A PROJECT REPORT

on

PREDICTING STOCK MOVEMENT

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE ENGINEERING

BY

UDIT AGARWAL	2005556
DIVYAM SINGH	20051346
KHUSHI KUMARI	20051912
ARYAN TRIPATHI	20051914

UNDER THE GUIDANCE OF
Dr. SURESH CHANDRA SATAPATHY

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

May 2023

A PROJECT REPORT

on

PREDICTING STOCK MOVEMENT

Submitted to

KIIT Deemed to be University

\

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE ENGINEERING

BY

UDIT AGARWAL	2005556
DIVYAM SINGH	20051346
KHUSHI KUMARI	20051912
ARYAN TRIPATHI	20051914

UNDER THE GUIDANCE OF
Dr. SURESH CHANDRA SATAPATHY

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAE, ODISHA -751024 May 2023

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is to certify that the project entitled

PREDICTING STOCK MOVEMENT

submitted by

UDIT AGARWAL	2005556
DIVYAM SINGH	20051346
KHUSHI KUMARI	20051912
ARYAN TRIPATHI	20051914

is a record of bona fide work carried out by them, in the partial fulfillment of the requirement for the Degree of Bachelor of Engineering Computer Science & Engineering award at KIIT Deemed to be University, Bhubaneswar. This work is done during the year 2022-2023, under our guidance.

Date: / /

(Dr. S.C. Satapathy)
Project Guide

Acknowledgments

We are profoundly grateful to **Dr. SURESH CHANDRA SATAPATHY** of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target from its commencement to its completion.

UDIT AGARWAL
DIVYAM SINGH
KHUSHI KUMARI
ARYAN TRIPATHI

ABSTRACT

Stock price analysis has been a critical area of research and is one of the top applications of machine learning. Stock prediction is basically the prediction of stock values based on past outcomes and patterns. There's a load of complicated financial indicators and also the fluctuation of the stock market is very violent. The objective is to predict stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades. However, as technology is getting advanced the opportunity to gain a steady fortune from the stock market is increased and it also helps experts to find out the most infinitive indicators to make a better prediction.

This abstract carries forward the idea of predicting the future importance of stocks based on the current scenario using Machine Learning and Deep Learning. The idea is to improve the efficiency of the current practices using specifications where required. LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down. We are using the random forest regression technique that'll help us provide higher accuracy even for the larger datasets as supposed to be received from the stock market.

Keywords: LSTM, Random Forest, CNN, ML, DL, Trade Open, Trade Close, Trade Low, Trade High

Contents

List of Figures

Chapter 1

Introduction

Stocks are essentially an equity investment that represents a portion of ownership in a corporation or company; as such, they entitle you to a share of the company's earnings and assets. As such, market value prediction is crucial to maximizing the profit of your stock option purchase while minimizing risk. This is significant because you must invest your money in a stock that will increase in value over time rather than decline. This program can assist them in becoming accustomed to the volatile market changes while limiting losses as a startling number of newcomers join the herd every day.

This can assist individuals in increasing their profit and reducing their losses. Although we cannot guarantee that the results will always be accurate due to the volatile nature of the stock market, which may also vary depending on factors like advertisements, public sentiments, market sentiments, etc., we can keep the values as close as possible in accordance with the historical data and patterns. For this, we are using the Random forest algorithm because it predicts output with a high degree of accuracy and operates effectively even for a large dataset.

It is challenging to forecast the movement of specific stocks on the stock market since it is a dynamic and complex system that is influenced by a wide range of factors. However, there has been an increase in interest in creating predictive models that can foretell stock prices as a result of the development of machine learning techniques and the wealth of available data. In this study, we want to investigate how machine learning techniques might be used to forecast stock movement. The research will use previous data to train the algorithm, which will then be used to forecast future stock movement. The study will also assess the prediction accuracy and evaluate how well it performed in comparison to other machine learning techniques.

Chapter 2

Literature Review

For the majority of us, "what other people think" has always been a crucial piece of knowledge when making decisions. Now, among other things, it is possible to learn about the perspectives and experiences of the enormous group of people who are neither our close friends nor well-known professionals in the field of criticism—that is, people we have never heard of. Conversely, an increasing number of people are sharing their ideas online with complete strangers. This field of interest is motivated by the individual users' interest in online reviews of goods and services as well as the potential power that these reviews may have. Additionally, there are numerous obstacles in the way of this process that must be overcome in order to produce the desired results. In this survey, we examined the fundamental technique that typically occurs in this process and the steps that need to be followed to get over the difficulties encountered.

2.1 Stock Market Analysis Using Supervised Machine Learning

The accuracy of supervised machine learning algorithms in forecasting stock prices is examined in this study by Kunal Pahwa and Neha Agarwal. For a period of ten years, from 2009 to 2018, the authors used a dataset of stock prices and financial ratios of 35 Indian companies listed on the National Stock Exchange. The technique for the analysis—which involves feature selection, data preprocessing, and the use of three supervised learning algorithms, Random Forest, Gradient Boosting, and Support Vector Machine—is discussed in the study.

The performance of these algorithms has been assessed by the authors using a number of measures, including Mean Absolute Error, Root Mean Squared Error, and R-squared value.
The results of the study indicate that the Gradient Boosting algorithm outperforms the other two algorithms in terms of accuracy in predicting stock prices.

According to the paper's findings, supervised machine learning algorithms can be useful for making stock price predictions, and their precision can be increased still further by utilizing more intricate feature engineering strategies and adding more data sources. In conclusion, the study offers insightful information about the use of machine learning in stock market analysis and its potential to improve investing decision-making.

2.2 Stock Analyzer and Bot Using Machine Learning

Professor Chintan Jethva, Saachi Dudani, Esa Malik, Manish Sonje, and Gaurav Tanna's research suggests a system that combines machine learning algorithms and natural language processing techniques to analyze stock market data and make investment suggestions.

A data collector, a machine learning module, and a chatbot are the three primary parts of the system's architecture, which is described in the article. The data gatherer preprocesses stock market data for analysis after obtaining it from multiple sources, including Google Finance and Yahoo Finance. The machine learning module analyses the data and produces investment suggestions using a variety of techniques, including Random Forest and XGBoost. Natural language processing techniques are used by the chatbot to communicate the recommendations to the user in a conversational manner.

On a dataset of five-year stock market data for 50 companies, the author evaluated the system's performance. The outcomes show that the algorithm is capable of making very accurate investment suggestions and stock price predictions. The approach can help investors make wise selections and increase their investment returns, the paper's conclusion states.

Overall, the study offers insightful information about the use of machine learning and natural language processing in stock market analysis, as well as the possibility for creating intelligent systems that can offer tailored investing advice.

2.3 Stock Price Prediction Using Machine Learning

The ability of machine learning algorithms to forecast the stock values of firms listed on the National Stock Exchange of India is explored in this research study by B N Varaprasad, G. Jeevan, Y. Kalyan Chakravarti, and Ch. Kundan Kanth. For a five-year period from 2013 to 2018, the authors used a dataset of daily stock prices for 15 companies from various industries.

The technique for the analysis—including feature selection, data preprocessing, and the usage of four machine learning algorithms—is discussed in the study. These are Random Forests, Support Vector Machines, Artificial Neural Networks, and K-Nearest Neighbor. The performance of these algorithms has been assessed by the authors using a number of measures, including mean absolute error, mean squared error, and R-squared value.

According to the study's findings, the Random Forest algorithm performs better than the other three algorithms at accurately forecasting stock values. The article indicates that utilizing more sophisticated feature engineering approaches and incorporating more data sources can significantly increase the accuracy of machine learning algorithms' ability to predict stock values. In conclusion, the study offers insightful information about the use of machine learning in stock market analysis and its potential to improve investing decision-making.

2.4 Prediction in Stock Price Using of Python and Machine Learning

The research article "Prediction in Stock Price Using Python and Machine Learning" by Anchal Vij, Komal Saxena, and Ajay Rana focuses on the application of machine learning algorithms in predicting the stock prices of businesses listed on the National Stock Exchange of India. The authors used a dataset of daily stock prices for five significant businesses from various industries across a five-year period from 2014 to 2018.

The analysis approach is covered in the paper, which also covers feature selection, data preprocessing, and the usage of four machine learning algorithms: linear regression, decision trees, random forests, and K-Nearest Neighbor. The performance of these algorithms has been assessed by the authors using a number of measures, including mean absolute error, mean squared error and R-squared value.

According to the study's findings, the Random Forest algorithm performs better than the other three algorithms at accurately forecasting stock values. The study finds that machine learning algorithms can be useful for making stock price predictions and that their precision can be increased any further by adding more data sources and applying more sophisticated feature engineering approaches. In conclusion, the study offers insightful information about the use of machine learning in stock market analysis and its potential to improve investing decision-making.

Chapter 3

Problem Statement / Requirement Specifications

In data analysis, time series forecasting and modeling are crucial. Time series analysis is a specialized area of statistics that is often employed in areas like operation research and econometrics. Time Series is a frequently used tool in data science and analytics. Stock prices are erratic and subject to a variety of influences. This project's primary goal is to forecast stock prices using the Random Forest algorithm and Long Short-Term Memory (LSTM).

3.1 Project Planning

Here are the steps to be followed while planning to execute the project development for predicting stock prices:

1. Defining the scope of the project and identifying the target audience: The first step is to define the scope of the project, its objectives, and the target audience. This will help in identifying the user requirements and features that need to be developed.

2. Collection and preprocessing of data: Gather relevant data for the project, including historical stock prices, financial statements, economic indicators, and news sentiment. Preprocess the data to ensure that it is clean, accurate, and in a format suitable for analysis.

3. Feature engineering: Feature engineering entails choosing the data's most pertinent features and converting them into a format that machine learning algorithms can use. This process is essential to figuring out how accurate the prediction models are.

4. Creating and evaluating models using LSTM and Random Forest: The preprocessed data and chosen features should be used to create these machine learning models. Compare the performance of the models using the proper evaluation metrics.

5. Improve the models' performance by adjusting the hyperparameters of the models. The learning rate, batch size, number of epochs, and other parameters can all be changed in this process.

6. Test and validate the project: Make sure the project is working as planned and that the predictions are correct by thoroughly testing it. To assess the performance of the models, compare the findings to the real stock prices.

3.2 Project Analysis

3.2.1. Introduction

This document's goal is to outline the specifications for a computer program that employs the Random Forest algorithm and long short-term memory (LSTM) to forecast stock values. The computer program will examine previous stock price data, carry out time series forecasting and modeling, and anticipate future stock values based on a number of variables. The software program will be created for use in data science and analytics.

3.2.2. Scope

The computer program will be created to carry out the following functions:
- Examine previous stock data
- Carry out time series modeling and forecasting
- Use the Random Forest method and Long short-term memory (LSTM) to forecast future stock values.

3.3.3. Functional Requirements

The software system will be able to analyze historical stock data, which satisfies the following functional requirements.
- Using the Random Forest method and Long short-term memory (LSTM), the software system will be able to forecast and model time series.
- The analysis and modeling results will allow the software system to forecast future stock prices.
- With a high degree of confidence, the software system will deliver precise stock price predictions.

3.3.4. Non-Functional Requirements

The software system will satisfy the non-functional requirements listed below:
- The software will be simple to use and user-friendly.
Large datasets can be handled well by the software system.
- The software system will have the ability to scale up and

3.3.5. Performance Requirements

The software system will be able to analyze historical stock data in real time, which satisfies the following performance requirements.
- The software system will be able to model and forecast time series data efficiently and accurately.
- With a high degree of accuracy, the software system will be able to forecast future stock values.

3.3.6. User Interface Requirements

The software system will have an intuitive interface that is simple to use and explore.

3.3.7. System Interface Requirements

The software program will have the ability to connect to other programs and databases.

3.3.8. Security Prerequisites

The software system will have suitable safeguards in place to secure data and thwart unauthorized access, and it will be developed with security in mind.

3.3.9. Summary

With the help of the Random Forest algorithm and Long short-term memory (LSTM), the software system will be able to predict stock values. It will be user-friendly and simple to use, and it will be built with scalability, dependability, and security in mind.

3.3 System Design

3.3.1 Schematic diagram

In software engineering and organizational theory, a structure chart (SC) is a diagram that depicts the breakdown of a system to its most basic, manageable components. In structured programming, they are employed to organize program modules into a tree. Each module is symbolized by a box that carries its name.

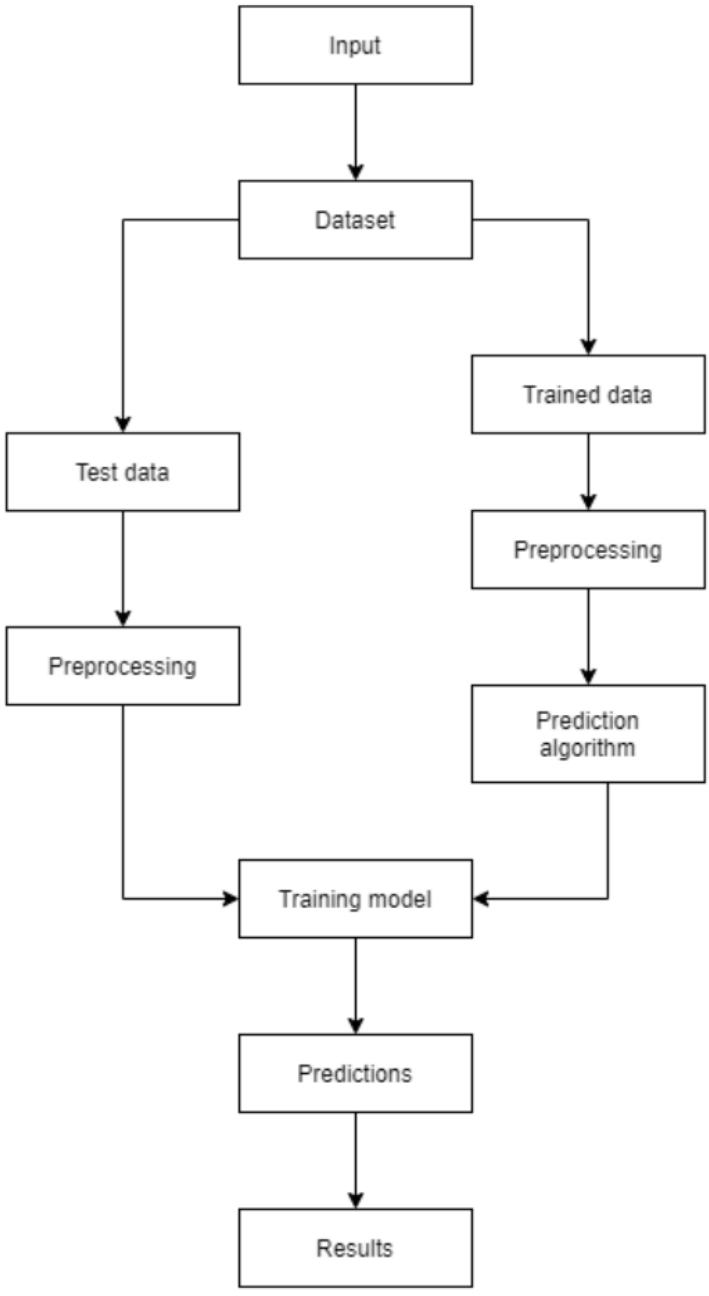

Fig. 3.1: Training and prediction

3.3.2 Design Constraints

Design constraints are restrictions or limitations that must be taken into account during a project's planning and development. The following design restrictions may need to be taken into account when creating a project that uses machine learning algorithms to anticipate stock prices:

1. Data availability and quality: The quality and amount of the data used for training and testing have a significant impact on the prediction models' accuracy. To ensure that the models are trustworthy and accurate, the project must take into account the availability and quality of data.

2. Computational resources: Machine learning algorithms may call for substantial amounts of memory and processing power. The project needs to ensure that it considers the accessibility of these resources and make sure the models can be operated effectively.

3. Time restrictions: Timing is crucial when attempting to predict stock prices. To guarantee that the predictions are pertinent and correct, the models must be created and tested within a certain timeframe.

4. Model complexity: The accuracy and computational needs of machine learning models can be impacted by their complexity. To ensure that the models can be used effectively, the project must strike a balance between the models' complexity, accuracy, and efficiency.

5. Ethical considerations: The project must take ethics into account, including how personal data is handled and whether there may be biases in the data and algorithms that are employed. It should make sure that the forecasts aren't utilized to manipulate the stock market or harm any specific group.

3.3.3 UML Diagrams

A partial graphical representation (view) of a model of a system that is being designed, being implemented, or that is already in use is called a UML diagram. A UML diagram contains graphical elements (symbols) that reflect the components of the intended system's UML model. These graphical elements are UML nodes connected with edges (also known as paths or flows). Additional documentation, such as use cases that are written as templated texts, may also be included in the system's UML model.

The basic graphical symbols used in the diagram determine its type. A class diagram, for instance, is a diagram where the classes are the main symbols in the content area. A use case diagram is a visual representation of actors and uses cases. A sequence diagram shows the sequence of message exchanges between lifelines.

UML specification does not preclude the mixing of different kinds of diagrams, e.g. to combine 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 do restrict the set of available graphical elements which could be used when working on a specific type of diagram. UML specification defines two major kinds of UML diagrams: structure diagrams and behavior 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.

Behavior diagrams show the dynamic behavior of the objects in a system, which can be described as a series of changes to the system over time.

3.3.3.1 Use Case Diagram

A use case diagram in the Unified Modelling Language (UML) can condense the specifics of your system's users (sometimes referred to as actors) and their interactions with the system. You'll need a specific set of connectors and symbols to construct one. Your team can discuss and visualize the following using a good use-case diagram:

• Scenarios in which your system or application interacts with individuals, groups, or external systems.

• Objectives that those entities (sometimes referred to as actors) are assisted in achieving by your system or application.

• How extensive is your system

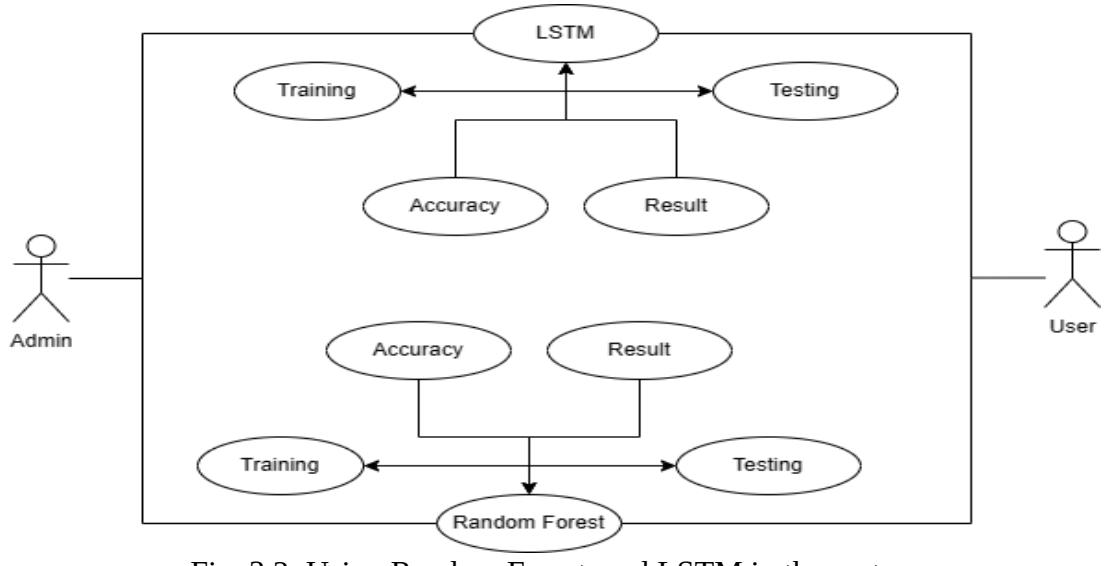

Fig. 3.2: Using Random Forest, and LSTM in the system

3.3.3.2 Sequence Diagram

A sequence diagram is a type of interaction diagram because it illustrates the interactions between a group of objects and the order in which they take place. Software engineers and business experts use these diagrams to comprehend the specifications for a new system or to describe an existing procedure. Event diagrams and event scenarios are other names for sequence diagrams.

For businesses and other organizations, sequence diagrams can be a useful resource. Create a sequence diagram to:
• Describe a UML use case's specifics.
• Recreate the logic of a complex process, feature, or operation.
• Observe how elements and objects work together to complete a process.
• Create a detailed functional plan for a current or potential circumstance.

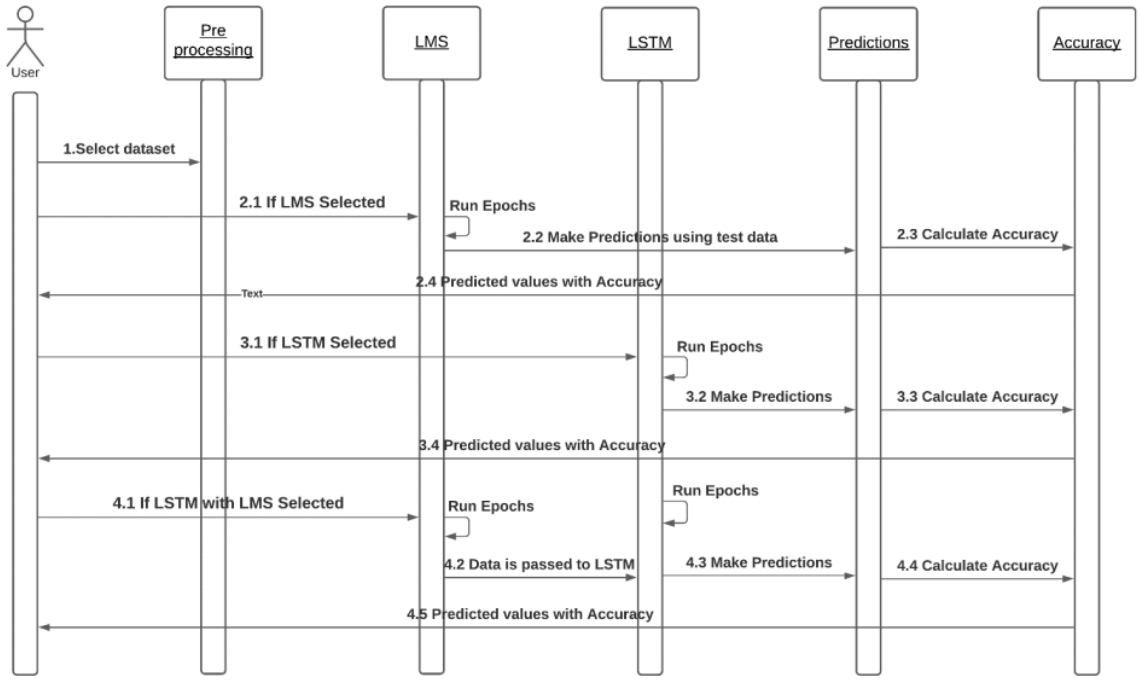

Fig. 3.3: Sequence Diagram for LSTM

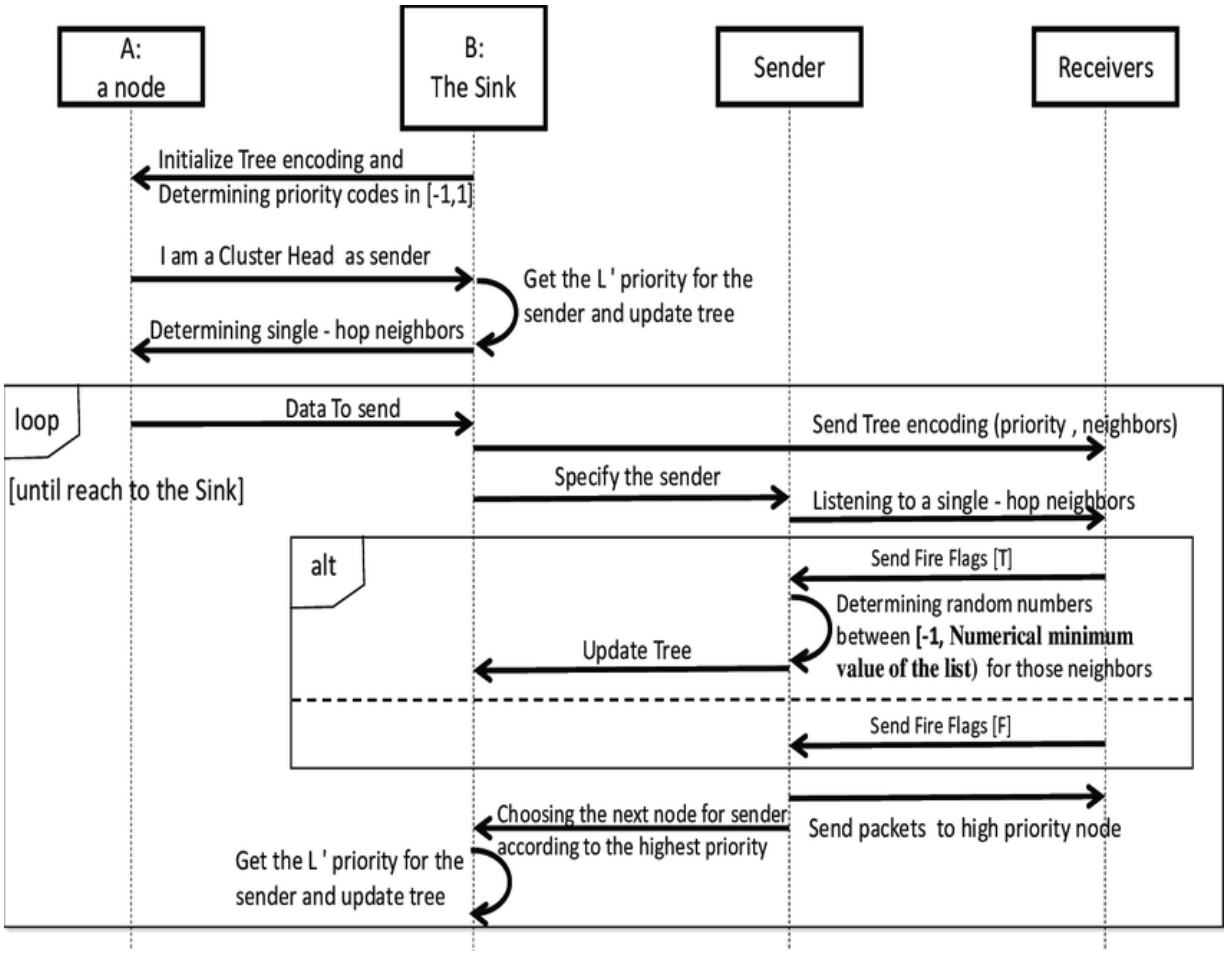

Fig. 3.4: Sequence Diagram for Random Forest

3.3.3.3 Activity Diagram

An activity diagram is a behavioral diagram, meaning it shows how a system behaves. An activity diagram shows the numerous decision routes that are available throughout the execution of an activity by depicting the control flow from a start point to an endpoint.

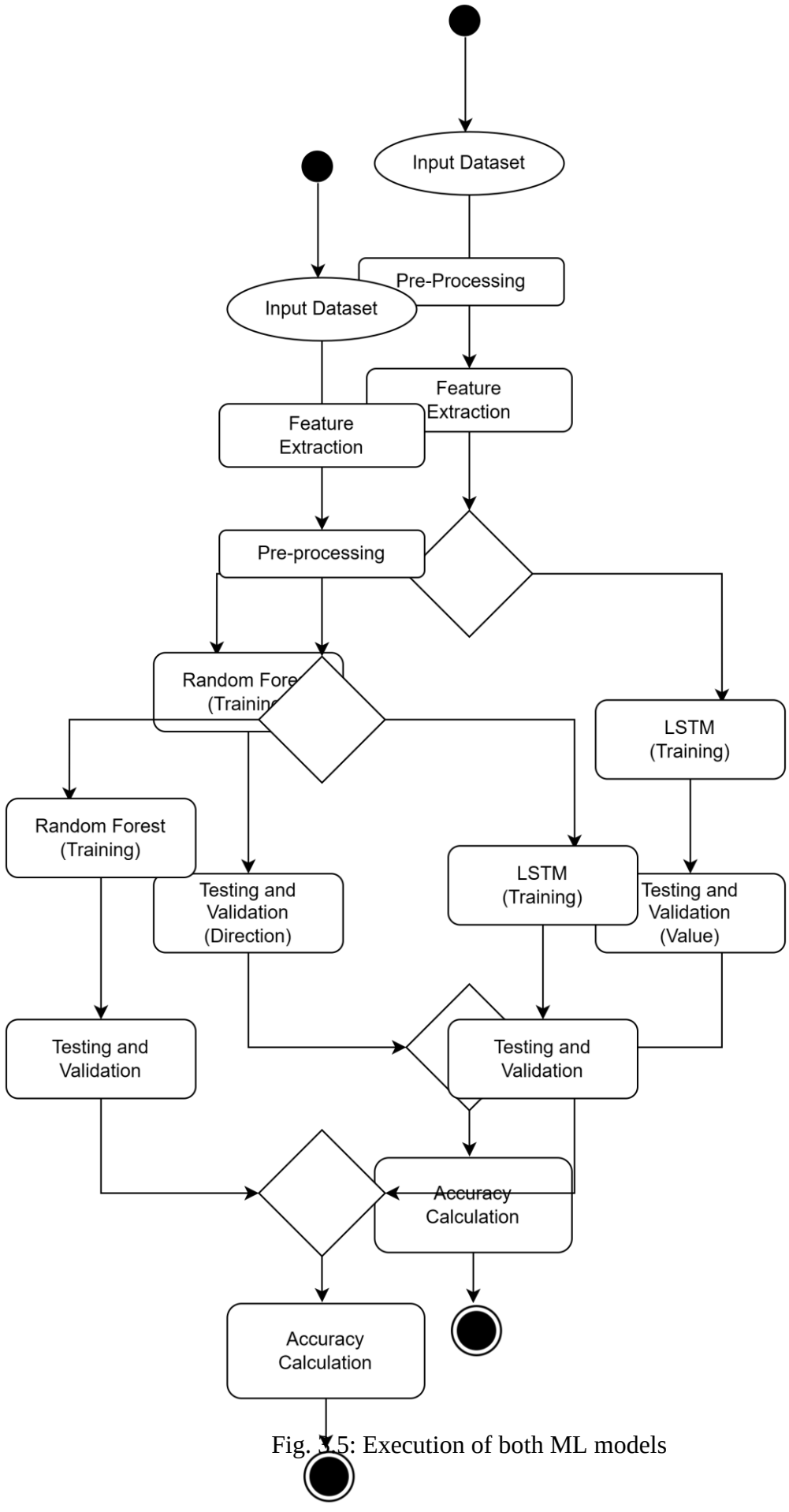

Fig. 3.5: Execution of both ML models

3.3.3.4 Collaboration Diagram

Collaboration diagrams are used to demonstrate how objects work together to carry out a specific use case's behavior or a section of a use case. Designers utilize collaboration in addition to sequence diagrams to specify and make clear the roles of the objects that carry out a specific sequence of actions in a use case. They serve as the main information source for determining the roles and relationships between classes.
When it is crucial to show the interaction between the objects, collaborations are employed. The information is the same in both the sequence and cooperation diagrams, but they depict it very differently. The analysis of use cases is best accomplished with collaboration diagrams.

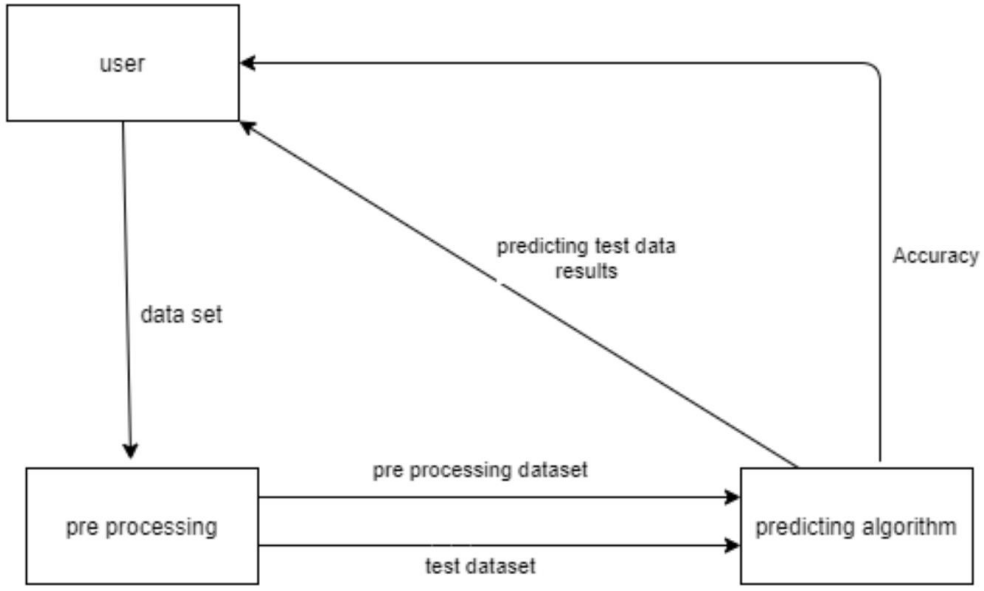

Fig. 3.6: Data transfer among Modules

3.3.3.5 Flow Chart

A diagram that depicts a workflow or process is called a flowchart. Another definition of a flowchart is a diagrammatic description of an algorithm or a step-by-step process for solving a problem. The flowchart displays the processes as various types of boxes and their order by joining the boxes together using arrows.

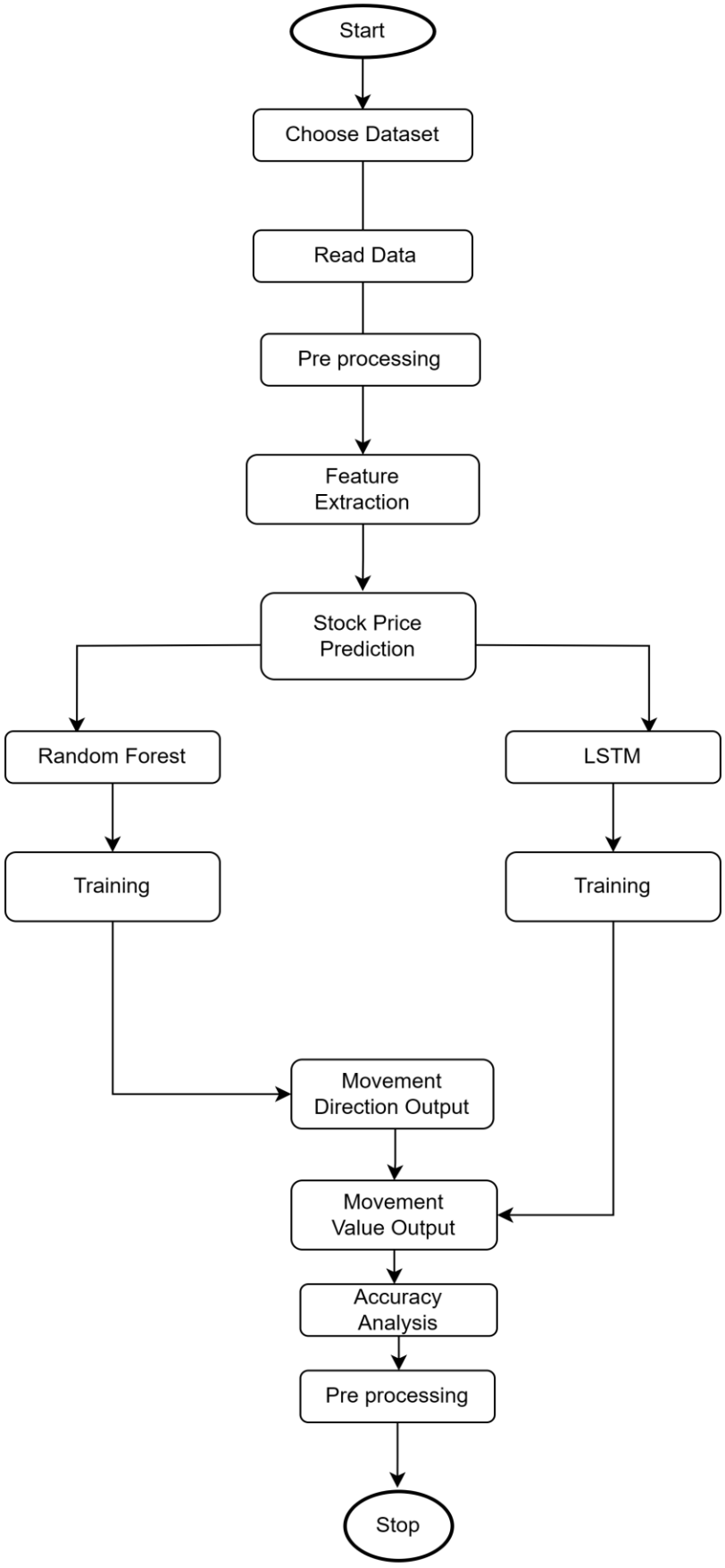

Fig. 3.7: Flow of execution

3.3.3.6 Component Diagram

A unique type of diagram in UML is the component diagram. In addition, the goal is distinct from the previous diagrams mentioned. Although it does not describe the system's functionality, it does describe the parts that go into creating that functionality.

In order to visualize, specify, and document component-based systems as well as to build executable systems through forward and reverse engineering, component diagrams are used to depict the physical features of object-oriented systems. Component diagrams are essentially class diagrams that concentrate on a system's individual components and are frequently used to describe a system's static implementation perspective.

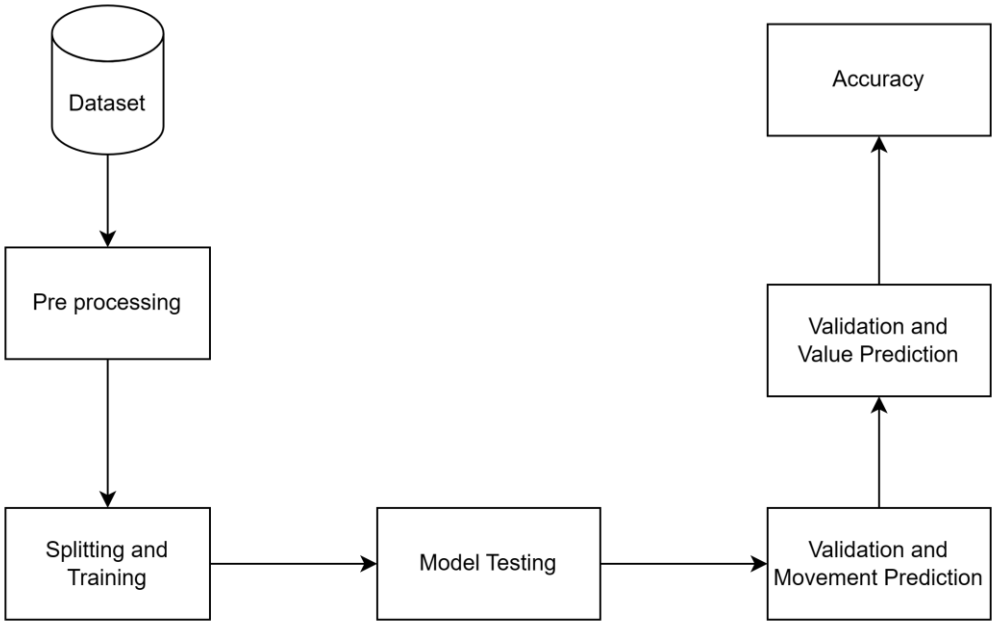

Fig. 3.8: Components present in the system

Chapter 4

Implementation

4.1 Methodology OR Proposal

The main goal of this project is to create a reliable and accurate stock prediction model using LSTM for time-series forecasting and Random Forest for feature selection. Future stock prices can be predicted using the suggested approach with great consistency and accuracy. The two categories of prediction techniques are Direction Prediction and the Value of Movement in that Direction. The Random Forest model is part of the direction prediction system. Single-layer LSTM, support vector machines, recurrent neural networks, and other techniques are used in value prediction. A long short-term memory network (LSTM) was employed.

To create the suggested model, we used the technique described below:

1. Data Gathering: We have gathered historical stock price data from sources that are open to the public, like Yahoo Finance.

2. Data Preprocessing: To get the data ready for analysis using Python libraries like Pandas was used, we performed data preprocessing procedures such as data cleaning, normalization, and feature engineering.

3. Model Training: To train the model on the chosen features, we utilized Random Forest and LSTM. In order to forecast future stock prices, we employed time-series data.

4. Model Evaluation: We used measures like mean absolute error and mean squared error to assess how well the suggested model performed.

5. Project Deployment: We have used Python GUI to make it accessible to the user with a simple and easy-to-use interface where the user just needs to input the name of the stock and will get the predicted output along with values, direction and graph.

4.1.1 Long Short Term Memory Network:

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

Working of LSTM:

Three "gate" structures make up the unique network structure known as LSTM. The input gate, forgetting gate, and output gate are the three gates that make up an LSTM unit. Information can be chosen by rules when it enters the LSTM network. Information that does not comply with the algorithm will be erased by the forgetting gate, leaving only the data that does.

The historical data that were collected from the Internet and used as experimental data in this study. The experiments made use of three data sets. Finding an optimization algorithm with a quicker convergence speed and less resource requirements is necessary.

• We use two LSTM layers in our model and implement drop out in between for regularization. The number of units assigned in the LSTM parameter is fifty. with a dropout of 10 %.
• Mean squared error is the loss function for optimizing the problem with adam optimizer. Mean absolute error is the metric used in our LSTM network as it is associated with time-series data.
• Min-max scaler is used for scaling the data so that we can bring all the price values to a common scale.
• The historical stock data table contains the information of opening price, the highest price, lowest price, closing price, transaction date, volume and so on.
•We have used inverse transformation to get back the original value with the transformed function and used it for visualization

LSTM Architecture

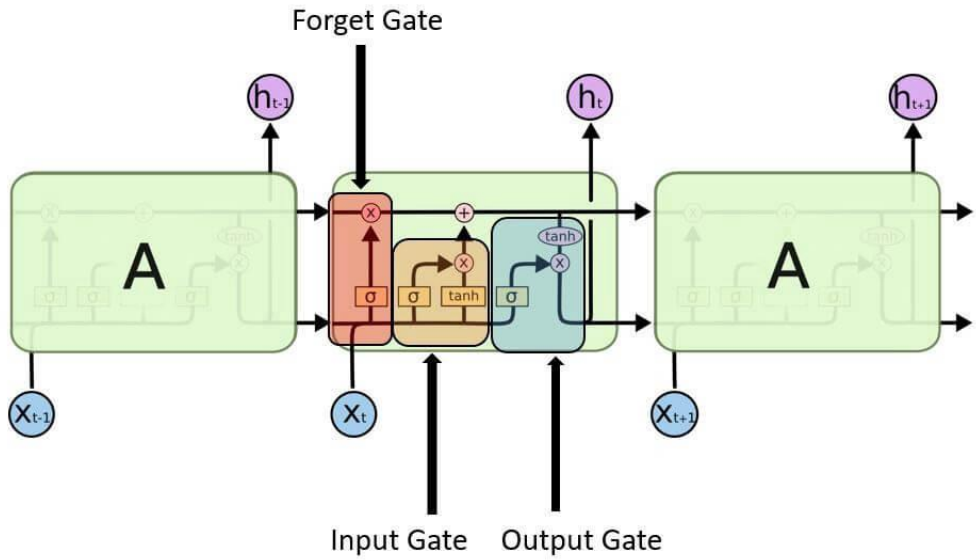

Fig. 4.1. LSTM Structure

Forget Gate:

Information from the cell state is eliminated by a forget gate.

• By multiplying a filter, the information that is no longer necessary for the LSTM to grasp things or that is of less importance is eliminated.

• The LSTM network's performance must be optimized, hence this is necessary.

• H_t-1 and x_t are the two inputs to this gate. The input at that particular time step is x_t, while the hidden state from the preceding cell, or its output, is h_t-1.

Input Gate:

1. Using a sigmoid function to control what values must be added to the cell state. This functions as a filter for all the data from hi-1 and x_t and is fundamentally very similar to the forget gate.

2. Making a vector that contains each value that could be added to the cell state (as determined by h_t-1 and x_t). The tanh function, which produces values between -1 and +1, is used for this.

3. Adding this valuable data to the cell state via an addition operation after multiplying the value of the regulatory filter (the sigmoid gate) by the vector that was formed (the tanh function).

Output Gate:

Again, the operation of an output gate can be divided into three steps:
• Building a vector from the cell state after scaling the values to lie between -1 and +1 using the tan h function.
• Creating a filter that can control the values that must be output from the vector created above by using the values of h_t-1 and x_t. The sigmoid function is used in this filter once more.
• Adding this regulatory filter's value to the vector produced in Step 1 and sending it there as an output as well as to the next cell's concealed state.

4.1.2. Random Forest

A well-liked machine-learning technique called Random Forest can be applied to both classification and regression applications. It is an ensemble learning technique that makes predictions by using several decision trees.

Building a forest of decision trees, each trained on a distinct part of the training data and a random collection of features, is how the algorithm accomplishes its goal. This randomness enhances the model's generalization capabilities and helps to lessen overfitting, a major issue with decision trees.

Here are the main steps involved in the Random Forest algorithm:

1. Data Preparation: The first step is to prepare the data by dividing it into training and testing datasets.

2. Tree Building: The next step is to build decision trees on different subsets of the training data. Each tree is built using a different random subset of features, which helps to reduce the correlation between the trees and improve the model's accuracy.

3. Prediction: Once the trees are built, they are used to make predictions on the testing data. For classification tasks, the final prediction is the mode of the predictions made by individual trees, while for regression tasks, it is the mean of the predictions.

4. Model Evaluation: Last but not least, the Random Forest model's performance is assessed using a variety of metrics, including accuracy, precision, recall, F1-score, etc. The performance of the model can be enhanced by adjusting its hyperparameters.

In conclusion, Random Forest is an effective algorithm that generates forecasts using a group of decision trees. Multiple trees predictions can be combined to improve the model's accuracy and robustness while reducing overfitting.

Decision Trees

Last but not least, the Random Forest model's performance is assessed using a variety of metrics, including accuracy, precision, recall, F1-score, etc. The performance of the model can be enhanced by adjusting its hyperparameters.

In conclusion, Random Forest is an effective algorithm that generates forecasts using a group of decision trees. Multiple trees' predictions can be combined to improve the model's accuracy and robustness while reducing overfitting.

Root Node: This node represents the entirety of the population or sample, which is then partitioned into two or more homogeneous sets. where we started.
Splitting: It is the process of breaking up a node into two or more sub-nodes. For instance, we might separate based on gender.
Decision Node: A decision node is a sub-node that has undergone further subdivision.
Leaf/Terminal Node: Leaf or Terminal Nodes are nodes that do not split.
Pruning: The removal of a decision node's sub-nodes is referred to as pruning. The process of separating in the opposite direction.
Branch/Sub-Tree: An individual branch or sub-tree is referred to as a branch or sub-tree.
Parent and Child Node: A node, which is divided into sub-nodes is called the parent node of sub-nodes whereas sub-nodes are the child of the parent node.

Procedure we followed:

1. We have created a Python function that utilizes historical stock data to train a Random Forest classifier model that can predict the future price trend of a stock.
2. The first step in this function is data cleaning and creation of a target variable based on price change. This ensures that the model is trained to predict the price direction correctly.
3. Once the data is cleaned and prepared, it is split into training and testing sets for the model. The Random Forest algorithm is then used to train the model on the training set and evaluate its accuracy on the testing set.
4. To verify the model's effectiveness, the function makes predictions on the last 5 entries in the data using the trained Random Forest model.
5. The Random Forest algorithm is an effective model that combines multiple decision trees to predict future trends based on a subset of features. This makes it a powerful tool for predicting stock prices with reasonable accuracy.

Ensemble Learning

An ensemble learning model is one in which choices from MANY MODELS are used to enhance the model's overall performance. Ensemble learning is best characterized by the proverb "Two minds are better than one." To better understand the true answer, we use the output from many models. Bagging is used by our Random Forest method to enhance performance.

Advantages of Random Forest

An ensemble learning model is one where options from MANY models are used to improve the performance of the model as a whole. The adage "two minds are better than one" perfectly describes group learning. We use the results from numerous models in order to more fully comprehend the real solution. Our Random Forest approach makes advantage of bagging to improve performance.

The overall layout of the decision tree can be drastically altered by even little modifications to the input data. They are frequently only slightly accurate. With comparable data, several other predictors do better. Decision tree information is biased in favor of attributes with higher levels for data containing categorical variables with various number levels. Calculations can get exceedingly complicated, especially if several numbers are ambiguous or if numerous results are interconnected.

Supervised Learning

There are two types of learning in machine learning. Learning that is guided and unguided. Unsupervised learning involves not supervising the model but rather letting it find out information on its own. We achieve this by giving the model a "UNLABELED" data set that does not specify whether category or value is the "correct" response.

When using supervised learning, we give the model a set of "LABELLED" data that informs it of the "correct" value. Because we give the model a set of labelled data, the supervised learning algorithm Random Forest is an example.

4.1.3. Python GUI

Python GUI stands for Python Graphical User Interface. It is a way to create visually appealing and user-friendly interfaces for Python applications. Python offers various libraries and frameworks to create GUIs such as Tkinter. These libraries allow us to create buttons, menus, text fields, and other interactive components that make it easier for users to interact with Python applications.

Python GUIs are used in a variety of applications, from desktop applications to scientific tools and games. GUIs can improve the user experience by providing a visual interface to the program and making it easier to interact with. Additionally, Python's simplicity and ease of use make it a popular choice for developing GUI applications.

Here we have given an example of our user interface created using The Tkinter Library.

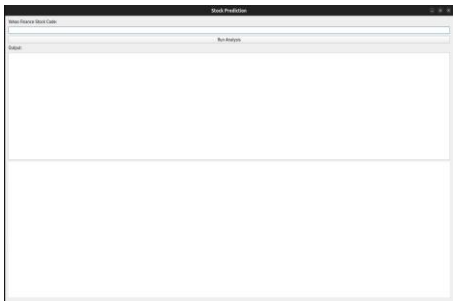

Fig. 4.2 – UI Design

4.2 Testing OR Verification Plan

A key component in the prediction of the movement in stock market prices was the verification plan. The model underwent a stringent testing procedure to make sure they predicted the values correctly. The strategy included functional and performance testing to ensure the model accurately predicted the movement and to ensure that it can handle large volumes of data and provide accurate predictions in a timely manner. For our model, this involves testing the application's response time to ensure that it can provide predictions quickly. Additionally, user testing was done to make sure that people can easily get accurate results and not face any serious monetary loss. Throughout the model, the verification plan was a continuous process to find and fix any problems prior to the release of the finished product.

Test ID	Test Case Title	Test Condition	Accuracy	Expected Accuracy
T01	Low Data	1year of data	65%	60%
T02	Adequate Data	10years of data	78%	80%
T03	More Data	20years of data	79%	80%

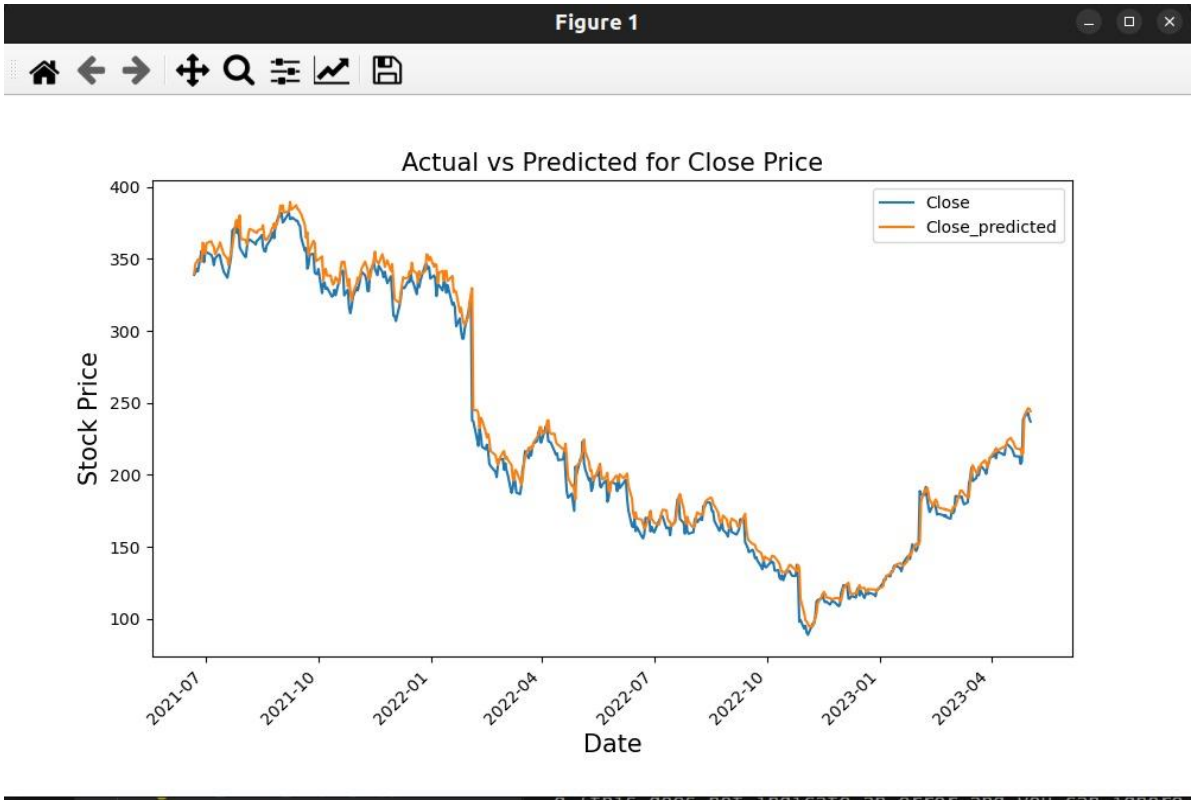

Fig. 4.3 – Model generated graph of the close values

4.3 Result Analysis

For now, since the project is still under construction, we are giving our final output using a very basic Python GUI. It takes the user input as the name of the stock and feeds it to our model which then fetches the data from Yahoo Finance, cleans the data, and gives the direction of movement first using the random forest model and then gives the output values using LSTM even though the two started simultaneously but the LSTM model is comparatively slower. Both models have high accuracy of greater than 78% for adequate data.

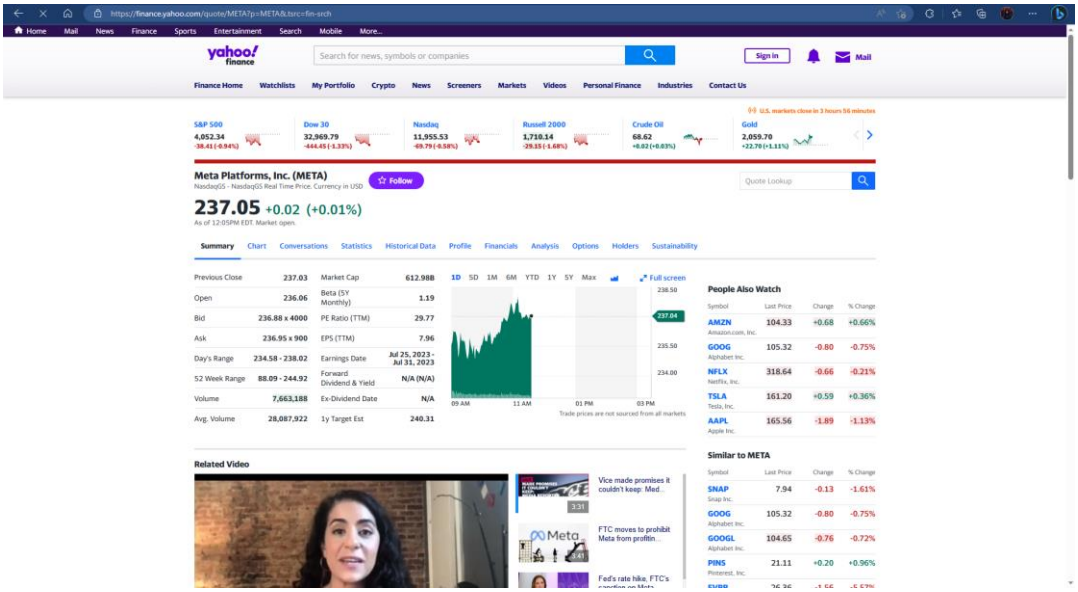

Fig 4.4 – Yahoo Finance

Fig 4.5 – Imported data

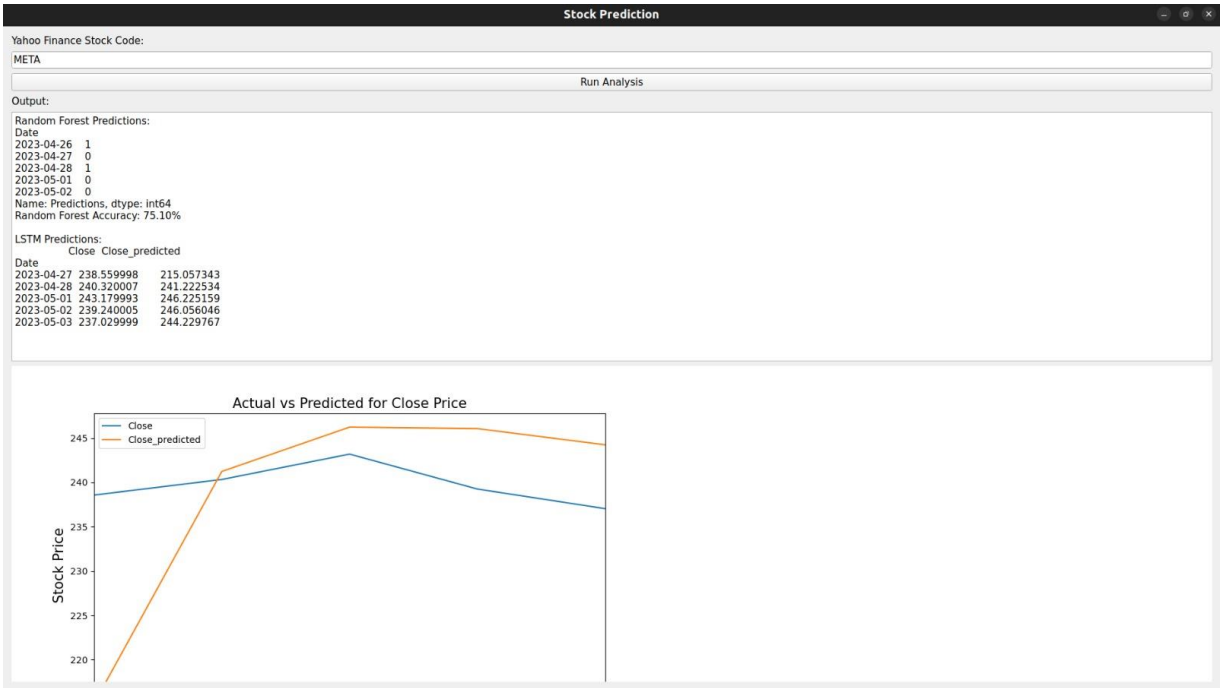

Fig 4.6 – Final Output for META

4.4 Quality Assurance

Quality assurance is a critical component of the development process for the Stock Movement Prediction Model. The goal of quality assurance is to ensure that the gloves meet or exceed the standards set by the project's stakeholders, which include accuracy, usability, and reliability. The quality assurance process for this project involved several steps, including the establishment of quality standards, the creation of test cases, and the implementation of a comprehensive testing plan. The testing plan encompassed both functional and non-functional testing, including usability testing to ensure that the model was easy to use and comfortable for individuals. The quality assurance process was an ongoing effort throughout the development of the model, with regular testing and monitoring to ensure that any issues were identified and addressed promptly. Through this process, the team was able to ensure that the final product was of the highest quality and met or exceeded the expectations of the project's stakeholders.

Chapter 5

Standards Adopted

The Time-Series Prediction model was designed, coded, and tested in accordance with best practices and standards set by organizations like IEEE and ISO. The system was created to be scalable and adaptable to allow for future advancements. We think that the system's robustness and dependability have been enhanced by the implementation of these standards.

5.1 Design Standards

We used the IEEE 1016-2009 standard for Software Design Documentation as our guide for the design guidelines we used for this project. The best practices for producing design documentation for software projects are outlined in this standard. It offers a structure for producing brief, understandable, and maintainable design papers. We adhered to this standard to make sure that our design documentation was of the highest caliber and adhered to accepted standards.

5.2 Coding Standards

The ISO/IEC 12207 standard, which offers a foundation for the software development life cycle processes, was followed whilst we were coding. We also followed the IEEE 829 standard for software testing, which establishes the structure for describing the procedures and outcomes of software testing.

5.3 Testing Standards

In terms of testing, we adhered to the IEEE 610 standard for software testing, which stipulates best practices for creating, implementing, and documenting software tests. The IEEE 29119 standard, which offers a foundation for software testing procedures and methods, was also employed.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

The proposed model has the potential to revolutionize the field of stock price prediction by combining the strengths of Random Forest and LSTM. The accurate and consistent prediction of stock prices will enable investors to make informed decisions about their investments.

Overall, the "Predicting Stock Movement using Random Forest and LSTM" project was successful in creating a software system that forecasts stock prices utilizing two potent machine learning algorithms: Random Forest and Long Short-Term Memory (LSTM). The major objective of the project was to create a tool that would allow traders, investors, and financial experts to predict stock prices with accuracy.

Data preprocessing, feature extraction, model selection, and model training were some of the stages that went into developing the system. The predictive models were constructed using the Random Forest and LSTM algorithms. Accuracy, precision, MSE, and RMSE were some of the measures that were used to assess the system's performance.

The outcomes demonstrated that the system performed with a high degree of accuracy, precision, and recall. Users of the system can receive trustworthy stock price predictions, which can aid in making wise investment choices and more effectively managing financial risk.

Overall, this study effectively illustrated how machine learning algorithms like Random Forest and LSTM can be used to reliably predict stock prices. The creation and assessment of the system can act as a starting point for further study and the creation of more complex and sophisticated predictive models.

6.2 Future Scope

Integrating Sentiment Analysis

Integrating public sentiments on the current historical data-based model will be a huge challenge. We will need to update the data based on news-based platforms and social media platforms such as Twitter, Reddit, etc. based on the use of mass hashtags and continuous monitoring of communities who are familiar with the stock market. To achieve this, we will need to apply concepts of deep learning like the convolutional Neural Network (CNN) model. We will then integrate this data into the current model to predict any significant changes that might affect the market in a drastic way.

Investment Bot

Our idea is to create a fully automatic investment bot that will help you attain the maximum possible profits without the need for constant supervision. The bot will analyze market trends and make investments on your behalf based on its analysis. By automating the investment process, the bot will be able to execute trades in real-time and take advantage of the fluctuations in the market to maximize your profits. This will save you time and effort while also giving you peace of mind, knowing that your investments are being managed by a reliable and efficient system.

References

1. Stock Market Analysis Using Supervised Machine Learning by Kunal Pahwa and Neha Agarwal.
2. Stock Analyzer and Bot Using Machine Learning by Professor Chintan Jethva, Saachi Dudani, Esa Malik, Manish Sonje, and Gaurav Tanna.
3. Stock Price Prediction Using Machine Learning study by B N Varaprasad, G. Jeevan, Y. Kalyan Chakravarti, and Ch. Kundan Kanth.
4. Prediction in Stock Price Using Python and Machine Learning by Anchal Vij, Komal Saxena, and Ajay Rana
5. https://finance.yahoo.com/
6. kaggle

INDIVIDUAL CONTRIBUTION REPORT:

PREDICTING STOCK MOVEMENT

Divyam Singh
20051346

Abstract: This abstract proposes a system for predicting stock prices using Machine Learning and Deep Learning, which aims to improve the accuracy of current practices. The system incorporates mathematical functions and external factors to achieve better prediction accuracy and profitable trades. LSTMs are utilized for sequence prediction and the random forest regression technique is used for higher accuracy with larger datasets. The objective is to predict whether the stock price will increase or decrease rather than predict the actual price**.**

Individual contribution and findings:

Our project began with brainstorming sessions and many failed ideas until I proposed the ideation of a Stock Price Prediction. From there my contribution as a member of the project team included the preparation of the code for data collection and finding appropriate libraries required for the project. I spent considerable time researching and identifying the best sources of data and libraries that would be required for the project. This involved studying different APIs and analyzing their features, benefits, and limitations to determine which ones would be most suitable for our purposes. I also played a key role in preparing the recorded data by cleaning and filtering it to ensure that it was usable for our analysis.

In addition, I was actively involved in the development of the final Python script for the project. Together with my team members, we worked tirelessly to integrate all the different parts of the project into a functional Python GUI application. We ensured that the application was user-friendly, intuitive, and could provide accurate predictions based on the data provided. We tested the script extensively to ensure that it was running smoothly and providing accurate results. Finally, we made sure that the code was well-documented and easy to understand for future reference.

Overall, I believe that my contributions to the project were instrumental in its success. The preparation of the data and the use of appropriate libraries were critical in enabling us to make accurate predictions. Furthermore, the development of the Python script was essential in bringing together all the different parts of the project and creating a functional and user-friendly application.

Individual contribution to project report preparation:

As a member of the project team, I made significant contributions to several chapters of the project report. In Chapter 1, I helped draft the abstract and introduction, which set the stage for the entire project and provided an overview of our goals and objectives.

In Chapter 2, I provided a summary of the research paper mentioned in sub-section 2.2 and absorbed the ideas proposed in the paper which later helped us with improving our outlook towards the project.

In Chapter 3, I prepared sub-section 3.3 and draw the UML diagrams as per the need. Furthermore, in Chapters 4 and 5, I helped in the methodology section and provided the team with the correct standards that suited our project.

Individual contribution for project presentation and demonstration:

My contribution to the presentation was in the Data Collection and Preprocessing section, as well as the Future Scope section. I explained how we obtained our data, the preprocessing steps we took to clean and prepare it for analysis, and the challenges we faced during this process. I also talked about our plans for future research and development in this area.

Full Signature of Supervisor: Full signature of the student:

……. ……..

INDIVIDUAL CONTRIBUTION REPORT:

PREDICTING STOCK MOVEMENT

Aryan Tripathi
20051914

Abstract: This abstract proposes a system for predicting stock prices using Machine Learning and Deep Learning, which aims to improve the accuracy of current practices. The system incorporates mathematical functions and external factors to achieve better prediction accuracy and profitable trades. LSTMs are utilized for sequence prediction and the random forest regression technique is used for higher accuracy with larger datasets. The objective is to predict whether the stock price will increase or decrease rather than predict the actual price**.**

Individual contribution and findings:

As part of this project, my main contribution was focused on building the Random Forest model for stock price prediction. First, I explored various libraries and tools that could be used for this task and identified the appropriate ones to use. Using the preprocessed data provided by my team members, I prepared the data for the Random Forest model and trained the model with the required parameters to get accurate predictions.

I also made sure to distribute tasks among my teammates, ensuring everyone was on the same page and working towards the same goal. I worked collaboratively with them, discussing the approach and ensuring that we were all in sync with each other.

Once the model was trained and tested, I integrated it with the other parts of the project, working with my teammates to create a final Python script that could be run as a GUI application. This involved carefully combining all the individual pieces of the project, ensuring that they worked seamlessly together.

Overall, my contribution to the project involved a combination of technical skills, team management, and collaboration. By working closely with my team members and leveraging my knowledge of Random Forest models and appropriate libraries, we were able to create a robust and accurate stock price prediction model that could be used as a tool for informed decision-making in the stock market.

Individual contribution to project report preparation:

I significantly contributed to various chapters of the project report as a member of the project team. I informed the team about the Random Forest method and covered all of its details in Chapter 4.

In Chapter 2, I provided a summary of the research paper mentioned in sub-section 2.1 and absorbed the ideas proposed in the paper which later helped us with improving our outlook towards the project.

In Chapter 3, I prepared sub-section 3.1 which was the project planning and described about the different steps involved in the process. Additionally, I contributed to the creation of the conclusion that was in accordance with our project in Chapters 5.

Individual contribution for project presentation and demonstration:

For the presentation, my role was to cover the slides on the Random Forest model, the Integrated Model, and the Conclusion. I started by giving an overview of the Random Forest algorithm and how it was used to make predictions in our project. Then, I talked about how we integrated the Random Forest model with the LSTM model to create a more accurate prediction system. Finally, I wrapped up the presentation by summarizing our findings and conclusions.
.

Full Signature of Supervisor: Full signature of the student:

……. ……..

INDIVIDUAL CONTRIBUTION REPORT:

PREDICTING STOCK MOVEMENT

Udit Agarwal
2005556

Abstract: This abstract proposes a system for predicting stock prices using Machine Learning and Deep Learning, which aims to improve the accuracy of current practices. The system incorporates mathematical functions and external factors to achieve better prediction accuracy and profitable trades. LSTMs are utilized for sequence prediction and the random forest regression technique is used for higher accuracy with larger datasets. The objective is to predict whether the stock price will increase or decrease rather than predict the actual price**.**

Individual contribution and findings:

As part of our project, my contribution was focused on preparing the LSTM model for prediction. I worked closely with my teammate who provided pre-processed data to use for the model training. Our goal was to build a model that could take in historical stock data and use it to predict future stock prices with high accuracy.

To begin, I first familiarized myself with the Algorithm and its various functions and features. I also researched different LSTM architectures that have been successful in predicting stock prices in the past. After selecting an appropriate architecture, I worked on preprocessing the data to make it compatible with the LSTM model. This involved scaling the data, splitting it into sequences, and creating labels for each sequence based on future stock price movements.

Once the data was preprocessed, I trained the LSTM model. I experimented with different hyperparameters, such as the number of LSTM layers, the number of nodes per layer, and the learning rate, to find the best combination that gave us the highest accuracy in predicting stock prices.

Finally, I collaborated with my teammates to integrate the LSTM model into the final Python script that would be used for our GUI application. We ensured that the LSTM model was seamlessly integrated with the other components of the script and that it worked seamlessly with the Random Forest model to provide the best predictions possible.

Overall, my contribution to the project involved a combination of research, data preprocessing, and model training, all aimed at improving the accuracy of our stock price prediction system.

Individual contribution to project report preparation:

I significantly contributed to various chapters of the project report as a member of the project team. I informed the team about the LSTM algorithm in Chapter 4 and covered all of its details in the methodology section.

I summarized the study paper described in subsection 2.3 in Chapter 2 and incorporated the concepts put out therein, which subsequently assisted us in sharpening our perspective on the project.

Additionally, I assisted with Chapter 4's Testing section and ensured that it was up to par. I also assisted with Chapter 5's Testing Standards section and gave the team the appropriate standards that were appropriate for our project.

Individual contribution to project presentation and demonstration:

For the presentation, I was responsible for the LSTM model, Real-time Testing, and Results sections. I talked about how we used the LSTM algorithm to analyze the time series data and make predictions about future stock prices. Then, I explained how we tested our model in real-time using current market data. Finally, I presented our results and discussed the strengths and limitations of our approach

Full Signature of Supervisor: Full signature of the student:

……. ……..

INDIVIDUAL CONTRIBUTION REPORT:

PREDICTING STOCK MOVEMENT

Khushi Kumari
2001912

Abstract: This abstract proposes a system for predicting stock prices using Machine Learning and Deep Learning, which aims to improve the accuracy of current practices. The system incorporates mathematical functions and external factors to achieve better prediction accuracy and profitable trades. LSTMs are utilized for sequence prediction and the random forest regression technique is used for higher accuracy with larger datasets. The objective is to predict whether the stock price will increase or decrease rather than predict the actual price**.**

Individual contribution and findings:

As part of our team project to develop a stock price prediction system using machine learning, my primary contribution was in developing the user interface for the final product using the Python GUI. I researched and utilized appropriate libraries to design and implement the GUI for the project, which enables the user to input the name of the stock and obtain its prediction.

I ensured that the GUI is user-friendly and easy to navigate, with clear instructions and feedback for the user. I also worked closely with my teammates to integrate the user interface with the rest of the project and ensure smooth functionality.

In addition to my work on the GUI, I also contributed to the development of the final Python script that integrates all parts of the project, including the Random Forest and LSTM models, the data preprocessing, and the user interface. I worked collaboratively with my teammates to ensure that the script was well-structured and efficient, and that all components worked seamlessly together to produce accurate predictions.

Overall, my contributions to the project helped to create a final product that is user-friendly and accessible, while also providing accurate and reliable stock price predictions.

Individual contribution to project report preparation:

As a team member, I made several significant contributions to various chapters of our project report. In Chapter 4, I introduced the Python GUI and included all the essential details about it in the methodology section.

In Chapter 2, I provided a summary of the study paper and incorporated its concepts to help us refine our perspective on the project. I also helped with the Quality Assurance section in Chapter 4 and ensured that it met our standards. Additionally, I contributed to Chapter 6's Future Scope section to outline potential areas for future research and development.

Overall, my contributions helped to enhance the quality of our project report and ensure that all the essential aspects were adequately covered.

Individual contribution for project presentation and demonstration:

My responsibility for the presentation was to cover the Introduction, Mission Statement, GUI, and References. I started by introducing the topic and outlining our objectives for the project. Then, I explained how we developed the GUI to make our prediction system more user-friendly. Finally, I shared the sources of information we used in our research.

Full Signature of Supervisor: Full signature of the student:

……. ……..

TURNITIN PLAGIARISM REPORT

(This report is mandatory for all the projects and plagiarism must be below 25%)