite direction, update the weight and bias of each layer, and go to step 4 to continue to train the network.

X. Save the model: save the trained model for forecasting.

XI. Input data: input the input data required for the forecasting.

XII. Forecasting: input the standardized data into the trained model of CNN-LSTM, and then get the corresponding output value.

XIII. Data standardized restore: the output value obtained through the model of CNN-LSTM is the standardized value, and the standardized value is restored to the original value.

XIV. Output result: output the restored results to complete the forecasting process.

ARIMA model.

As the stock data is noisy, we must first perform a stationarity test on the stock sequence. The test method is to observe the sequence diagram, autocorrelation diagram, and partial autocorrelation diagram of the sequence first, and then do a unit root (ADF) test to test its P If the sequence is non-stationary, then choose the difference for smoothing. After determining the order of the difference, confirm that it is a stationary sequence, which can be used to determine the order of the model, that is, p, q. After the determination is completed, the model is tested, mainly the LB test, to confirm whether the residual is white noise. If it is, then the model passes the test and we can make predictions.

Development tools.

The development tool that will be used for this improvement of this undertaking is Microsoft visual studio. It will be utilized as the IDE for the development of the machine learning prediction model. It is suitable and it is open source, so it is free. Not just that, Microsoft Visual Studio is likewise viable to most programming dialects out there. This is an extraordinary IDE as the improvement of this undertaking utilizes different programming dialects to begin coding. The programming dialects include Python for machine learning model advancement and Dash framework from Python for the User Interface (UI) of this model.

Figure 11: Microsoft Visual Studio Logo

Jupyter notebook

The Jupyter Notebook is an open-source web application that enables to making and sharing of documents that contain visualizations, narrative text, live code and equations. Uses include data visualization, data transformation, statistical modelling, machine learning, numerical simulation, and data cleaning.

Figure 12: Jupyter

Hardware Requirements

I. RAM: 4 GB and above
II. Storage: 500 GB
III. CPU: 2 GHz or faster
IV. Architecture: 32-bit or 64-bit

Software Requirements

I. Python 3.5 in Microsoft visual studio is used for data pre-processing, model training and prediction.
II. Operating System: Windows 7 and above or Linux-based OS or MAC OS

Functional requirements

Functional requirements describe what the software should do (the functions).

I. The software shall accept the NFLX.csv dataset as input.
II. The software shall do pre-processing (verifying for missing data values) on input for model training.
III. The software shall use LSTM ARCHITECTURE as the main component of the software.
IV. It processes the given input data by producing the most possible outcomes of a CLOSING STOCK PRICE.

Non-Functional requirements

I. Usability: It defines the user interface of the software in terms of simplicity of understanding the user interface of stock prediction software, for any kind of stock trader and other stakeholders in the stock market.

II. Efficiency: maintaining the possible highest accuracy in the closing stock prices in the shortest time with available data.

III. Performance: It is a quality attribute of the stock prediction software that describes the responsiveness to various user interactions with it.

Data abstraction.

Data abstraction is a step of finding the resource to the very best in categorizing the datasets and learning the best out of it. It is to customize the datasets in finding the best constraints to consider and the unwanted resources are the dump which will be dumped. Datasets are cleared on this progression for the starting system. The significant dataset is the set that carries the worth to the dataset for a superior agreement and gives a superior yield and creation by assessing something very similar. This is a component deliberation model to remove the inclusion of the dataset. This is a component model cycle where every one of the plausible assets is arranged and a similar one will be used for the featuring.

Training and testing the dataset for prediction.

The following stage is to split the dataset into training and test sets to keep away from overfitting and to have the option to examine the speculation capacity of the model. Officially, overfitting alludes to the circumstance where a model learns the data yet, in addition, the noise that is important for preparing data to the degree that it contrarily impacts the presentation of the model on new unseen data. In other words, the noise mentioned above (for example random fluctuations) in the preparation set is learned as rules/patterns by the model. In any case, these loud learned portrayals don't matter to new unseen data and consequently, the model's presentation (for example accuracy, MSE, MAE) is adversely affected. MSE, (Mean Squared Error), MAE, (Mean Absolute error).

The "'target" value to be predicted in the model will be the " Adj Close" price stock price value. It's a smart thought to standardize the dataset before model fitting. This will support the performance. While Implementing any LSTM, we ought to consistently reshape our X train in dimension, add 1 to the purpose for the time step and the 1 is given to the LSTM.

trainX =np.array(X_train)

testX =np.array(X_test)

X_train = trainX.reshape(X_train.shape[0], 1, X_train.shape[1])

X_test = testX.reshape(X_test.shape[0], 1, X_test.shape[1])

Then, at that point, import the required modules for the stacked LSTM. Sequential for initializing the neural network, Dense for adding a densely connected neural network layer, LSTM for adding the Long Short-Term Memory layer, and Dropout for adding dropout layers that prevent overfitting. The modules are tensorflow.keras.models, tensorflow.keras.layers, tensorflow.keras.layers import LSTM. Then, utilize a successive model and add the layers of the LSTM as said, in the above sentence. Then, will add the LSTM layer and later add a few Dropout layers to prevent the datasets from overfitting.

When characterizing the Dropout layers, we determine 0.1, implying that 10% of the layers will be dropped. From there on, we add the Dense layer that determines the result of 1 unit. After this, I accumulate the model utilizing the well-known Adam streamlining agent and set the misfortune as the mean_squarred_error. This will process the mean of the squared blunders. Then, fit the model to run on 100 ages with a group size of 32. It is contingent upon the specs of the PC; this may require a couple of moments to complete the process of running.

```python
#Building the LSTM Model
lstm = Sequential()
lstm.add(LSTM(32, input_shape=(1, trainX.shape[1]), activation='relu', return_sequences=False))
lstm.add(Dense(1))
lstm.compile(loss='mean_absolute_error', optimizer='adam')
```
Python

Figure 13: building LSTM model

System Architecture

Preprocessing of data

Figure 14: Preprocessing of data

Overall Architecture

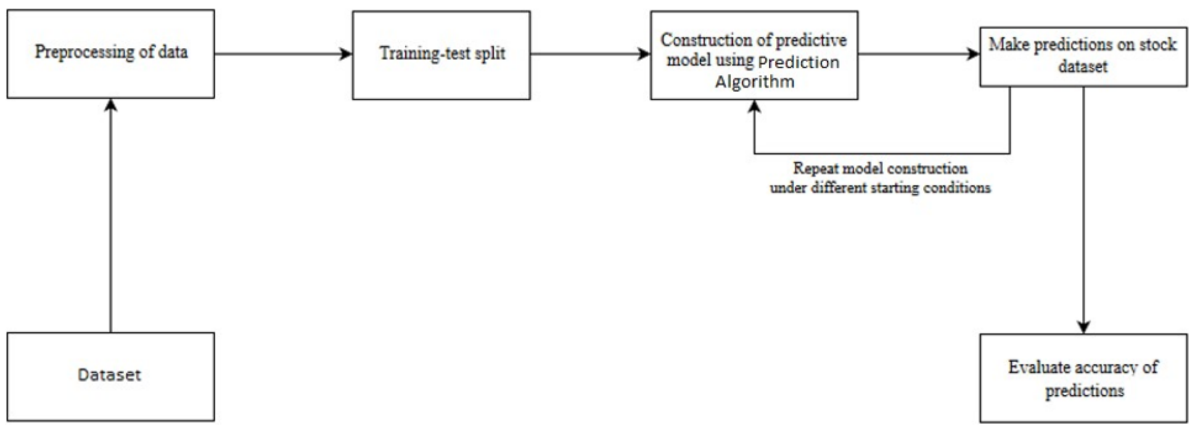


Figure 15: Overall Architecture

RESULTS

Dataset Analysis

```
df.info()
✓ 0.1s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1009 entries, 0 to 1008
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       1009 non-null   object
 1   Open       1009 non-null   float64
 2   High       1009 non-null   float64
 3   Low        1009 non-null   float64
 4   Close      1009 non-null   float64
 5   Adj Close  1009 non-null   float64
 6   Volume     1009 non-null   int64
dtypes: float64(5), int64(1), object(1)
memory usage: 55.3+ KB
```

Figure 16: Stock Dataset Information

Firstly, I have performed Data analysis for the stock price of Netflix company. Fig 13.1 represent the date, open, close, high, low, adjusted close and volume of stocks details.

Read Dataset

```
df=pd.read_csv(r"C:\Users\Bogere\OneDrive\Desktop\Research Project\NFLX.csv")
df.head()
```

[2] ✓ 0.1s Python

	Date	Open	High	Low	Close	Adj Close	Volume
0	2018-02-05	262.000000	267.899994	250.029999	254.259995	254.259995	11896100
1	2018-02-06	247.699997	266.700012	245.000000	265.720001	265.720001	12595800
2	2018-02-07	266.579987	272.450012	264.329987	264.559998	264.559998	8981500
3	2018-02-08	267.079987	267.619995	250.000000	250.100006	250.100006	9306700
4	2018-02-09	253.850006	255.800003	236.110001	249.470001	249.470001	16906900

Figure 17: Read Dataset

After performing data analysis, I read the dataset and this shows the dataset information table starting from the head.

Graph of Close Price history

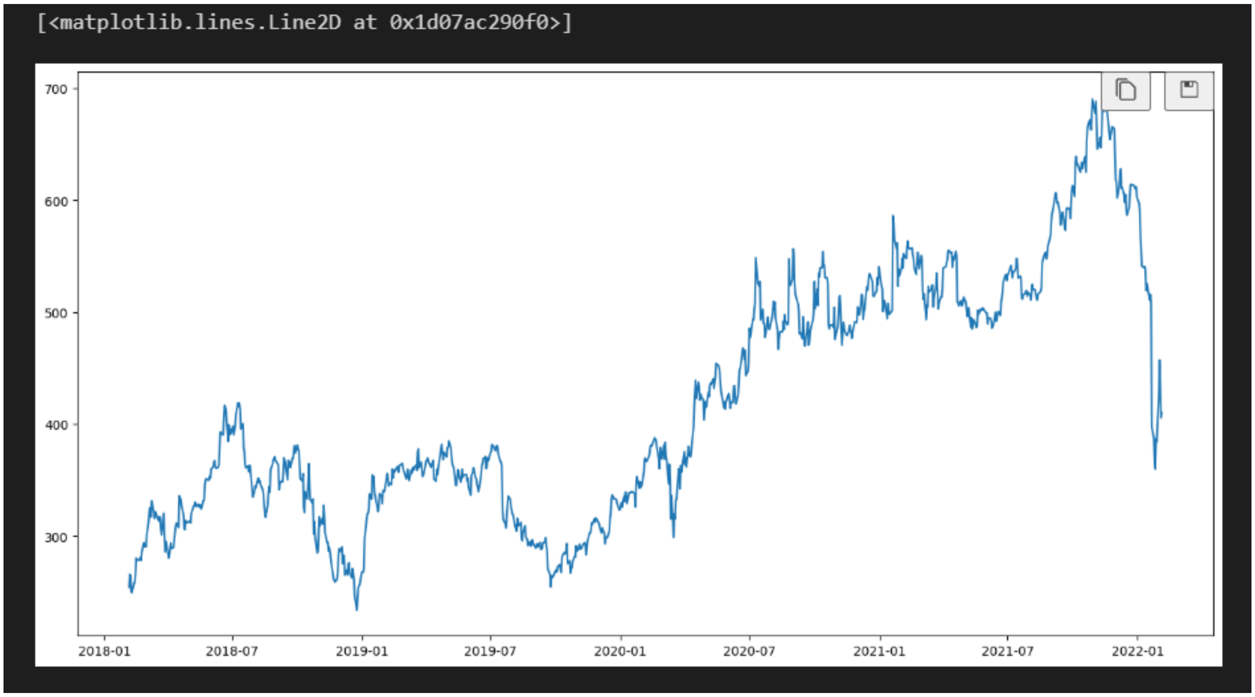


Figure 18: Graph of Close Price history

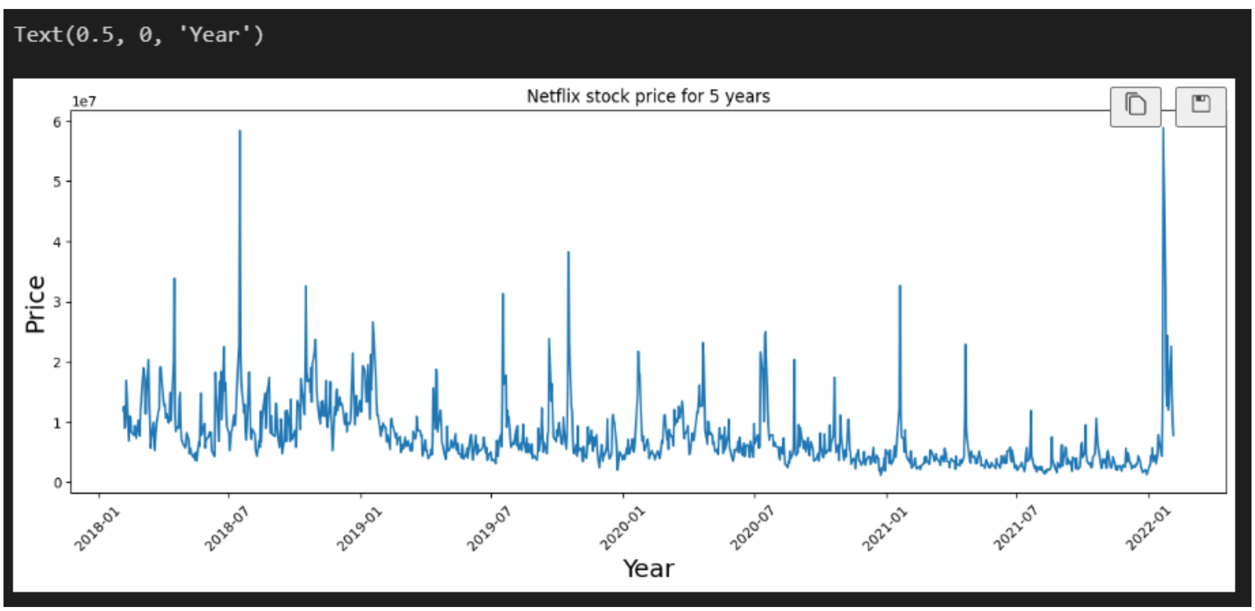
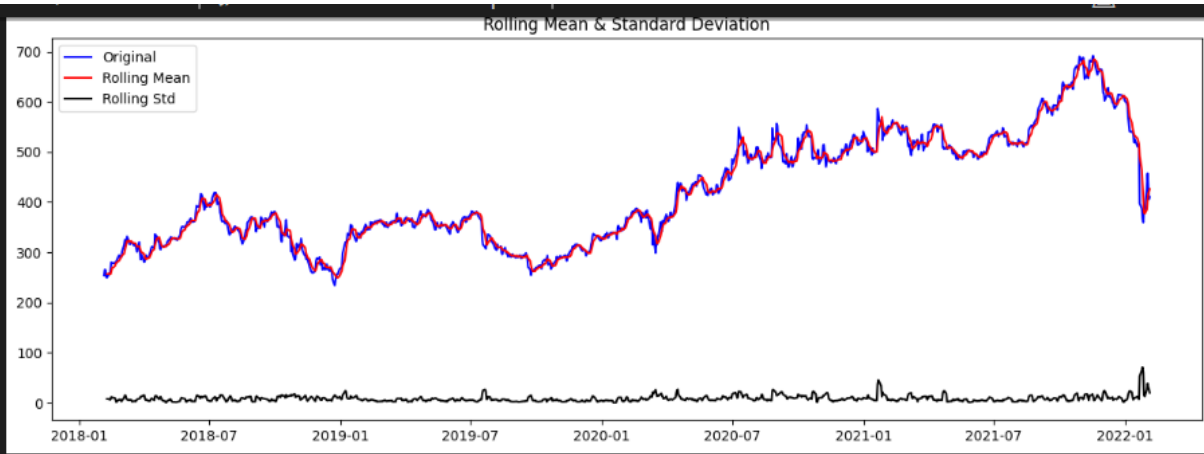


Figure 19: Netflix stock price for 5 years

Stationary test for closing price




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.812591
p-value                          0.374229
#Lags Used                       6.000000
Number of Observations Used   1002.000000
Critical Value (1%)             -3.436893
Critical Value (5%)             -2.864429
Critical Value (10%)            -2.568308
dtype: float64
```

Figure 20: Stationary test for the closing price

Model fitting of Long Short Term Memory architecture, Convolution Neural Network architecture & Hybrid Approach of LSTM+CNN and ARIMA

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 60, 50)            10400

 lstm_1 (LSTM)               (None, 50)                20200

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 30,651
Trainable params: 30,651
Non-trainable params: 0
_________________________________________________________________
```

Figure 21: LSTM summary

SARIMAX Results

Dep. Variable:	Adj Close	No. Observations:	1009
Model:	ARIMA(6, 1, 6)	Log Likelihood	-3838.362
Date:	Mon, 10 Apr 2023	AIC	7702.723
Time:	23:30:10	BIC	7766.628
Sample:	0	HQIC	7727.002
	- 1009		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2227	0.065	3.449	0.001	0.096	0.349
ar.L2	0.6266	0.062	10.139	0.000	0.505	0.748
ar.L3	-1.1329	0.069	-16.536	0.000	-1.267	-0.999
ar.L4	0.5392	0.062	8.669	0.000	0.417	0.661
ar.L5	0.2455	0.056	4.353	0.000	0.135	0.356

Figure 22: ARIMA summary

```
Output exceeds the size limit. Open the full output data in a text edito
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 time_distributed (TimeDistr  (None, 1, 98, 64)        256
 ibuted)

 time_distributed_1 (TimeDis  (None, 1, 49, 64)        0
 tributed)

 time_distributed_2 (TimeDis  (None, 1, 47, 128)       24704
 tributed)

 time_distributed_3 (TimeDis  (None, 1, 23, 128)       0
 tributed)

 time_distributed_4 (TimeDis  (None, 1, 21, 64)        24640
 tributed)

 time_distributed_5 (TimeDis  (None, 1, 10, 64)        0
 tributed)

 time_distributed_6 (TimeDis  (None, 1, 640)           0
```

Figure 23: CNN-LSTM summary

Predicted graph

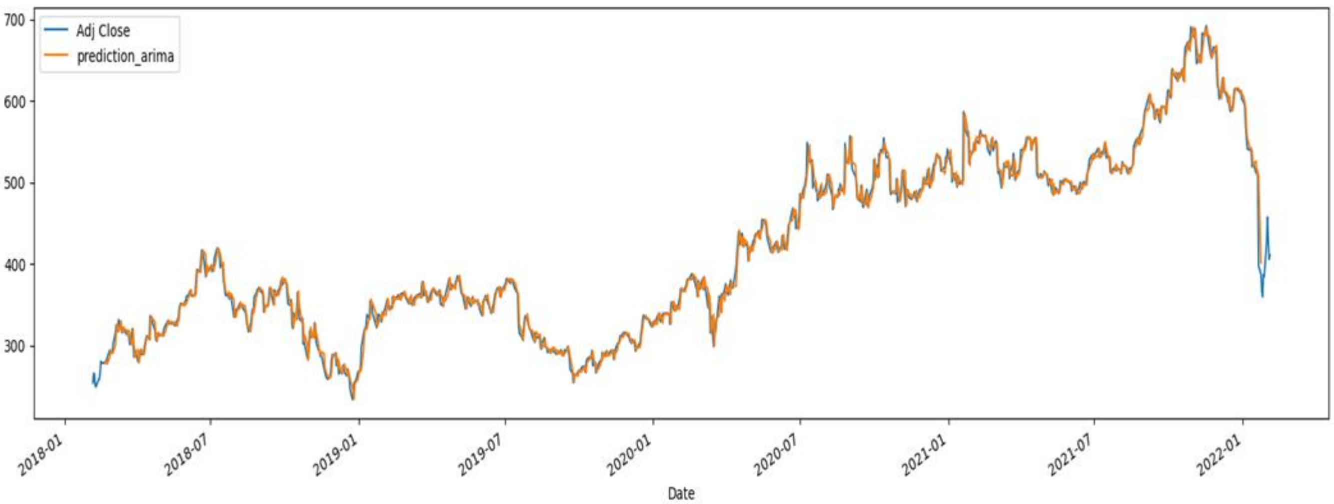


Figure 24: ARIMA prediction result for Adj Close

The graph above shows the ARIMA prediction result against Adj Close and the increase in Adj Close increases the prediction of Arima.

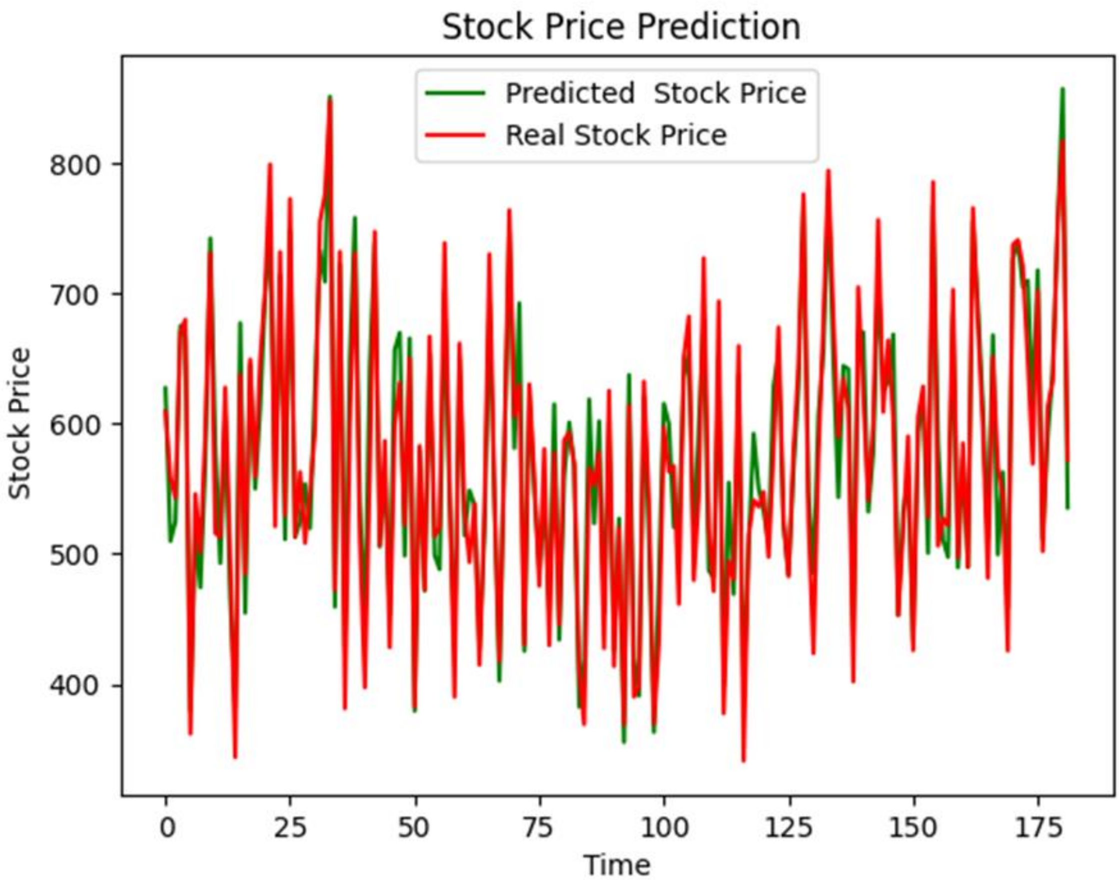


Figure 25: CNN-LSTM Training

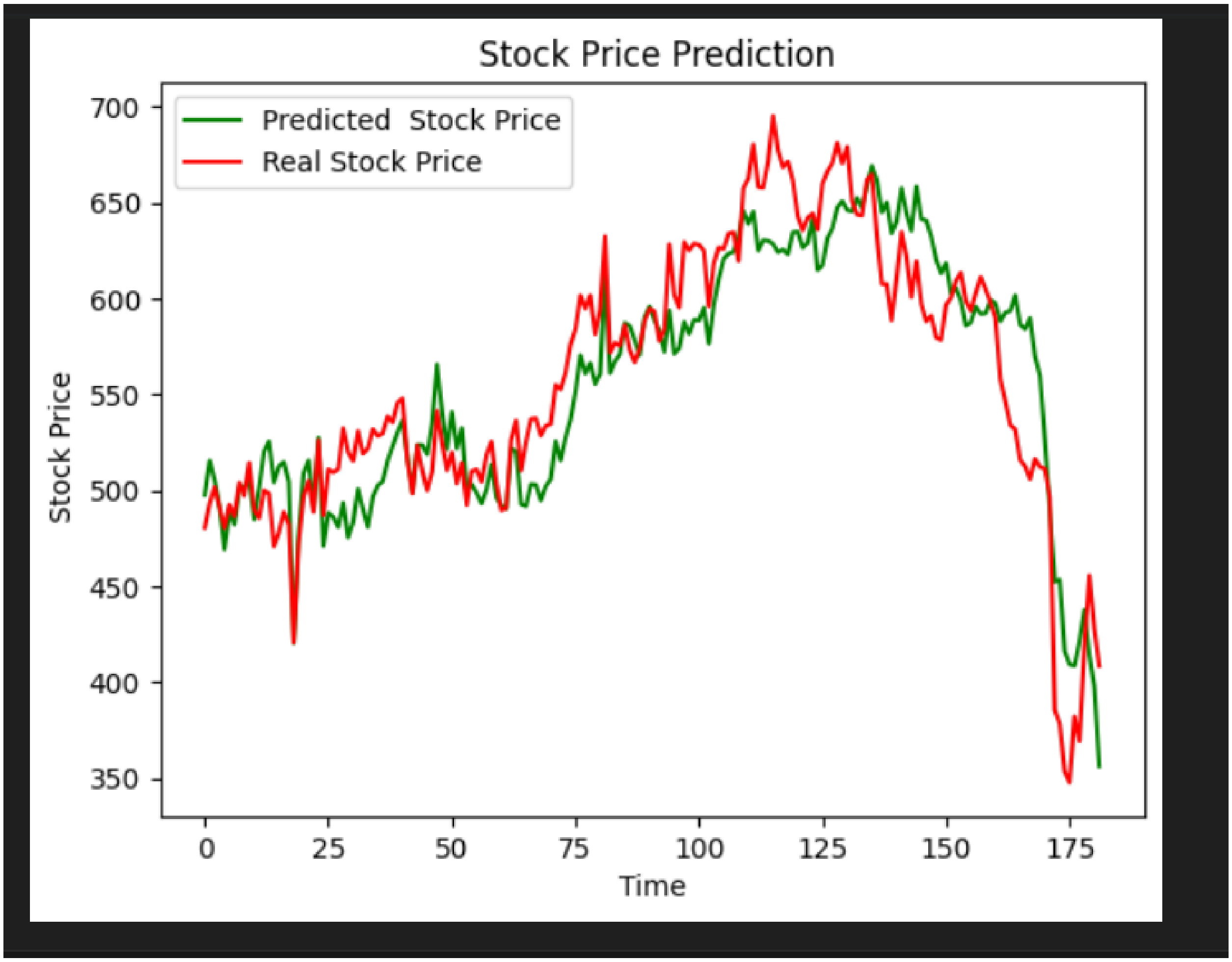


Figure 26: CNN-LSTM Testing

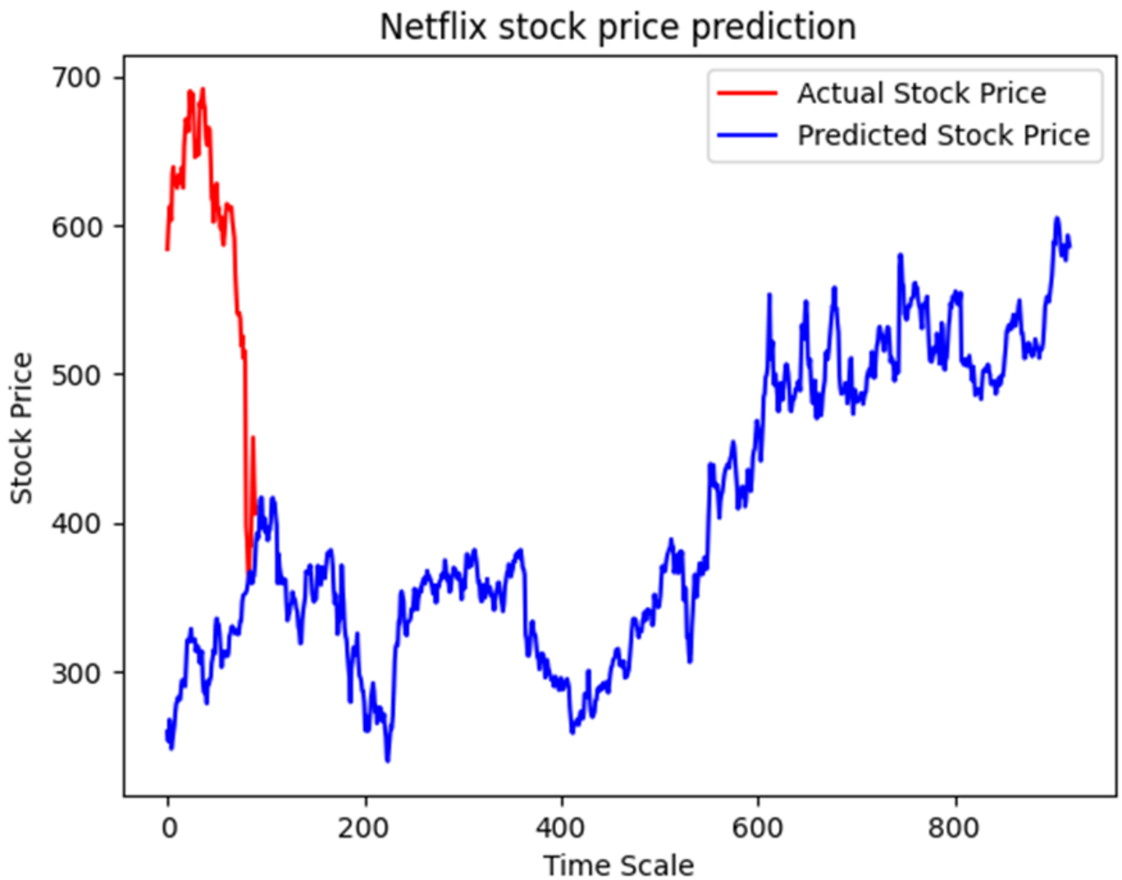


Figure 27: LSTM prediction

The above image shows the Comparison of the LSTM prediction and the actual value of the Adjusted for Close price. From the graph, as the actual stock price of Netflix draws down, due to the time scale, the predicted stock price increases and this implies that whenever the actual stock prices decrease, they will again raise due to the time aspect.

Table 1 Table showing the performance of each algorithm

ALGORITHM	**MSE**	**RMSE**	**MAE**
HYBRID MODEL	315979.5239378678	562.1205599672261	557.535075415432
ARIMA	115.65055028533058	10.754094582312849	7.390266340641206
LSTM	61.55320776897409	7.845585240692634	6.047683842955873

Conclusion

Stock price movements vary in a lot of aspects. That includes economic, political, academic, and company situations and more. The prediction is just one of the ways that can be used to know the stock's price based on current movements. This is to ease the work of investors, individuals, analysts and maybe politicians. Stock market performance can affect how companies perform. If the price of a certain stock goes up, we can conclude that said company is performing at its best at that current time. The model that this project has created is interactive for users to use. It has a feature for users to view the predicted price for the stock chosen. Although the predicted stock in this development project is the one that was personally chosen, the behaviour of the chosen stock can determine how other stocks in the exchange performs. So, technically the model that is developed is predicting all the stocks in the exchange.

CHAPTER FOUR. SYSTEM DESIGN

Structure Chart.

A structure chart (SC) in software engineering and organizational theory shows the breakdown of a system to its lowest manageable levels. They are used in structured programming to arrange program modules into a tree. Each module is represented by a box, which contains the module's name

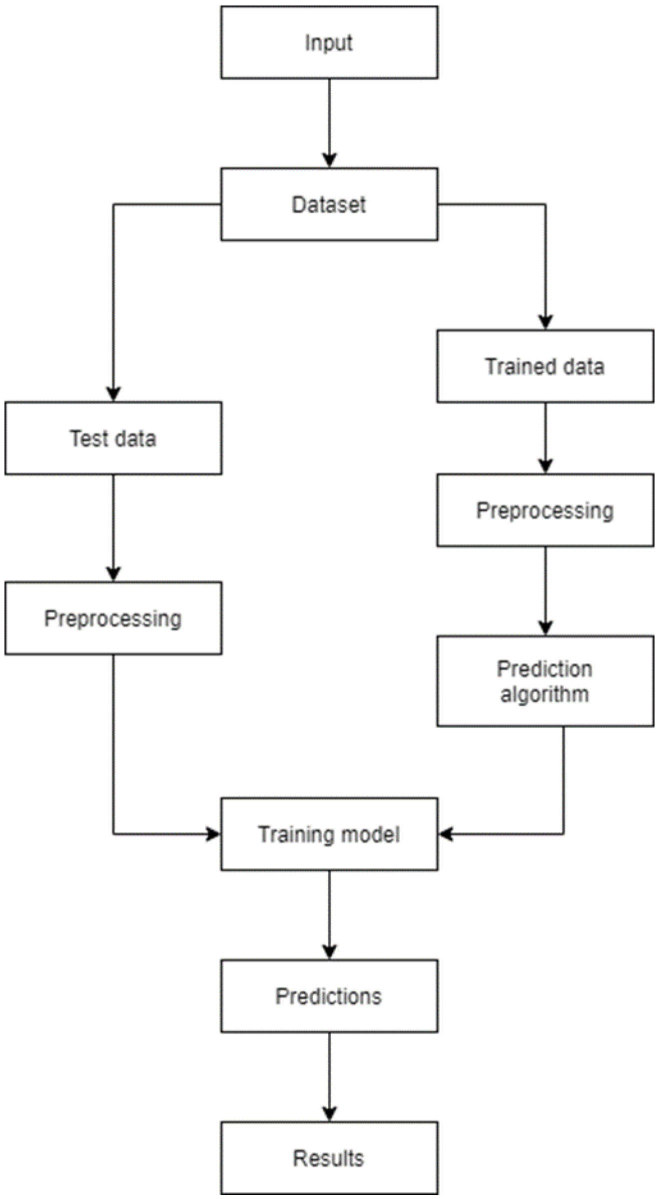


Figure 28: Training and prediction

UML Diagrams.

A UML diagram is a partial graphical representation of a model of a system under design, implementation, or already in existence. UML diagram contains graphical elements (symbols). UML nodes are connected with edges (also known as paths or flows) that represent elements in the UML model of the designed system. The UML model of the system might also contain other documentation such as use cases written as template texts. UML specification does not preclude the mixing of different kinds of diagrams, e.g. combining structural and behavioral elements to show a state machine nested inside a use case. Consequently, the boundaries between the various kinds of diagrams are not strictly enforced. At the same time, some UML Tools restrict the set of available graphical elements that could be used when working on a specific type of diagram. UML specification defines two major kinds of UML diagrams: structure diagrams and behaviour diagrams. Structure diagrams show the static structure of the system and its parts on different abstraction and implementation levels and how they are related to each other. The elements in a structure diagram represent the meaningful concepts of a system and may include abstract, real-world and implementation concepts. Behaviour diagrams show the dynamic behaviour of the objects in a system, which can be described as a series of changes to the system over time.

Use Case Diagram.

In the Unified Modelling Language (UML), a use case diagram can summarize the details of your system's users (also known as actors) and their interactions with the system. To build one, you'll use a set of specialized symbols and connectors. An effective use case diagram can help a team discuss and represent:

I. Scenarios in which your system or application interacts with people, organizations, or external systems.
II. Goals that your system or application helps those entities (known as actors) achieve.
III. The scope of your system.

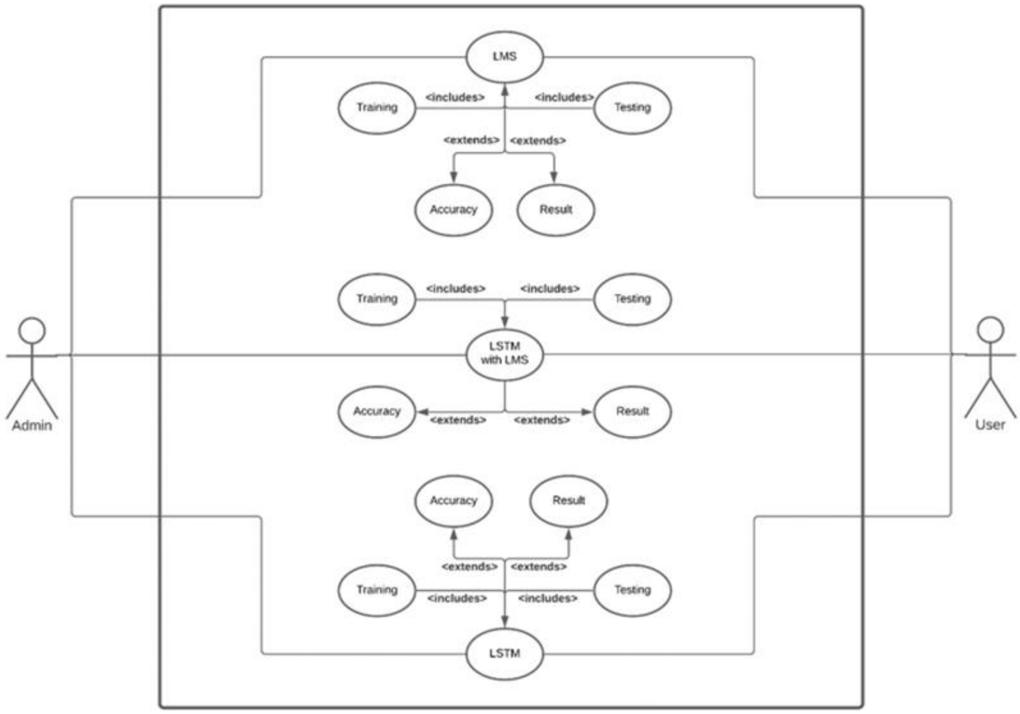


Figure 29: Using LMS, LSTM and LSTM with LMS in the system

Sequence Diagram.

A sequence diagram is a type of interaction diagram because it describes how and in what order a group of objects works together. These diagrams are used by software developers and business professionals to understand requirements for a new system or to document an existing process. Sequence diagrams are sometimes known as event diagrams or event scenarios. Sequence diagrams can be useful references for businesses and other organizations. Try drawing a sequence diagram to:

I. Represent the details of a UML use case.

II. Model the logic of a sophisticated procedure, function, or operation.

III. See how objects and components interact with each other to complete a process.

IV. Plan and understand the detailed functionality of an existing or future scenario.

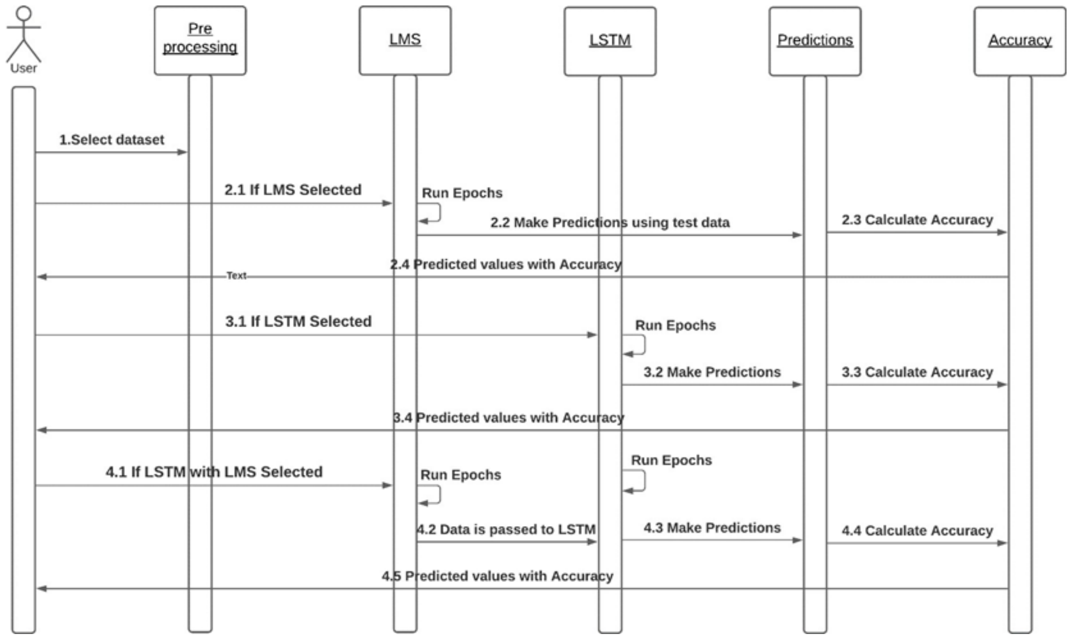


Figure 30: Execution based on model selection

Activity Diagram.

An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system.

An activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed.

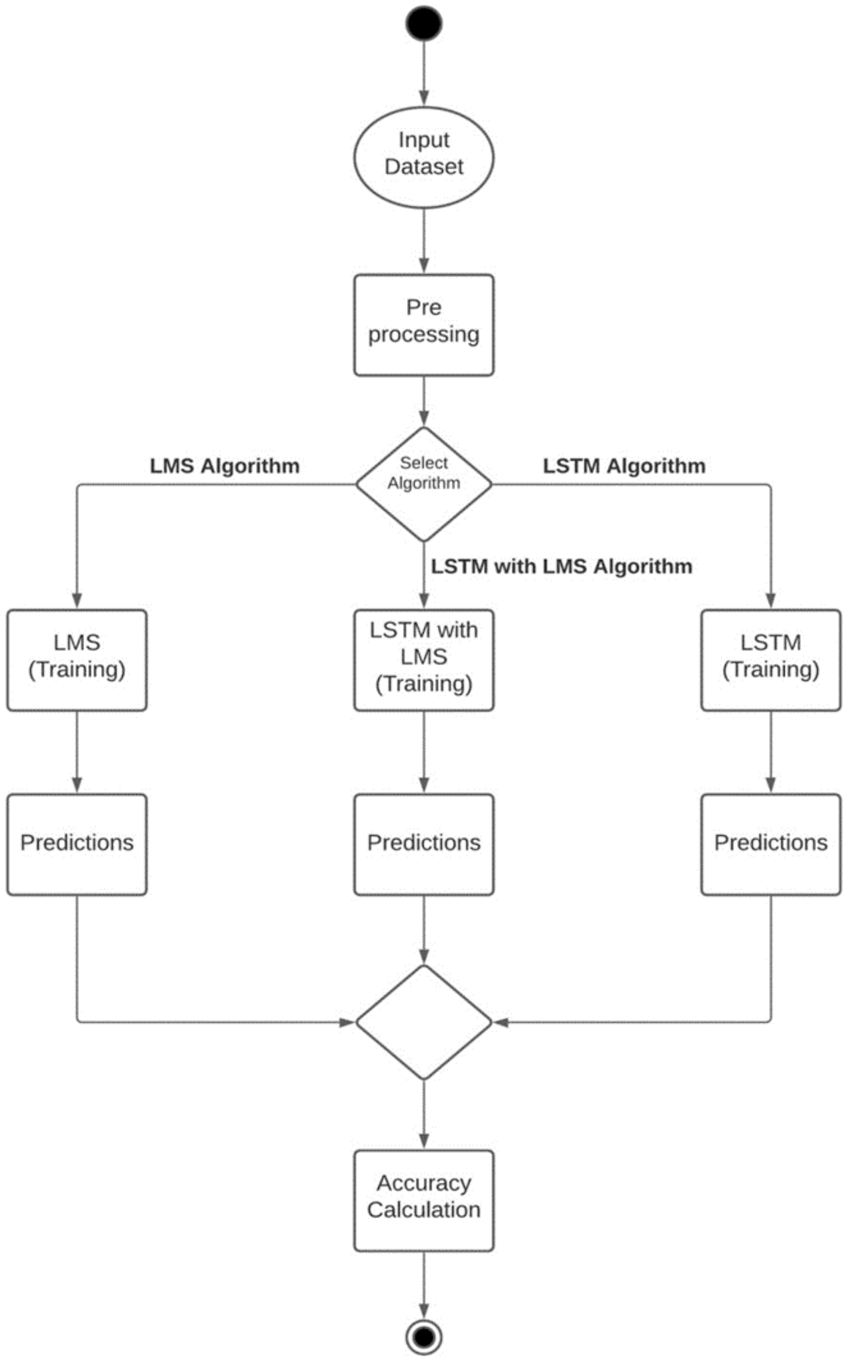


Figure 31: Execution based on algorithm selection

Collaboration Diagram.

Collaboration diagrams are used to show how objects interact to perform the behaviour of a particular use case or a part of a use case. Along with sequence diagrams, collaborations are used by designers to define and clarify the roles of the objects that perform a particular flow of events in a use case. They are the primary source of information used to determine class responsibilities and interfaces. The collaborations are used when it is essential to depict the relationship between the object. Both the sequence and collaboration diagrams represent the same information, but the way of portraying it quite different. The collaboration diagrams are best suited for analyzing use cases.

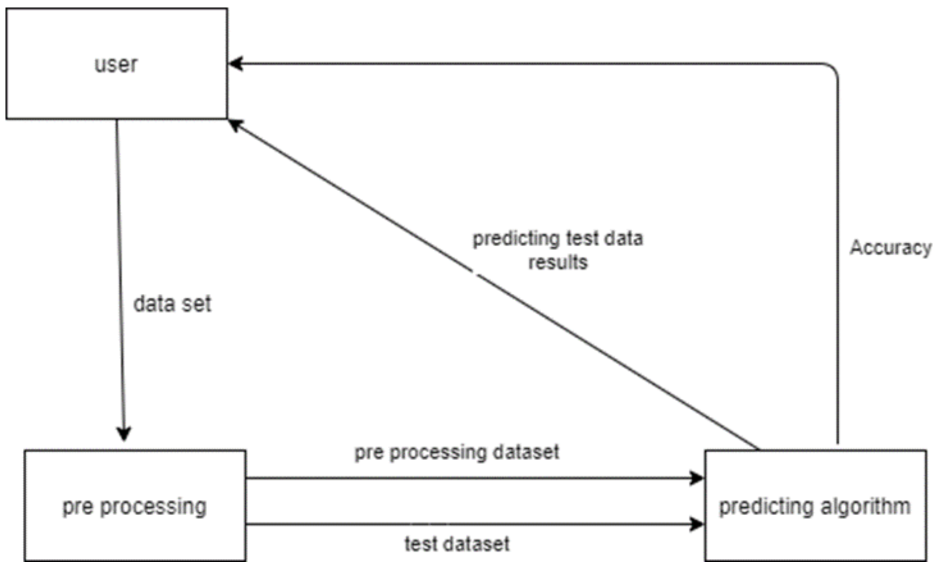


Figure 32: Data transfer between modules

Flow Chart.

A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows.

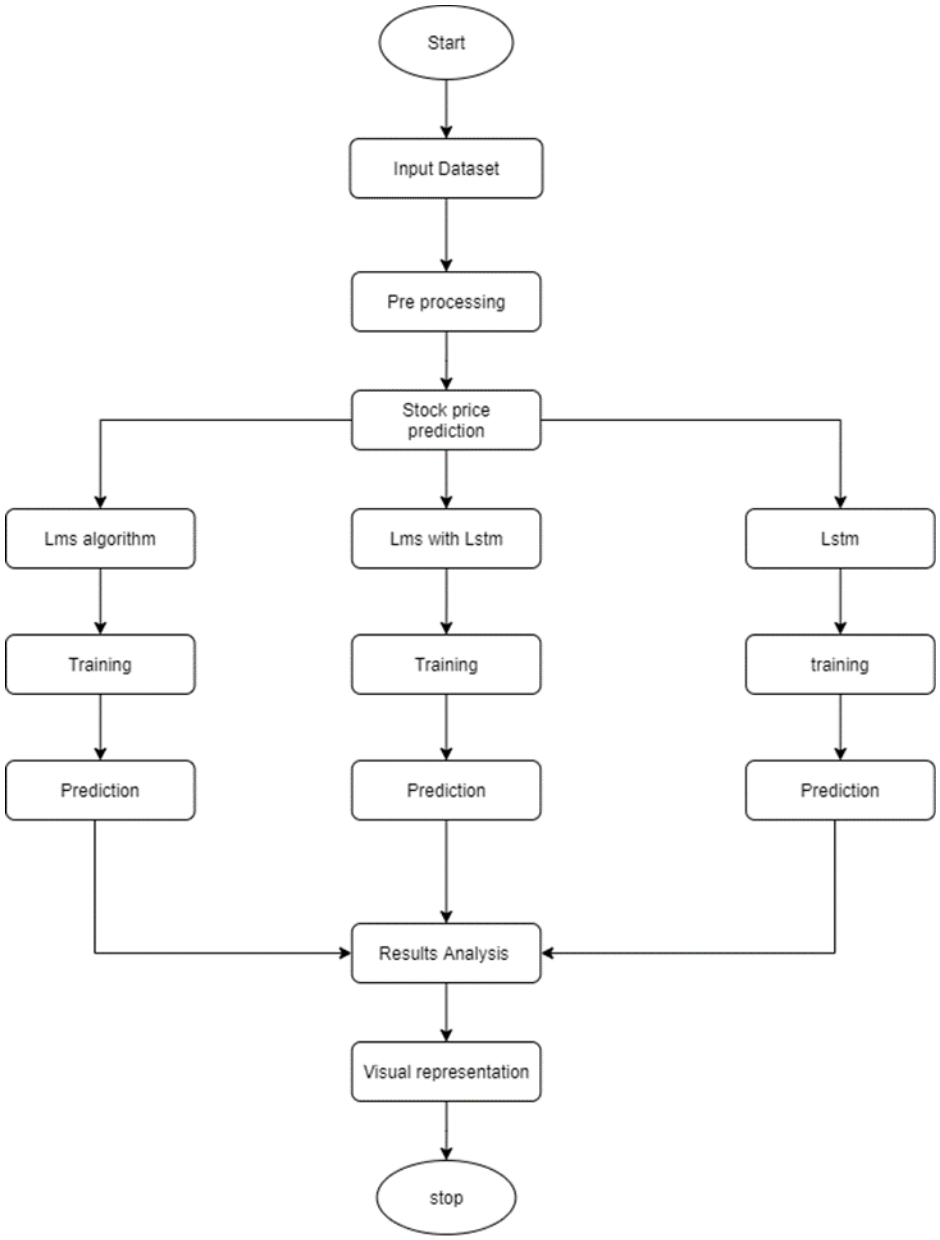


Figure 33: Flow of execution

Component Diagram.

The component diagram is a special kind of diagram in UML. The purpose is also different from all other diagrams discussed so far. It does not describe the functionality of the system but it describes the components used to make those functionalities. Component diagrams are used in modelling the physical aspects of object-oriented systems that are used for visualizing, specifying, and documenting component-based systems and also for constructing executable systems through forward and reverse engineering. Component diagrams are essentially class diagrams that focus on a system's components that often used to model the static implementation view of a system.

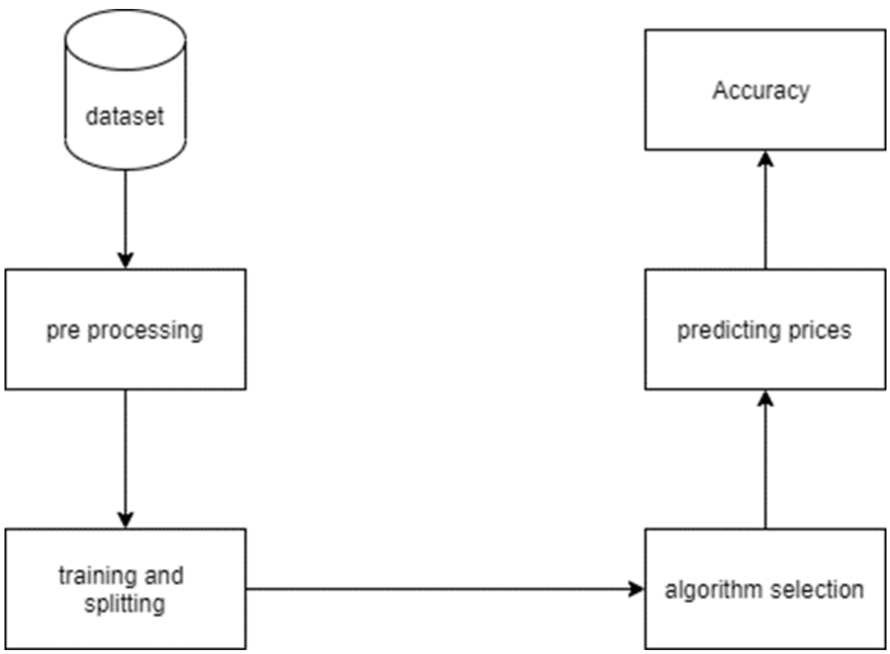


Figure 34: Components present in the system

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

CONCLUSION.

The importance of the stock market to a country's economy will make the types of stock price forecasting methods continue to develop and grow and will continue to be derived from the development of other disciplines. In the development process of the follow-up forecasting method, it is necessary to continuously explore and deeply study the characteristics of the stock market, so as to make the model closer to reality, expand the applicability of the method, and obtain better forecasting accuracy. Since stock data is affected by economic factors, political factors or environmental factors, the law of its change is elusive, and the cycle of the law of change is difficult to determine. Therefore, the model still needs a lot of historical data and the selection of appropriate variables for analysis to obtain the desired results. In the traditional ARIMA model, when analyzing complex stock markets, its prediction results are not particularly ideal, and there are still certain errors in price prediction. As a technology in the field of deep learning, neural networks can solve non-linear problems well. LSTM neural network is optimized on traditional neural network and introduces the concept of "gate", which enhances the long-term memory ability of the model, which enhances its generalization ability. Therefore, the application of the LSTM neural network in analyzing financial-related time series data is promising.

Based on the understanding of traditional time series analysis and RNN and LSTM neural network, this paper constructs a stock price prediction model based on LSTM neural network. For better comparison, we also established a traditional ARIMA model for comparison. As the neural network has a good predictive effect on nonlinear problems, this article chooses the optimized neural network-LSTM model and also chooses the use of single-feature and multi-feature input models to seek better prediction results. The traditional time series model focuses on the role of time in stock forecasting. However, certain errors will occur when the model deals with complex nonlinear stock data, and the model does not consider other factors, such as economics and politics, so the prediction error of the ARIMA model will be large. this thesis considers the ARIMA model, the CNN-LSTM model, and the single-feature input LSTM model that can handle nonlinear data, but they all have unconsidered problems, and their prediction results will also appear to be certain. The multi-feature input LSTM model not only takes into account the influence of external factors, but also can process non-linear data, and its prediction performance is better. Through the result of the prediction, we can see that the prediction result of the mixed model is the best.

The following points may be summarized from this research work;

I. Carry out the steps of smoothing, model order, and model checking on our stock data, and finally establish the ARIMA model and predict the stock price.
II. Construct an LSTM neural network, determine the number of neural networks layers and neurons, and train the parameters.
III. Construct a mixed model to predict stock prices.
IV. Compare the prediction results of the above models.

RECOMMENDATIONS.

Stock markets are the best alternative for businesses to grow and it's a side-way income for the individuals who are ready to invest and earn from the same. The term stock had been in the picture ever since and it's growing in bulk every day. There are thousands of investors investing in the same and making a fortune out of it. Stock market business sectors give an exceptional stage to exchanging and contributing, where exchanges can be executed from any gadget that can associate with the Internet. With the coming of stock market exchanges, individuals have the chance to have numerous roads to make their speculations develop. However, it brought about various sorts of assets like common assets, speculative stock investments, and list assets for individuals. State-run administrations of most nations contribute a piece of their medical care, work, or retirement assets into financial exchanges to accomplish better returns for everybody. Web-based exchange administrations have effectively upset the way individuals purchase and the monetary business sectors have advanced quickly into a solid and interconnected worldwide commercial center. These headways deliver new freedom and the information science methods offer many benefits yet they represent the entire arrangement of new difficulties. In this paper, I propose a scientific categorization of computational ways to deal with securities exchange examination and forecast present a nitty-gritty writing investigation of the cutting-edge calculations and techniques that are regularly applied to stock market predictions and examine a portion of the proceeding with difficulties in the space that require more consideration and give freedom to future turn of events and exploration. In contrast to conventional frameworks, the security exchanges today are assembled utilizing a mix of various advances, for example, ML and enormous information which speak with each other to work with more educated choices. Simultaneously, worldwide client networks on the web have delivered the financial exchange helpless to client opinions because of creating the news and inclined to vindictive assaults. This is the point where further examination can assume a significant part in making ready how financial exchanges will be dissected furthermore. A promising exploration course is to investigate different calculations to assess whether they are adequately amazing to foresee the more extended term since business sectors act like gauging machines over the long haul having not so much commotion but rather more consistency. Half breed draws near that join factual and ML procedures will presumably end up being more helpful for the stock predictions.

REFERENCES

[1] Zhi S U, Man L U, Dexuan L I. Deep Learning in Financial Empirical Applications: Dynamics, Contributions and Prospects[J]. Journal of Financial Research, 2017.

[2] Bin Weng, Ahmed M A, Megahed F M. Stock Market One-day ahead Movement Prediction Using Disparate Data Sources [J]. Expert Systems with Applications, 2017, 79(2): 153–163.

[3] B. Lu, X. Liu, X. Zhu, J. Wang and J. Lu, "Research on Stock Price Forecasting Based on Machine Learning," 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, China, 2018, pp. 1622-1626, doi: 10.1109/ICMA.2018.8466028.

[4] DOSKI, S.A.M., THE ROLE OF RISKTAKING IN MODERATING THE INFLUENCE OF OWNERSHIP STRUCTURE AND SELECTED ECONOMIC FACTORS ON STOCK MARKET IN TURKEY.

[5] Shen, S., Jiang, H. and Zhang, T., 2012. Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, pp.1-5.

[6] Jin, X., Guo, D. and Liu, H., 2014, September. Enhanced stock prediction using social network and statistical model. In 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA) (pp. 1199-1203). IEEE.

[7] Nengovhela, M., 2022. Forecasting stock market returns using machine learning models (Doctoral dissertation, University of Johannesburg (South Africa)).

[8] Rapach, D. and Zhou, G., 2013. Forecasting stock returns. In Handbook of economic forecasting (Vol. 2, pp. 328-383). Elsevier.

[9] Campbell, J.Y. and Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review, 94(5), pp.1249-1275.

[10] Weerakody, P.B., Wong, K.W., Wang, G. and Ela, W., 2021. A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing, 441, pp.161-178.

[11] Kumar, R.D., STOCK MARKET PREDICTION USING MACHINE LEARNING.

[12] Marwala, L.R., 2010. Forecasting the stock market index using artificial intelligence techniques.

[13] Y. Lei, K. Zhou, and Y. Liu, “Multi-Category Events Driven Stock Price Trends Prediction,” in 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nanjing, China, Nov. 2018, pp. 497–501, doi: 10.1109/CCIS.2018.8691392.

[14] Dharmaraja, Selvamuthu., Vineet, Kumar., Abhishek, Mishra. (2019). Indian stock market prediction using artificial neural networks on tick data. Financial Innovation, 5(1), 1-12. Available from: 10.1186/S40854-019-0131-7

[15] B. Jeevan, E. Naresh, B. P. V. kumar, and P. Kambli, “Share Price Prediction using Machine Learning Technique,” in 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739647.

[16] Mariam, Moukalled., Wassim, El-Hajj., Mohamad, Y., Jaber. (2019). Automated Stock Price Prediction Using Machine Learning. 16-24.

[17] Du, J., Liu, Q., Chen, K. and Wang, J., 2019, March. Forecasting stock prices in two ways based on LSTM neural network. In 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 1083-1086). IEEE.

[18] Choi, H.K., 2018. Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. arXiv preprint arXiv:1808.01560.

[19] S. E. Gao, B. S. Lin, and C.-M. Wang, “Share Price Trend Prediction Using CRNN with LSTM Structure,” in 2018 International Symposium on Computer, Consumer and Control (IS3C), Dec. 2018, pp. 10–13, doi: 10.1109/IS3C.2018.00012.

[20] Z. Yeze and W. Yiying, “Stock Price Prediction Based on Information Entropy and Artificial Neural Network,” in 2019 5th International Conference on Information Management (ICIM), Cambridge, United Kingdom, Mar. 2019, pp. 248–251, doi:10.1109/INFOMAN.2019.8714662.

[21] Duong, M.H. and Siliverstovs, B., 2006. The stock market and investment. unpublished Manuscript, Available at http://www. finprop. de/Paper5_The_Stock_Market. pdf (accessed 2 June 2010).

[22] Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin, “An innovative neural network approach for stock market prediction”, 2018.

[23] Sidra, Mehtab., Jaydip, Sen., Abhishek, Dutta. (2020). Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models. arXiv: Statistical Finance, doi: 10.1007/978-981-16-0419-5_8

[24] Sidra, Mehtab., Jaydip, Sen. (2020). A Time Series Analysis-Based Stock Price Prediction Using Machine Learning and Deep Learning Models. International Journal of Business Forecasting and Marketing Intelligence, 6(1):272-. doi: 10.13140/RG.2.2.14022.22085/2

[25] Jaydip, Sen., Sidra, Mehtab. (2021). A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing. Research Papers in Economics, Available from: 10.36227/TECHRXIV. 15023361.V1

[26] Abe, M. and Nakagawa, K., 2020, May. Cross-sectional stock price prediction using deep learning for actual investment management. In Proceedings of the 2020 Asia Service Sciences and Software Engineering Conference (pp. 9-15).

[27] Rebwar, M., Nabi., Soran, Saeed., Habibollah, Harron. (2020). A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-wFE). Science, 5(1), 28-48. Available from: 10.24017/SCIENCE.2020.1.3

[28] Davoodi Kasbi, A., Dadashi, I. and Azinfar, K., 2021. Stock price analysis using machine learning method (Non-sensory-parametric backup regression algorithm in linear and nonlinear mode). Advances in Mathematical Finance and Applications, 6(2), pp.285-301.

[29] Hamed, Hamedinia., Reza, Raei., Saeed, Bajalan., Saeed, Rouhani. (2022). Analysis of Stock Market Manipulation using Generative Adversarial Nets and Denoising Auto-Encode Models. 7(1), 1-22. Available from: 10.22034/AMFA.2021.1933112.1608