Customer Adaptive Automated Trading System with Capital Risk Analysis using Machine Learning

submitted in partial fulfillment of the requirement
for the award of the Degree of

Bachelor of Technology
in
Computer Engineering

by

Harsh Agarwal
Aditi Kandoi
Bhavya Ahir

under the guidance of

Dr. Sudhir Dhage

Department of Computer Engineering
Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Munshi Nagar, Andheri-West, Mumbai-400058
University of Mumbai
October 2020

Certificate

This is to certify that the Project entitled ”Customer Adaptive Automated Trading System with Capital Risk Analysis using Machine Learning” has been completed to our satisfaction by Mr. Harsh Agarwal, Mr. Bhavya Ahir and Ms. Aditi Kandoi under the guidance of Dr. Sudhir Dhage for the award of Degree of Bachelor of Technology in Computer Engineering from University of Mumbai.

Certified by

Dr. Sudhir Dhage
Project Guide

Dr. Anant V. Nimkar
Head of Department

Dr. Bhalchandra Chaudhari
Principal

Department of Computer Engineering
Bharatiya Vidya Bhavan’s
Sardar Patel Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai)
Munshi Nagar, Andheri(W), Mumbai-400058
University of Mumbai
April 2019

Project Approval Certificate

This is to certify that the Project entitled "Customer Adaptive Automated Trading System with Capital Risk Analysis using Machine Learning" by Mr. Harsh Agarwal, Mr. Bhavya Ahir and Ms. Aditi Kandoi is found to be satisfactory and is approved for the award of Degree of Bachelor of Technology in Computer Engineering from University of Mumbai.

External Examiner	**Internal Examiner**
(signature)	**(signature)**
Name:	**Name:**
Date:	**Date:**

Seal of the Institute

Statement by the Candidates

We wish to state that the work embodied in this thesis titled "Customer Adaptive Automated Trading System with Capital Risk Analysis using Machine Learning" forms our own contribution to the work carried out under the guidance of Dr. Sudhir Dhage at the Sardar Patel Institute of Technology. We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission.

Name and Signature:

1. **Harsh Agarwal**
2. **Aditi Kandoi**
3. **Bhavya Ahir**

Acknowledgments

It is really a pleasure to acknowledge the help and support that has gone to in making this thesis. We express our sincere gratitude towards our guide, Dr. Sudhir Dhage for his invaluable guidance, insights and encouragement which brought out our full potential and not limited it by any notional or technical constraints. Without his encouragement this work would not be a reality. With the freedom he provided,we really enjoyed working under him.

We would also like to thank our Head of Department, Dr. Anant V. Nimkar and the whole of Computer Engineering Department of Sardar Patel Institute of Technology for giving us all the facilities required to carry out this research work.

Lastly we would like to thank all our family members and well wishers for their constant encouragement for all these years, without which I could not have completed this work.

Contents

List of Figures

List of Tables

List of Abbreviations

HFT	High Frequency Trading
LSTM	Long-Short Term Memory
RCNN	Recurrent Convolutional Neural Network
NLP	Natural Language Processing
ML	Machine Learning
ReLU	Rectified Linear Unit
SMA	Simple Moving Average
EMA	Exponential Moving Average
DEMA	Double Exponential Moving Average
TEMA	Triple Exponential Moving Average
CCI	Commodity Channel Index
RoCE	Return on Capital Employed
ROI	REturn on Investment

Abstract

Stock market plays a huge role in the economy of our country. Several attempts have been made to analyse and predict the stock market. While the existing systems try to exploit the patterns of stock prices using historical data, they do not take into the account the poor performance of the system. Moreover, there is no system which provides user specific trading strategies. The proposed solution explores filtration and different trading strategies using RoCE and Machine Learning to solve the problem and predict the portfolio values. It also takes into consideration the sentiment aspect of trading using NLP and combines the two to efficiently to perform trading for even those users who have smattering knowledge about stock market thereby making it suitable for everyone.

Chapter 1

Introduction

In India, prediction and analysis of stock market is one of the most researched topics. Stock market acts as an interface for people to perform trading i.e. buy and sell stocks. Trading of stocks is one of the highly used means of investment for an individual and the country's economy is highly dependent on stock market. With the right investments at the right time, stocks can be the most effective place to invest for a person looking to earn money. However, due to the highly volatile nature of stock market, no human can impeccably predict the direction of stock market. As a result, this acts as a huge disadvantage for novice investors who have a smattering knowledge in trading activities. Every human participating in market must be at least aware of the variations taking place. Therefore, there is a need for systems and techniques that address these demands, understand the market, and extend the scope of market to novices too.

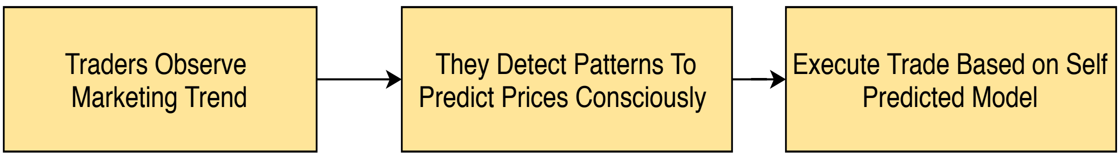


Figure 1.1: Process of Traditional Trading

These limitations by traditional trading have circumscribed the stock market penetration to less than five percent as stated in the article by financial express. The scanting removal of these limitations has been possible by making the best use of technological development. Human activity has been replaced by automated trading systems which use predefined rules that allow computers to perform and monitor the trades. The biggest advantage of rule based trading automation is that it can help the trading activities to function without the consideration of emotion. However, the existing automated systems often give rise to additional problems. These problems mainly include the inaccuracy in performance, high computation power etc. Hence, there is a need to improve the such systems by elimination of these problems and to make the system accessible for everyone.

The main objective of this paper is to develop a client-adaptive automated trading system which provides the privilege of simplicity in a way that it provides ease even to those, who do not understand its intricacy. The uniqueness of the proposed solution is that it introduces an initial filter based on the parameter: Return on

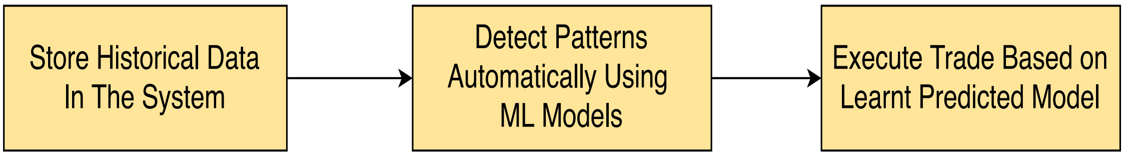


Figure 1.2: Process of Automated Trading System

Capital Employed (RoCE), to narrow down the scope of the stocks in order to reduce the computation time that leads to poor performance. Not only this but it also takes into consideration the trending stocks to further filter out the non profitable stocks. Moreover, implementing user specific trading strategies based on their risk score to increase profitability for each user and making the system entirely automated allow the system to take decisions regarding appropriate trades to generate maximum profits.

1.1 Motivation

- According to a financial express article, Indian stock market penetration is less than five percent
- Manual trading requires sufficient knowledge for people to participate in trading activities
- Trading systems consider the emotions of the market thereby increasing the chances of error
- The market due to its unpredictable and volatile nature needs to be monitored 24/7

1.2 Objectives

The objective of this paper is to develop a Client-Adaptive Automated Stock Trading System

- Filtration in the initial stage to narrow down scope of the stocks
- Prediction of stocks and determination of stocks based on sentiment analysis
- Implementation of different trading strategies for users based on their risk handling capability

1.3 Problem Statement

The limitations by traditional trading has restricted the stock market penetration to less than five percent of the population. The scanting removal of these limitations by the existing automated systems often result in giving rise to additional problems like inaccurate performance, computation power, etc.

1.4 Contributions

There has been a lot of research conducted on automated trading system and stock market using Machine Learning to suggest trading actions to users. Likewise, research has been done on predicting stock prices too. Our contribution to the financial sector includes bringing both these areas of research together along with sentiment aspect of trading. Moreover, introducing a system that is robust as well as user-friendly, making it easily available to all masses who do not have any prior knowledge in stock market. There has not been much development on the amalgam of the two researches.

1.5 Layout of the Report

Chapter 1: A brief introduction of the topic to be studied is presented here.
Chapter 2: A literature review of the existing solutions on automated trading system is presented.
Chapter 3: System Design along with the application flow of the system will be described in this chapter.
Chapter 4: In this chapter, a description of the implementation is carried out which involves the different modules of the system.
Chapter 5: The results obtained from the implementation of our work is stated in this chapter with respect to the objectives stated. Chapter 6: The takeaways from the project are mentioned in this chapter in the form of a conclusion.
Chapter 7: The future scope of this research work is stated in this chapter.
Chapter 8: Research Publication will be presented in this chapter.

Chapter 2

Literature Survey

The literature review consists of a number of papers on RoCE components, Sentiment Analysis, Price Prediction, Risk Analysis and Trading Bot. The features that are common across the previous work include study stock trends and prediction of stock prices.

Eli Amir and Itay Kama [1] examined the investor reaction to return on capital equity and its components during its quarterly earnings. The importance of net profit margin (NPM), asset turnover (ATO) and leverage (LEV) was considered relative to each other using Fama–MacBeth regression and portfolio analysis techniques. Using this, the beta and the risk values were found to determine the asset prices. It was observed that a high level of NPM led to a more positive market reaction as compared to the other components and vice versa. Any levels of the other components did not affect the market reaction.

Joshi et al. [2] proposed a solution to discover future trends of a stock using news articles about a company with the help of fundamental analysis techniques. The news was classified as good(positive) and bad(negative). The sentiment analysis along with the determination of the relationship between stock price and the news articles was done using supervised machine learning and text mining techniques. Polarity was assigned to an article based on the stop words mentioned in a dictionary. They compared three classification models: Random Forest which gave the highest accuracy ranging from 88% to 92%, SVM which worked well giving an accuracy of 86% and Naive Bayes having 83% accuracy. However, the dictionary created consisted of a limited financial terms which results in less efficient performance by the models. Unlike Joshi, Guangyu Ding and Liangxi Qin [3] proposed a solution to predict multiple value outputs using deep recurrent neural network having multiple inputs and outputs using a long short term memory (LSTM) network. The backtesting was done using multiple data sets and then compared with the two models individually. Since, the network gave multiple outputs, the model could predict open price, lowest price and highest price of any stock at the same time thereby reaching an accuracy of over 95%. However, this model does not take into account the relationship between the user and their loss handling capability which makes the system less precise.

The aim of the paper [4] is to analyse the risk and return values of the banking by considering Nifty Index and compared fifty stocks. They found the relationship between returns and volatility with Standard Deviation. As a result, it was observed that all the trending and top stocks had a positive beta value based on the market index and the less volatile stocks had a low beta value. However, the system did not successfully prove to be accurate because it did not analyse the market continuously to select the top stocks.

In recent times, with the increase of complex algorithms and the availability of powerful computers, there is a lot of demand for algorithmic trading. We studied various trading strategies for our system. In [5], trading decision methods were proposed based on a multiple classifier system and candlestick patterns. The trading strategies were evaluated using Bollinger Bands and Parabolic SAR indicators. The performance of the proposed decision strategies was tested by developing a web application which uses the end of day stock prices, and the stocks were used to represent the up and down trends. The best strategy for up trend stocks gave around 17% profit and around 1% for sideway market trend. Similarly, in down trend stocks, the loss was minimized to 2.62%. The paper [6] compared the use of three famous trading strategies that is SMA, MACD and PIVOT. The strategies were backtested using the d Backtest PS method. The systems were run for 1.5 years and closing price of the data was considered. They concluded that MACD>PIVOT>SMA in terms of the profits generated by each.

Due to the unpredictable events that completely change the trend of the market, most algorithmic traders opt for High Frequency Trading (HFT). In this the brokers or the traders execute a large number of trades in a short period of time to extract profits from low changes in prices. The paper [7] proposes the use of Grid Trading Algorithms to be used on FOREX markets to execute these trades. The system places orders in regular time intervals to generate steady profits from these deals. Regression network and trend classifier was used to predict which buying and selling of a particular stock. The paper achieved a good Return on Investment(ROI) of 13.76%.

L. Chen and Q. Gao [8] aim at suggesting right decisions in stock market using Deep Q-network (DQN) and Deep Recurrent Q-network (DRQN). The two methods were compared and it was observed that DQN easily learns profitable patterns from the historical stock market data. However, introduction of recurrence to DQN led to an improvement in the performed in the trading actions. Akhil Raj et. al. [9] proposed the use of deep reinforcement learning model in an automated trading system. RCNN was used to predict sentiment from the stock market news data which was fed into the reinforcement learning neural network model. This predicted the trend and made Buy/Sell decision to maximize the profit. The trading bot then executes those decisions in a robust manner. The model had an accuracy of 96%, this means the model was extremely efficient in predicting the sentiment and the trend of the market.

In [10], Prasetijo et al. proposed trading strategies using Bollinger Bands and Parabolic SAR indicators. It also built a web application to perform backtesting of the strategies. By taking into consideration the end-of-day stock price for different LQ-45 stocks, the system categorized into up, down and sideway trends. Bollinger Bands generated profit 17.06% for up-trending stocks while it gave profit of 1.19% for sideway market trend. Parabolic resulted in losses for all the trends hence, it worked the worst in comparison to Bollinger Bands. C. N. W. Tan [11] simulates a trading based on the two years NYSE stocks data using an artificial neural network based financial trading system. The system fit quite well giving an accuracy of 83%

with a tolerance of 5%. The reliability of the system is poor as it did not consider several parameters affecting the neural network.

The literature survey indicates that an automated trading system is very efficient and effective to generate huge profits. Therefore, by implementing an automated trading system, this paper proposes an all in one system that reduces the computation time and make it less expensive, suggest different trading strategies to users, analyse risk and return, perform trades and predict the portfolio values to make the system customer adaptive.

2.1 Research Gap Identified

1. **No filters in the initial stages**
   While performing automated trading, the entire market is analysed to determine the best stocks. Market data consists of irrelevant data during the trading process. Analysing the entire market results in poor process performance due to high computation time. System becomes expensive and cannot be used in customer applications

2. **Do not consider company's profitability**
   Selection of stocks is done by analysing the market trend and sentiment. Important factor that affects the stock prices includes a company's profitability or its earnings. Focusing on just the market trend does not perform selection of stocks accurately

3. **Lack of user specific trading strategy**
   Existing systems are based on generalised trading strategies for all users. A user might have a different risk handling capability than some other user. Therefore, having the same trading strategy for every user will not result in maximising the profit

4. **Requires user's input to perform trading**
   Existing systems give suggestions regarding the strategies and the decision to the user. Final decision has to be taken by the user to perform trades. However, this may give rise to errors while setting up the input by the users

Chapter 3

Automated Trading System Design

3.1 Trading Bot System Design

Fig 3.1represents the architecture of our system. The system is divided into 8 components

- **Bot Database:** Bot Database is the central store that contains data from all the modules, comprising of stock data, user data, transaction data, user profile, and other metadata. It's the core of the trading bot.
- **Market Data API:** This module gets data for daily trending stocks from different websites and passes it to the ROCE module to filter out the stock tickers.
- **Market News Data:** Market sentiment (also known as investor attention) is the general prevailing attitude of investors as to anticipated price development in a market. This attitude is the accumulation of a variety of fundamental and technical factors, including price history, economic reports, seasonal factors, and national and world events.
- **Calculate ROCE:** Return on capital employed is an accounting ratio used in finance, valuation, and accounting. It is a useful measure for comparing the relative profitability of companies after taking into account the amount of capital used. It helps in filtering out the stocks which are passed o from Market Data API.
- **Web app Front-End:** Web application for users to use this Trading Bot and see all transactions being made by the bot, keeping track of user portfolio, performance on each stock, and other related information.
- **Order Management API:** This module is responsible for placing the stock order on the basis of the decision made by the stock trading bot by using a paper trading platform.
- **Evaluation and Backtesting:** Backtesting is a term used in modeling to refer to testing a predictive model on historical data. Backtesting is a type

of retrodiction, and a special type of cross-validation applied to the previous time period(s). This module performs Backtesting on the selected Trading Algorithms for a period of 2 years. Evaluation refers to the process of calculating the risk associated with the algorithms. ROI, Sortino ratio, and Sharpe ratio are the measures that are used to define the risk and we will see these factors in-depth in the implementation section.

- **Trading Algorithms:** In finance, trading algorithms is a fixed plan that is designed to achieve a profitable return by going long or short in markets. The main reasons that a properly researched trading strategy helps are its verifiability, quantifiability, consistency, and objectivity. Trading algorithms are based on fundamental or technical analysis or both. They are usually verified by backtesting, where the process should follow the scientific method, and by forwarding testing (a.k.a. 'paper trading') where they are tested in a simulated trading environment.

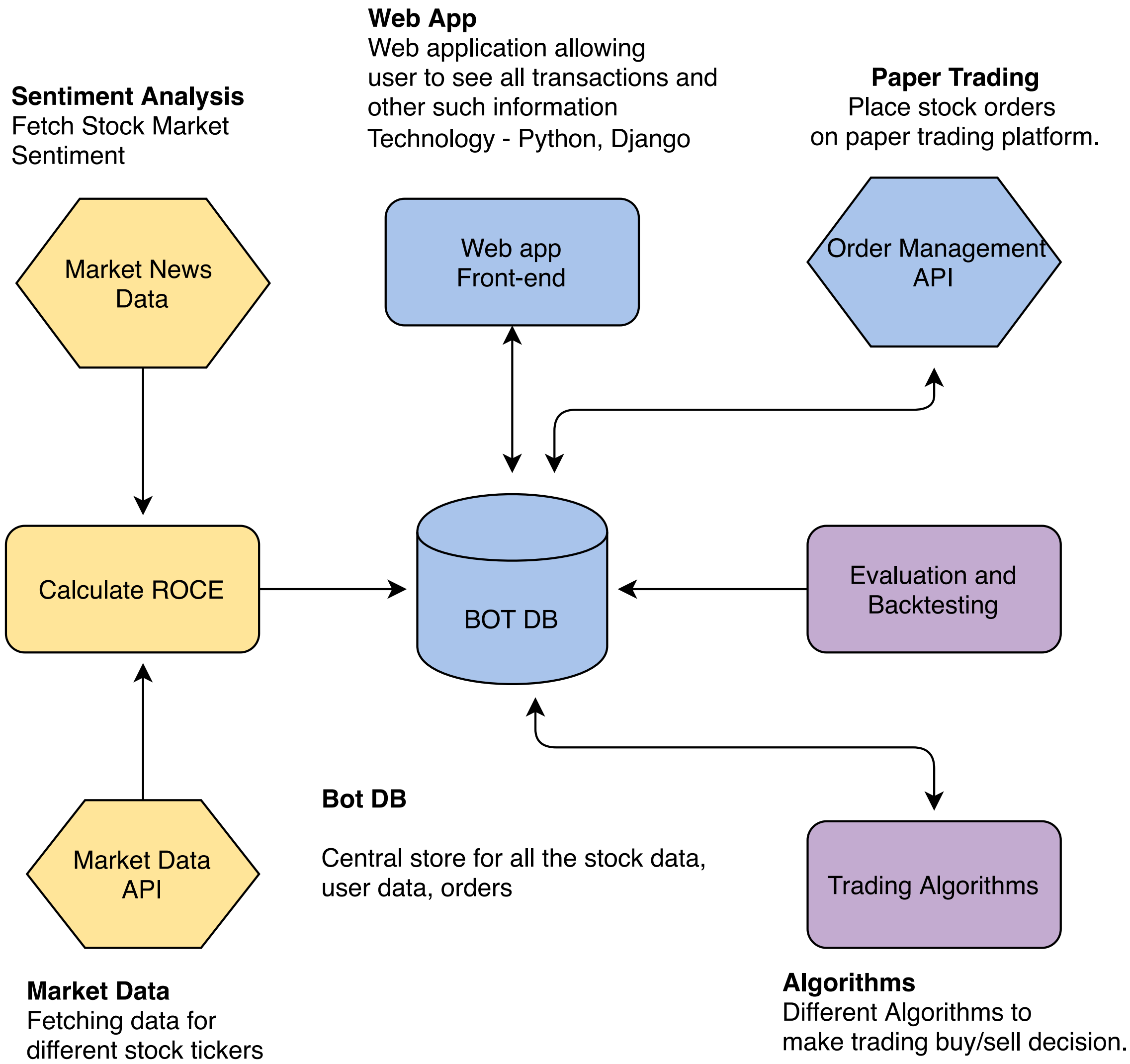


Figure 3.1: Proposed System Architecture for Trading Bot

3.2 Internal Architecture of Trading System

The proposed system comprises of six modules as shown in Figure 3.2. Each module is explained below.

Data Scraping Module

The system requires a lot of financial real-time and accurate data to work properly. As of now, NSE consists of 1837 listed companies, and manually acquiring all the data is difficult. We have defined a Data Scraping Module to help us collect this data. The script web-scrapes all the data from reliable sources to generate a database. Two types of data are collected using this script:

1. Financial Records of the Companies
   The Financial Records consist of the quarterly bank sheet, cash flow statement, and the income statement of all the stocks. These records are required for accurately predicting and trading stocks.

   Beautiful Soup with Selenium and YahooFinancials is used in the module to collect all this data. YahooFinancials is a python library that scrapes the yahoo finance website to generate financial records. The data was gathered in 10 different batches to prevent the IP from being blocked by the website. Total 1466 companies were scrapped as the remaining tickers were unavailable on yahoo. For the remaining set of companies, BeautifulSoup with Selenium was used to scrape from different sources like moneycontrol.com, Finviz India, or Screener. BeautifulSoup parses the HTML page to get a particular ID which consists of required data and then Selenium repeats this task over different threads over the stock ticker list. In total financial data of **1573** companies was collected and stored.

2. Latest News Data
   The news articles related to the stock market play a major role in determining the trending stocks. The data obtained from news articles is required to precisely performing trading of stocks.

   The live news articles are scraped from 'Finviz India' website using BeautifulSoup. The data consists of the headlines of the news articles, its date and time. This is then stored in the form of data frames using the python library 'Pandas'.

Data Filtering Module

Due to the huge size of data, the time and computation power needed to process it increases exponentially making the application useless as the trading orders needs to be executed as fast as possible. In the proposed application, we have defined a unique filter to address this issue. We use RoCE (Return on Capital Employed) as the basis of our filter. This ratio indicates how profitable a particular company is and how effective their investments are. The ratio is given by:

$$RoCE = \frac{\text{Earning Before Interest and Tax (EBIT)}}{\text{Total Assets - Current Liabilities}} \tag{3.1}$$

According to the latest research, a company is considered to be profitable if the RoCE value is greater than 20. The higher the value is the higher company's profitability and in turn can be considered a good investment. RoCE is calculated from these financial records using Equation 3.1. From all the companies we get a final list of 30 companies, and trading is performed on these companies.

Also, the use of this module is for removing unnecessary data from the Live Market Data. Data is cleaned of any abnormalities and preprocessed for future prediction modules. Preprocessing involves converting to DateTime data, normalization of the stock price, and cleaning missing and noisy values. Normalized values of the portfolio are required in the LSTM neural net module to predict the user's portfolio. We use MinMaxScaler from Scikit-Learn preprocessing library. The feature_range is set between 0 and 1. These scaled values improve the performance of the prediction module hence providing better accuracy in forecasting.

Natural Language Processing Module

The stock market's volatile nature is due to the fluctuations in demand and supply. The relation between demand and supply is extremely sensitive to the news articles at that instant. By analysing the sentiments of news articles, information on the trending stocks can be known and contributes to a good stock picking strategy. In the proposed system, an NLP module is implemented to help in assigning sentiments to the articles as well as obtaining the trending stocks.

The news articles' data obtained is cleaned initially to remove the stop words which do not play any role in assignment of sentiment to the article. Natural Learning ToolKit Vader is used to express sentiment polarity and intensity to the news articles. Once the data is cleaned, the lexicon approach and rules come into picture to scan the data to find lexicons. It then finds the sentiment features and calculates the news article's compound score which ranges from -1 to 1. Higher the value of the score, more positive is the sentiment of the article and vice versa.

Prediction Module

The website has a dashboard where the users have an option to view how their trading strategies will perform in the future. This prediction is done by using LSTM (Long-Short Memory Model) neural net models. This model improves the traditional RNN models and is been known to achieve the best accuracy in predicting time-series forecasting.

The model is trained with the historical data of users portfolio worth from a particular trading strategy. For our purpose, we used the data obtained from back-testing the strategies. The data consisted of 676 days of trading from 1st January, 2018 to 1st October, 2020. For testing of the model we used the trades of last 60 days. The RELU fuction is used as the activation function with Adam optimizer used to optimize the output. Mean squared error is used as the loss function. The model is then fitted and run for 150 epochs. Then the predictions are made and the graphs are plotted and shown on the website dashboard.

Decision Making Module

The Decision-making module is responsible for making buy/sell decisions. It's the core of the Trading Bot. It comprises 3 sub-parts namely User Portfolio Data which is a storage unit for user data, Trading Strategies that helps to make buy/sell decisions, and Risk analysis which help improve the decisions made by the Trading Strategies.

User Portfolio Data: The User Portfolio is a central database for storing all the data received from different modules. It stores information about all the stock held by the users, the transactions made by the system, the risk analysis, and evaluation of the trading strategies. On every new financial day, the trending stock tickers which are scrapped by the module are saved in this database, and these tickers are stored temporarily as every new day we will get new tickers. These tickers are then passed on to the Trading Strategies which perform analysis on these tickers and make buy/sell decisions accordingly.

Trading Strategies: Trading Bot can make transactions and analyze the market at a speed which is impossible for a human to achieve and hence can make profits quickly. The program mainly focuses on price, variation in prices, quantity and stock trends, and various such factors to make buy/sell decisions. For this project the Trading Strategy that we are using are:

1. **Simple Moving Average(SMA) Strategy**
   It calculates the avg of a range of closing prices, by the number of periods in that range. This can help to analyse a bull or bear trend.

2. **Exponential Moving Average(EMA) Strategy**
   EMA analyzes the predominance of a trend in the market. EMA is a line on a price chart that smoothens out the price action by using mathematical formulae. It puts more emphasis on the recent price, which makes it more reliable as it reacts faster to the latest changes in price data. The calculation for EMA is done as followed:

$$\text{Initial SMA} = 20 - periodsum/20 \tag{3.2}$$

$$\text{M}ultiplier = \frac{2}{Timeperiod + 1} \tag{3.3}$$

   EMA = Close - EMA(previous day) $\times$ $multiplier + EMA(previousday)$

3. **Double Exponential Moving Average (DEMA) Strategy**
   The traditional moving averages experience a lag that can be removed by using DEMA since it smoothens price fluctuations. This helps to confirm uptrend when price moves above average and similarly downtrend. DEMA can be calculated as followed:

$$DEMA = (2 \times EMA_n) - (EMA\ of\ EMA_n) \tag{3.4}$$

$$where \text{ n= Look-back period} \tag{3.5}$$

4. **Triple Exponential Moving Average(TEMA) Strategy**
   The traditional moving averages experience a lag that can be removed by using TEMA since it smoothens price fluctuations. TEMA can be calculated as followed:

$$TEMA = (3 \times EMA_1) - (3 \times EMA_2) + EMA_3 \tag{3.6}$$

$$EMA_1 = EMA\ of\ EMA_1 \tag{3.7}$$

$$EMA_3 = EMA\ of\ EMA_2 \tag{3.8}$$

5. **Bollinger Bands Strategy**
   Bollinger Bands are envelopes plotted at a standard deviation level above and below a simple moving average of the price. They adjust to volatility swings in the price since the distance of the band is based on standard deviation. Period and Standard Deviations are the two main parameters of Bollinger Bands, the default value being 20 and 2 respectively. They help to predict whether the prices are low or high on a relative basis. Both upper and lower bands are used in pairs and in conjunction with Standard Deviation.

   The steps to calculate Bollinger Bands are as follows:

   (a) Calculate a simple moving average.
   (b) Calculate the standard deviation over the same number of periods as the simple moving average.
   (c) Add the standard deviation to the moving average, for the upper band.
   (d) Subtract the standard deviation from the moving average, for the lower band.

6. **Commodity Channel Index Strategy (CCI):** CCI Strategy calculates something called as CCI line – which is a function of the last price, mean deviation of past prices, and simple moving average (SMA) of past prices. It is dependent on mean deviation impact, a number of periods, and user-defined parameters.

$$CCI = \frac{\text{Last price} - \text{SMA of last n prices}}{\text{Impact factor * mean deviation of n prices}} \tag{3.9}$$

   Where n is equal to the Number of Periods parameter. Mean deviation (after receiving *(k+n)* prices, $k >= 0$ is defined as follows:

$$\frac{\sum_{i=k}^{k+n} |Price(i) - \text{SMA of last n prices}}{n} \tag{3.10}$$

   When the CCI line crosses the overbought level or oversold level it indicates that the strategy should make an order.

Risk Analysis Module: Risk analysis module calculates the risk associated with the trading decision made by the bot on historical stock data. It is a type of cross-validation applied to the previous time period. This module performs analysis on the selected Trading Algorithms for a period of 2 years. It refers to the process of calculating the risk associated with the algorithms. ROI, Sortino ratio, and Sharpe ratio are the measures that are indicators that were used for this purpose.

Trade Execution Module

This module is responsible for placing stock orders passed onto it by the Decision-making module on the paper trading platform. A paper trade is a simulated trade that allows an investor to practice buying and selling without risking real money.

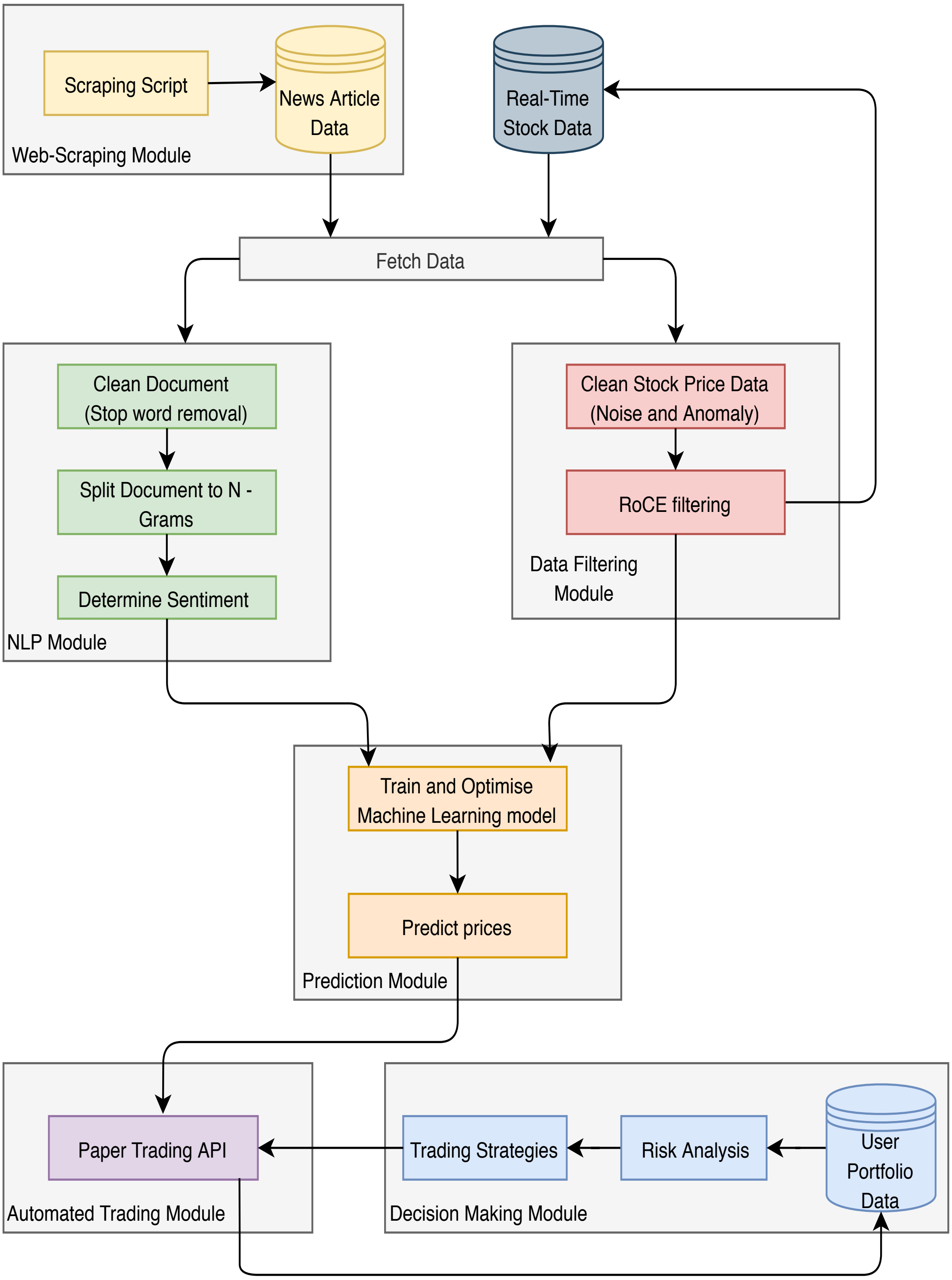


Figure 3.2: Internal Architecture with Various Modules

3.3 User Interaction Diagram

Figure 3.3 explains how the user will interact with the application. The main purpose of the project is to reduce the user's workload and perform trades automatically. Our application has a minimal number of steps that the user needs to perform to start earning money. The system is supposed to operate as follows:

1. The application checks if the user is registered or not. If the user is not registered the application redirects to the CREATE PORTFOLIO PAGE or the HOME PAGE if the user is registered.

2. A portfolio defines the user's net worth with respect to the stocks and cash they hold. After successfully creating the portfolio, the user needs to add money to the application.

3. The user can then define and set limits according to their wishes. These limits can be the duration they want to test on, the expected ROI, the loss they are ready to bear, and the initial ticker(if any) they want to try on.

4. These parameters go into the *Trading Strategy Decision Module*. This module recommends a trading strategy according to the parameters set by the user. After successfully choosing a given strategy the user is again directed to the HOME PAGE.

5. Once on the HOME PAGE, the user has an option to perform 3 tasks, that is, VIEW PORTFOLIO or START TRADING or RESET THE STRATEGIES.

6. VIEW PORTFOLIO option allows the user to check their current worth and also allow them to run a prediction. The *Prediction Portfolio Module* enables the user to check what their profile might look like in the future. These graphs from the predicted values help the user decide if they want to keep using the same strategy or invest in another one.

7. START TRADING option allows the user to start investing if the market is open at that time. The user does not need to perform any action. There are 4 MODULES that handle trading automatically. As explained above, the *Data Scrapper Module* gives news and stock data which passes to the *Filtering and Sentiment Module* to give the filtered data with sentiment value of the tickers. This data is then passed to the *Decision-Making Module* where the trading orders are generated according to the strategy selected by the user initially. After which the orders are executed on the *Executing Orders Module* where the paper trading APIs manage in real-time.

8. RESET TRADING option redirects the user to the Define Limits and Parameters Page where the user can again calculate RISK SCORE and select a new Trading Strategy

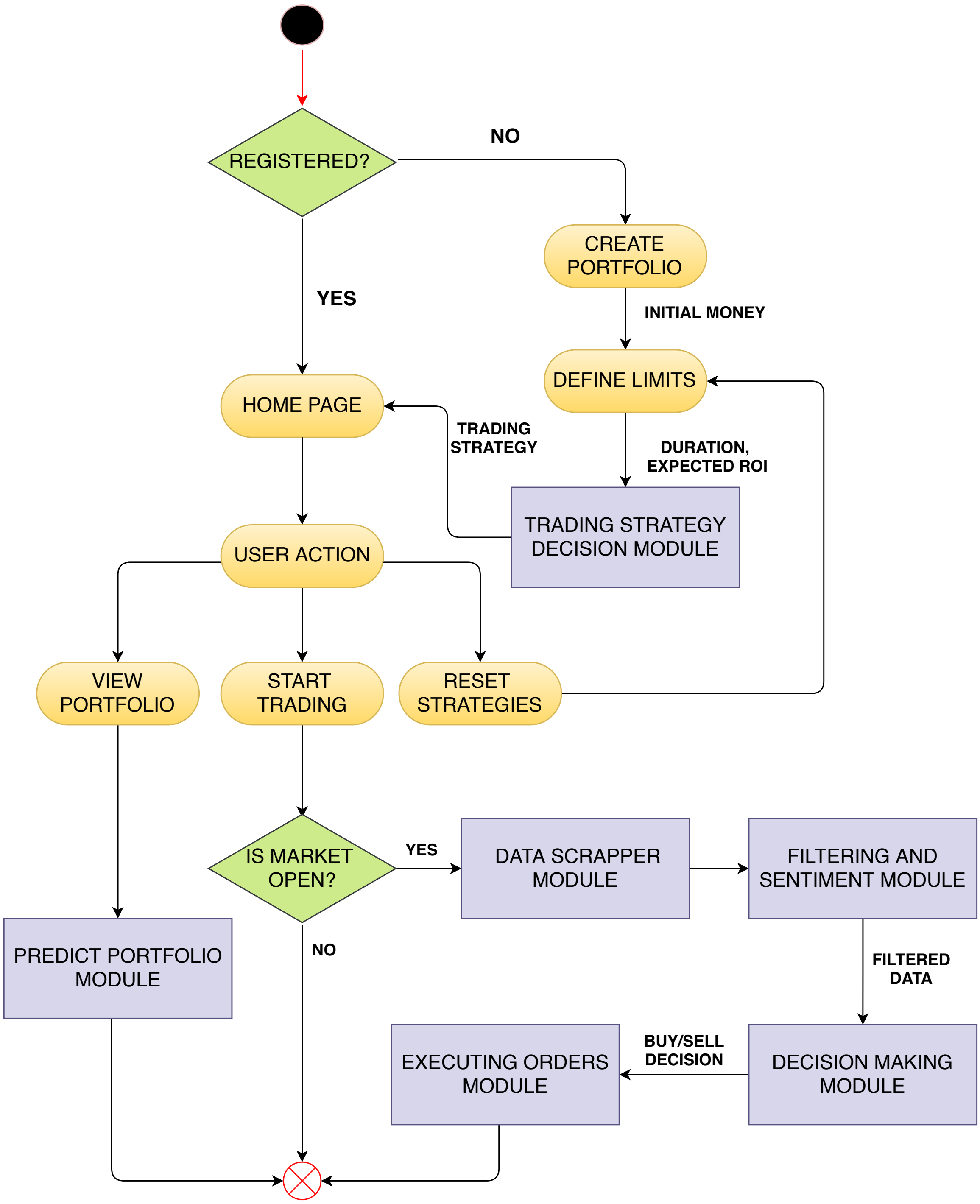


Figure 3.3: User Interaction Diagram

Chapter 4

Results and Discussion

The aim of this research was to fulfill the objectives mentioned earlier. In order to achieve those objectives a system consisting of different modules was put together and following results with the respect to the mentioned objectives were achieved.

4.1 Sentiment Analysis on News Articles

News articles play an important role in predicting the stock prices and performing buy and sell actions in trading. By using the Vader analyzer, the compound scores of all tickers were calculated and compared. The compound scores for three Indian stocks: HDFC Bank, Infosys and Tata Motors was calculated based on their news articles on four days.

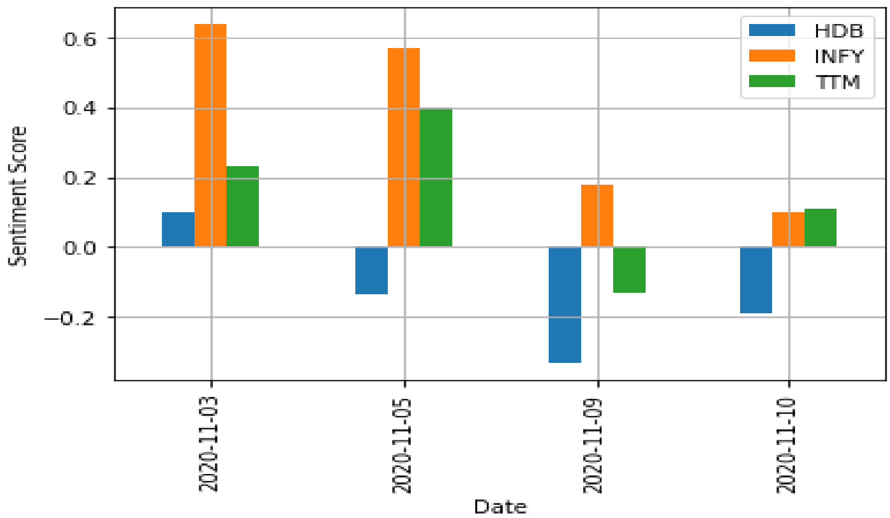


Figure 4.1: Comparison of Tata Motors, HDFC Bank and Infosys Stocks

It can been observed from Figure 4.1 that Infosys had a very high score on November 3, 2020, this means that there were several positive news articles about that company. Hence, it can be infered that the stock price of INFY would be high based on its demand. Similarly, it can also be viewed that Tata Motors and HDFC Bank has negative score on November 9, 2020. This shows that there were negative sentimental articles on those companies which contributed to the low prices of their stocks.

4.2 Prediction of stock prices

Future values of the user portfolio can be calculated using LSTM neural net model. Figure 4.2 represents the user's portfolio over time. Blue line indicates the training data for the model. Yellow and orange lines indicate the training and validation over the period of 2 months. As seen in Figure 4.3, it can be clearly inferred that the model is accurate in predicting results in the future and has a train loss of 0.001 and validation loss of 0.02. The model predicted that the user's portfolio will increase by 12% over the next week.

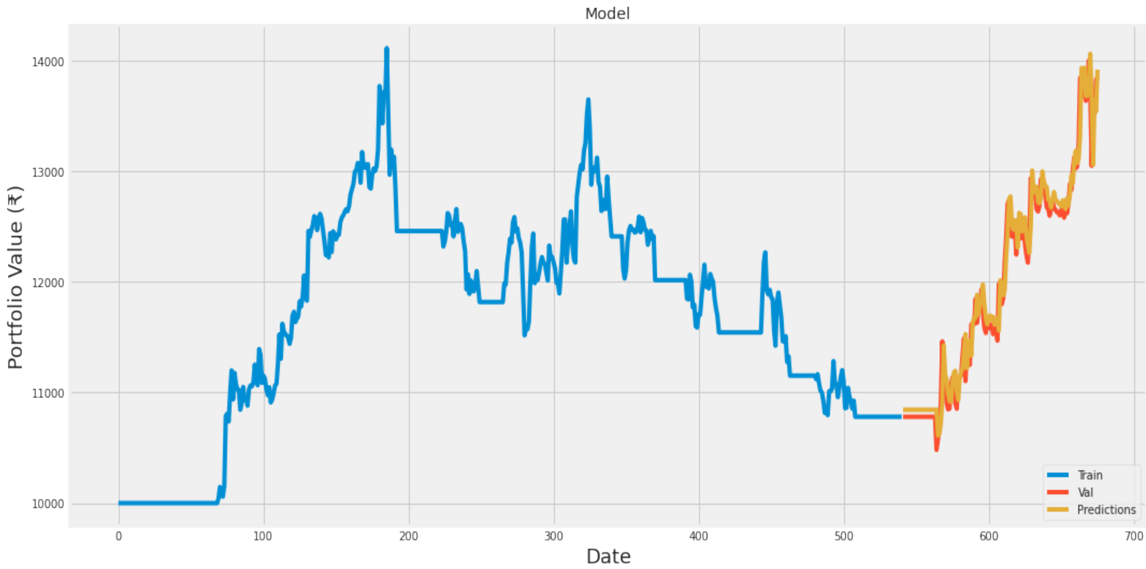


Figure 4.2: User Portfolio Prediction

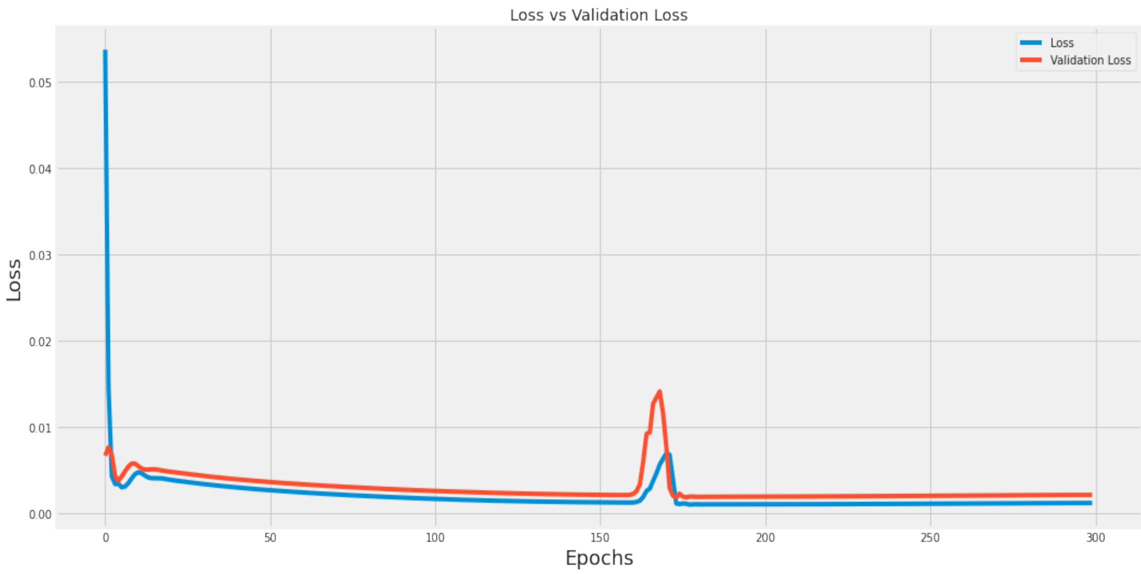


Figure 4.3: Loss and Validation Loss

4.3 User specific Trading Strategies

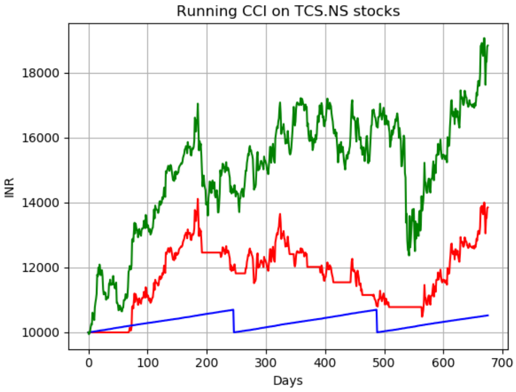
One of the primary objectives was to implement different trading strategies and find the best fit according to the risk associated with those strategies. Bollinger bands, CCI, Moving momentum and various moving averages based Trading Strategies were implemented in the system. Backtesting was done on these trading strategies by initialising the amount at hand as Rs 10,000 and for a period of 2 years for stock TCS. Figure 4.4 compares backtesting of various trading strategies. This comparison shows how the strategies will react to the change in price. Green line indicates the company's stock price, red line indicates the user's portfolio and the blue line is the risk free return which is set to 7% as the standard market rate.

As seen in Table 4.1, TEMA yields the best result for the period of time taken into consideration. This results might change with change in duration and other influencing factors that affect the stock market.

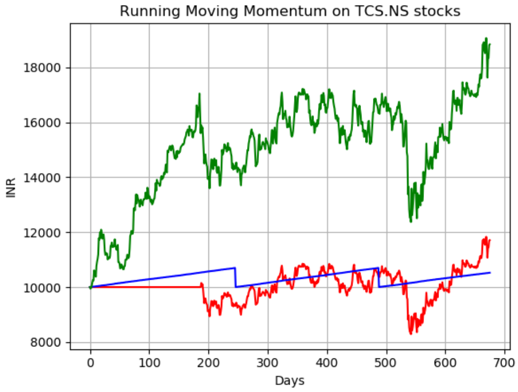
Table 4.1: PROFIT GENERATED AFTER BACKTESTING

Strategy Name	**Average Profit**	**Total Profit**	**Backtesting Period**
CCI	5.5	3851	700
Momentum	2.4	1710	700
SMA	4.7	3350	700
EMA	7.5	5252	700
DEMA	7.9	5573	700
TEMA	8.6	6068	700

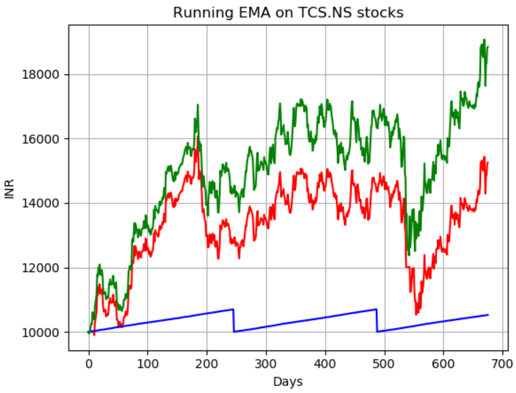
In figure 4.4 green line indicates the price of stock during the backtesting period, red line indicates the user's portfolio worth and blue line indicates risk free trade valuation during the same duration.



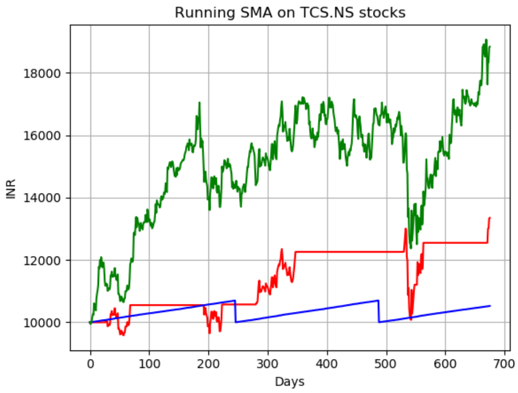
(a) CCI



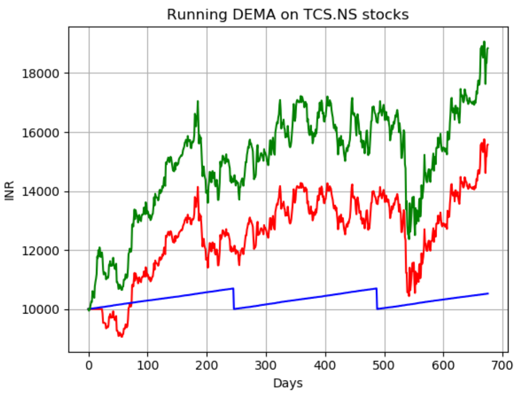
(b) Moving Momentum



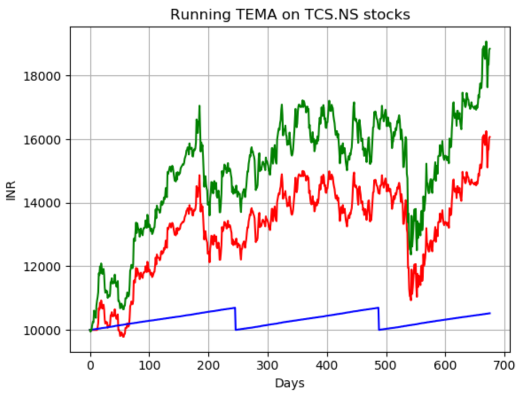
(c) EMA



(d) SMA



(e) DEMA



(f) TEMA

Figure 4.4: Evaluation graphs for Trading Strategies

Chapter 5

Conclusions

Stock market is an important aspect in the financial industry. A lot of research has been carried out in the past to automate the trading system. With recent advancements in the technology sector, there is an increasing impetus to optimise automated system and make it user adaptive. The proposed system uses both stock market data and news articles related to them to perform trading. From this, using RoCE filter, the stocks that are not profitable and unnecessary are removed. The sentiment analysis of stocks is also performed using NLP to generate the trending stocks. Trending stocks are the ones whose news articles have high positive sentiment values. Using this combined filtered data, the system performs trading system using one of the several trading strategies that satisfy the user requirements. With the help of this, the solution can help to perform trading easily and efficiently. In addition, the system also predicts future value of users' portfolios using LSTM. The validation loss of the model was 0.02, hence, it fit well with the system. These novel features make the proposed system robust, user-friendly and one of a kind application which is available for all masses.

Chapter 6

Future Scope

The most difficult challenge we faced was the abundance of financial data that needs to be scraped before implementing the strategies. This data is not available freely and the scraping module in the system can only retrieve data from specific websites as of now. As a part of our future scope, we can increase the efficiency of the scraping module and process more data for the system. Also the backtesting of trading strategies take a lot of time and manpower and this can be deployed on the cloud to remove all the manual steps in the process.

Due to ever increasing demand of Machine Learning in finance sector, new and improved algorithms will be kept on developing for automated trading. This poses a challenge in future where in we have to adapt and incorporate the strategies in our project.

Chapter 7

Research Publication

Customer Adaptive Automated Trading System with Capital Risk Analysis using Machine Learning

Harsh Agarwal, Aditi Kandoi, Bhavya Ahir, Sudhir Dhage
Department of Computer Engineering,
Sardar Patel Institute of Technology,
Mumbai, 400058
harshagarwal2703@gmail.com, kandoiaditi0@gmail.com, bhavya.ahir@spit.ac.in, sudhir_dhage@spit.ac.in

***Abstract*—Stock market plays a huge role in the economy of our country. Several attempts have been made to analyse and predict the stock market. While the existing systems try to exploit the patterns of stock prices using historical data, they do not take into the account the poor performance of the system. Moreover, there is no system which provides user specific trading strategies. The proposed solution explores filtration and different trading strategies using RoCE and Fuzzy Logic to solve the problem and predict the portfolio values. It also takes into consideration the sentiment aspect of trading using NLP and combines the two to efficiently to perform trading for even those users who have smattering knowledge about stock market thereby making it suitable for everyone.**

***Index Terms*—Automated System, Algorithmic Trading, Price Prediction, Data Analysis.**

I. Introduction

In India, prediction and analysis of the stock market are the most researched topics. The stock market acts as an interface for people to perform buying and selling of stocks. With the right investments at the right time, stocks can be the most effective place to invest for a person looking to earn money. Moreover, the country's economy is highly dependent on the stock market. However, due to the volatile behaviour of the market, no human can impeccably predict the direction of the stock market. As a result, it is a disadvantage for novice investors who have only a smattering knowledge in trading activities. Every human participating in trading activities must be at least aware of the variations taking place in the stock market. Therefore, there is a need for systems and techniques that address these demands, understand the market, and extend the scope of the market to novices.

The limitations by traditional trading have circumscribed the stock market penetration to less than five percent as stated in the article by financial express [1]. The scanting removal of these limitations has been possible by making the best use of technological development. Human activity has been replaced by automated trading systems which use predefined rules that allow computers to perform and monitor the trades. The biggest advantage of rule-based trading automation is that it can help the trading activities to function without the consideration of emotion. However, the existing automated systems often give rise to additional problems. These problems mainly include the inaccuracy in performance, high computation power etc. Hence, there is a need to improve such systems by the elimination of these problems and to make the system accessible to everyone.

Therefore, a novel solution is needed which can:

- Provide an efficient system that maximizes profits
- Provide an easy interface for amateur users who require no prior knowledge
- Perform trading in a robust manner with near to zero errors
- Provide custom trading strategies to users based on their risk score

The paper aims to develop a client-adaptive automated trading system which provides the privilege of simplicity in a way that it provides ease even to those, who do not understand its intricacy. The uniqueness of the proposed solution is that it introduces an initial filter based on the parameter: Return on Capital Employed (RoCE), to narrow down the scope of the stocks in order to reduce the computational time that leads to poor performance. Moreover, implementing user-specific trading strategies based on their risk score to increase profitability for each user and making the system entirely automated allows the system to make decisions regarding appropriate trades to generate maximum profits.

The content of the paper is divided into five sections. In Section II, the literature review of the existing automated trading systems and its limitations are summarized. Section III introduces the proposed solution which discusses the architecture of the proposed automated system and its flow. Section IV discusses the obtained results of the work. The paper concludes with Section V which details the conclusion.

II. Literature Survey

The literature review consists of various papers related to RoCE components, Sentiment Analysis, Price Prediction, Risk Analysis and Trading Bot. The features that are common across the previous works include study stock trends and prediction of stock prices.

Eli Amir and Itay Kama [2] examined the investor reaction to return on capital equity and its components during its quarterly earnings. The importance of net profit margin (NPM), asset turnover (ATO) and leverage (LEV) was considered relative to each other using Fama–MacBeth regression and portfolio analysis techniques. Using this, the beta and the risk values were found to determine the asset prices. It was observed that a high level of NPM led to a more positive market reaction as compared to the other components and vice versa. Any levels of the other components did not affect the market reaction.

Joshi et al. [3] proposed a solution to discover future trends of a stock using news articles about a company with the help of fundamental analysis techniques. The sentiment analysis along with the determination of the relationship between stock price and the news articles was done using supervised machine learning and text mining techniques. Polarity was assigned to an article based on the stop words mentioned in a dictionary. They compared three classification models: Random Forest which gave the highest accuracy ranging from 88% to 92%, SVM which worked well giving an accuracy of 86% and Naive Bayes having 83% accuracy. However, the dictionary created consisted of a limited financial terms which resulted in poor performance by the models. Unlike Joshi, Guangyu Ding and Liangxi Qin [4] proposed a solution to predict multiple value outputs using deep recurrent neural network having multiple inputs and outputs using a long short term memory (LSTM) network. The backtesting was done using multiple data sets and then compared with the two models individually. Since, the network gave multiple outputs, the model could predict open price, lowest price and highest price of any stock at the same time thereby reaching an accuracy of over 95%. However, this model did not take into account the relationship between the user and their loss handling capability which made the system less precise.

The paper [5] aimed to analyze the risk and return values of the banking by considering Nifty Index and compared fifty stocks. They found the relationship between returns and volatility with Standard Deviation. As a result, it was observed that all the trending and top stocks had a positive beta value based on the market index and the less volatile stocks had a low beta value. However, the system did not successfully prove to be accurate because it did not analyse the market continuously to select the top stocks.

In [6], trading decision methods were proposed based on a multiple classifier system and candlestick patterns. The trading strategies were evaluated using Bollinger Bands and Parabolic SAR indicators. The performance of the proposed decision strategies was tested by developing a web application which used the end of day stock prices, and the stocks were used to represent the up and down trends. The best strategy for up trend stocks gave around 17% profit and around 1% for sideway market trend. Similarly, in down trend stocks, the loss was minimized to 2.62%. The paper [7] compared the use of three famous trading strategies: SMA, MACD and PIVOT. The strategies were backtested using the d Backtest PS method. The systems ran for 1.5 years and closing price of the data was considered. They concluded that MACD>PIVOT>SMA in terms of the profits generated by each.

Due to the unpredictable events that completely change the trend of the market, most algorithmic traders opt for High Frequency Trading (HFT). In this, the brokers or the traders execute a large number of trades in a short period of time to extract profits from low changes in prices. The paper [8] proposed the use of Grid Trading Algorithms on FOREX markets to execute such trades. The system placed orders in regular time intervals to generate steady profits from these deals. Regression network and trend classifier were used to predict which buying and selling of a particular stock. The paper achieved a good Return on Investment(ROI) of 13.76%.

L. Chen and Q. Gao [9] used Deep Q-network (DQN) and Deep Recurrent Q-network (DRQN) for decision making in stock market. The two methods were compared and it was observed that DQN easily learned profitable patterns from the historical stock market data. However, introduction of recurrence to DQN led to an improvement in the performed in the trading actions. Akhil Raj et al. [10] proposed the use of Deep Reinforcement Learning model in an automated trading system. RCNN was used to predict sentiment from the stock market news data which was fed into the reinforcement learning neural network model. This model predicted the trend and made Buy/Sell decisions to maximize the profit. The model having an accuracy of 96%, was extremely efficient in predicting the sentiment and the trend of the market.

In [11], Prasetijo et al. proposed trading strategies using Bollinger Bands and Parabolic SAR indicators. They also built a web application to perform backtesting of the strategies. By taking into consideration the end-of-day stock price for different LQ-45 stocks, the system categorized into up, down and sideway trends. Bollinger Bands generated profit 17.06% for up-trending stocks while it gave profit of 1.19% for sideway market trend. Parabolic resulted in losses for all the trends hence, it worked the worst in comparison to Bollinger Bands. C. N. W. Tan [12] simulated a trading based on the two years NYSE stocks data using an artificial neural network based financial trading system. The system fit quite well giving an accuracy of 83% with a tolerance of 5%. The reliability of the system is poor as it did not consider several parameters affecting the neural network.

Most of the previous research on automated trading system focus on predicting stock prices. These systems perform trading by considering the whole stock market without categorizing and filtering out non-profitable stocks. Moreover, a generalized trading strategy is used for all kinds of user portfolios which result in inefficient performance due to the ignorance of the risk handling capabilities of users. Hence, the proposed system resolves this by introducing a filter stage to remove the non required stocks in carrying out trading. Additionally, a fuzzy logic controller has also been implemented to suggest users different trading strategies based on their portfolio.

III. Methodology

A. Architecture Model Description

The heart of the research involves providing a customer adaptive automated trading system to extend the scope of the stock market in India by making it friendly for everyone. The aim is to eliminate the limitations in the present automated trading systems like high computational power and time. Therefore, a filtering stage is included in the system to narrow down the scope of the stocks. User-specific trading strategies are also suggested based on their risk handling capabilities. Data from market data API and news articles on stocks enter the filter stage. News and articles' data is used by evaluating the sentiment for all the articles related to stocks. The filtration is performed by calculating the Return on Capital Employed for all the available stocks. The filtered stocks are stored in the System Bot database which consists of the users' transactions and the portfolio details as illustrated in Figure 1.

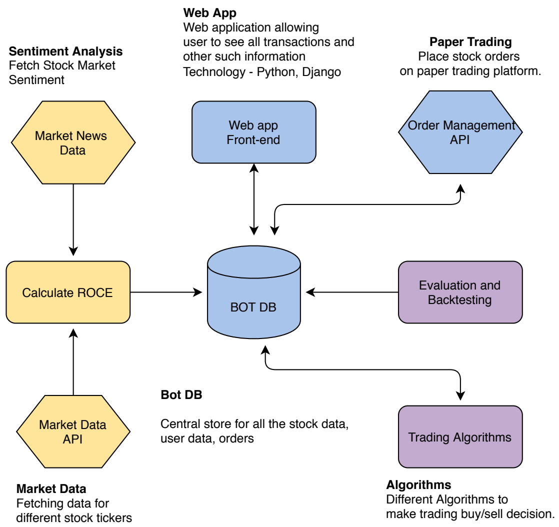


Fig. 1: System Architecture for Trading Bot

Information regarding the users' transactions and portfolio is extracted from the database and sent to the Django server. It is displayed on the web portal by the users. The system comprises of six different trading algorithms of which one is performed using Order Management API for a user based on his requirements. The evaluation and backtesting is done to assess the viability of the strategies implemented.

B. Automated Trading System

The proposed system comprises of six modules as shown in Figure 2. Each module is explained below.

*1) **Data Scraping Module**:* All the financial data affect the volatility of the stock market. The National Stock Exchange currently consists of 1837 listed companies, and manually acquiring data for all the companies is difficult. Hence, the Data Scarping Module in the system is used to collect the data. The script web-scrapes data from reliable sources to generate a database. Two types of data are collected using this script:

(i) Financial Records of the Companies
The financial records of a company give details of its stocks which help in predicting and trading stock. It consists of the quarterly bank sheet, cash flow statement, and the income statement of all the stocks.
The extraction of the financial records is performed using Beautiful Soup with Selenium and YahooFinancials. YahooFinancials is a python library that scrapes the yahoo finance website to generate financial records. There are 1466 companies available on yahoo which were gathered in batches of ten to prevent from blocking the IP.
The remaining companies were scraped from sources like moneycontrol.com, Finviz India or Screener using BeautifulSoup with Selenium. BeautifulSoup parses through the HTML page to extract IDs which consist of the required data while Selenium repeats the task over different threads to get the stock ticker list. A total of **1573** companies' financial data was collected and stored.

(ii) Latest News Data
The news articles related to the stock market play a major role in determining the trending stocks. The data obtained from news articles is required to precisely performing trading of stocks.
The live news articles are scraped from 'Finviz India' website using BeautifulSoup. The data consists of the headlines of the news articles, its date and time. This is then stored in the form of data frames using the python library 'Pandas'.

*2) **Data Filtering Module**:* The large size of the data results in increasing computational time and power of the system. Because of this, the system becomes less efficient. Hence, the proposed system uses a unique filter - Return on Capital Employed (RoCE). This parameter is a ratio which indicates how profitable a company is and how effective its investments are. The ratio is given by:

$$\text{RoCE} = \frac{\text{Earning Before Interest and Tax (EBIT)}}{\text{Total Assets - Current Liabilities}} \tag{1}$$

Based on the latest research, companies having RoCE values greater than 20 are considered profitable. The high value of RoCE means high profitability which in turn can be considered a good investment. A company's RoCE can be calculated from its financial records using Equation 1. From the financial records of companies extracted in the Data Scraping Module, a list of thirty companies is generated after passing through the Data Filtering Module.

Also, the module removes unnecessary data from real-time market data. The redundant values and abnormalities are cleaned from the data followed by preprocessing. It is done by converting the data to DateTime format and normalizing the stock prices using MinMaxScaler from Scikit-Learn preprocessing library. The normalized values of the portfolio are required in the LSTM neural net module to predict the user's portfolio. The feature_range is set between 0 and 1. These scaled values improve the performance of the prediction module hence providing better accuracy in forecasting.

*3) **Natural Language Processing Module**:* The stock market's volatile nature is due to the fluctuations in demand and supply. The relation between demand and supply is extremely sensitive to the news articles at that instant. The analysis of the sentiments of news articles gives information on trending stocks that help in a profitable trade. Hence, a Natural Language Processing module is implemented that assigns a sentimental value to the news articles extracted in the Data Scraping module and gives a list of the trending stocks.

The news articles' data obtained is cleaned initially to remove the stop words which do not play any role in assignment of sentiment to the article. Natural Learning ToolKit Vader is used to express sentiment polarity and intensity to the news articles. Once the data is cleaned, the lexicon approach and rules come into picture to scan the data to find lexicons. It then finds the sentiment features and calculates the news article's compound score which ranges from -1 to 1. Higher the value of the score, more positive is the sentiment of the article and vice versa.

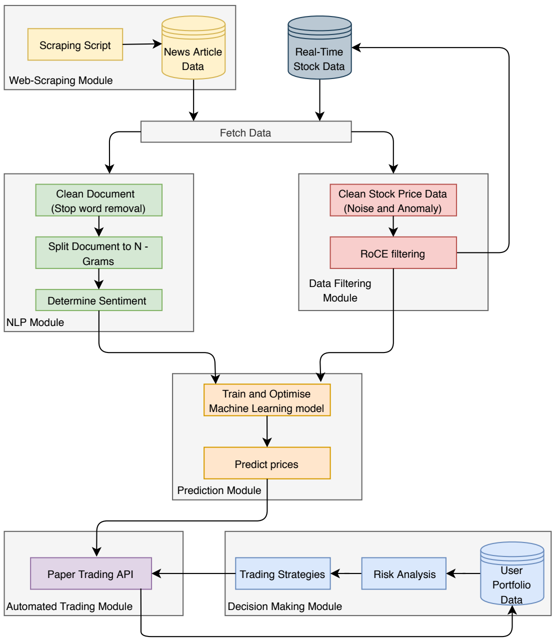


Fig. 2: Trading System with Different Modules

*4) **Prediction Module**:* The proposed system allows users to view the predictive performance of stocks based on their trading strategies. The prediction is performed using the LSTM (Long-Short Memory Model) neural net models.

The model is trained on the data of the user's portfolio based on a particular trading strategy. To attain accuracy, the model used the data obtained from backtesting the trading strategies. It consists of the trading activities for 676 days from 1st January, 2018 to 1st October, 2020 for training and the last 60 days for testing. The model used the ReLU activation function with Adam optimizer to optimize the output and ran for 150 epochs. Finally, the results of the graphs are viewed on the dashboard of the website by the users.

*5) **Decision Making Module**:* The decision making module is the most important module of the proposed system. It is responsible for determining user specific strategies and making buy and sell decisions. The module is divided into three parts: User Portfolio Data which acts as a storage unit for user data, Trading Strategies which help to execute the trading decisions and Risk Analysis which determines which strategy is suitable for a particular user.

User Portfolio Data: The user portfolio is a central database that stores data such as the information about all the stocks held by users, transactions made by the system, risk analysis and evaluation of the trading strategies. Every financial day, the database is updated with the current trending stock tickers from the Natural Language Processing module. These tickers are used with the user's trading strategies for analysis and to carry out buy/sell decisions.

Trading Strategies: Algorithmic trading strategies use a set of predefined rules and indicators to analyze the market in order to perform trades. The strategies use different parameters like price variations, stock volume, market trend, historical data, etc. for making effective trading decisions. There are six trading strategies which are stored in the system.

(i) **Simple Moving Average (SMA) Strategy**
Simple Moving Average is calculated by adding the closing prices and dividing the sum by the number of time periods in the average. This helps in analyzing a bull or bear trend.

(ii) **Exponential Moving Average (EMA) Strategy**
Exponential Moving Average analyzes the predominance of a trend in the market. It is a line on a price chart that smoothens the price action by using the mathematical formulae. It puts emphasis on the recent price which makes it more reliable as it reacts faster to the latest changes in the price data.

$$\text{Initial SMA} = 20 - \text{Period Sum}/20 \tag{2}$$

$$\text{Multiplier } (k) = \frac{2}{\text{Time period+1}} \tag{3}$$

$$\text{EMA} = \text{Close Price} * k + \text{EMA}_\text{n} * (1 - k) \tag{4}$$

where,
n = Previous day

(iii) **Double Exponential Moving Average (DEMA) Strategy**
The Double Exponential Moving Average reduces the amount of lag present in the traditional moving averages. A change of trend is observed when the prices crosses the average. Hence, it confirms the uptrends when the

price moves above average and downtrends when the price moves below average.

$$\mathrm{DEMA} = (2 * \mathrm{EMA_n}) - (\mathrm{EMA\ of\ EMA_n}) \quad (5)$$

where,
n = Look-back period

(iv) **Triple Exponential Moving Average (TEMA) Strategy** Triple Exponential Moving Average reduces the lag time further as compared to DEMA. It has a complex formula which involves calculating EMA of an EMA of an EMA. Like DEMA, it also helps to find the trend changes by comparing with the average.

$$\mathrm{TEMA} = (3 * \mathrm{EMA_1}) - (3 * \mathrm{EMA_2}) + \mathrm{EMA_3} \quad (6)$$

$$\mathrm{EMA_2} = \mathrm{EMA\ of\ EMA_1} \quad (7)$$

$$\mathrm{EMA_3} = \mathrm{EMA\ of\ EMA_2} \quad (8)$$

(v) **Bollinger Bands Strategy** Bollinger Bands are trendlines which are plotted at two stand deviation levels above and below a simple moving average (SMA) of a price. A middle band which is the simple moving average, an upper band and a lower band are the three lines of the Bollinger Bands. These can be modified based on the user preferences. The main parameters are period and the standard deviation which help to predict the prices on a relative basis.

$$\mathrm{BOLU} = \mathrm{MA(TP,n)} + m * \delta[\mathrm{TP,n}] \quad (9)$$

$$\mathrm{BOLD} = \mathrm{MA(TP,n)} - m * \delta[\mathrm{TP,n}] \quad (10)$$

where,
BOLU = Upper Bollinger Band
BOLD = Lower Bollinger Band
MA = Moving average
TP (typical price) = (High+Low+Close)÷3
n = Number of days in smoothing period
m = Number of standard deviations
δ[TP,n] = Standard Deviation over last n periods of TP

(vi) **Commodity Channel Index (CCI) Strategy** CCI calculates the difference between the historical average and the current price. It also helps to determine the overbought and oversold levels for an asset. If the CCI is above zero then the price is higher than the average and if it is below zero then the price is lower than the historical average.

$$\mathrm{CCI} = \frac{\mathrm{Last\ price} - \mathrm{SMA\ of\ last\ n\ prices}}{\mathrm{Impact\ factor} * \delta} \quad (11)$$

$$\delta = \frac{\sum_{i=k}^{k+n} \mathrm{Price}(i) - \mathrm{SMA\ of\ last\ n\ prices}}{\mathrm{n}} \quad (12)$$

where,
δ = Mean Deviation
n = Period
k >= 0

Risk Analysis: The analysis of the risk associated with the trading decision that is made by the system is calculated on the stored historical data. Every user has a specific trading strategy which is determined by using Fuzzy Logic Controller. The user enters his risk handling capabilities and expected returns. The inputs for fuzzification are the Sortino ratio, Sharpe ratio and the Return on investment. The rule base derives the estimated risk and return range which maps to each of the trading strategies. By taking into consideration the users capabilities, a specific trading strategy is assigned to the user. On the assigned trading strategy, analysis is performed for a period of two years.

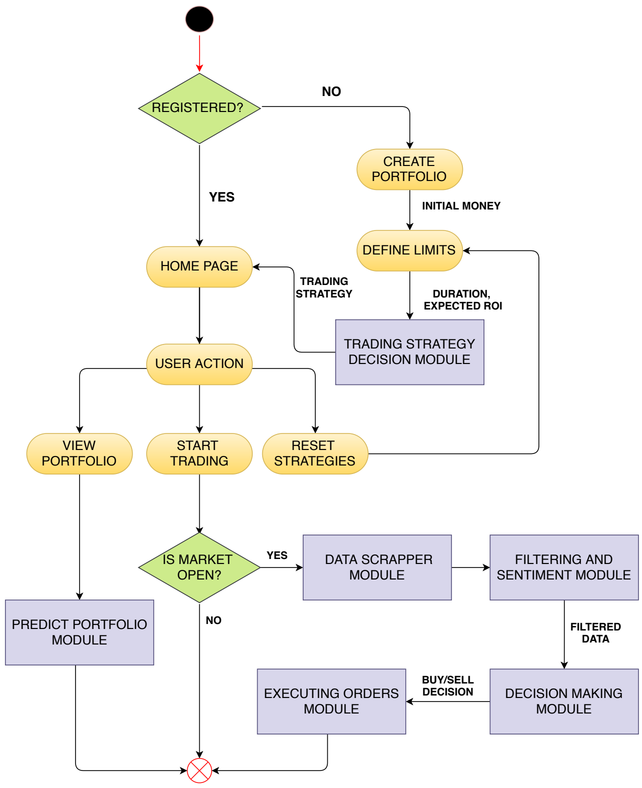


Fig. 3: User Interaction Diagram

*6) **Trade Execution Module:*** Trade Execution module is the final module where actual trading activities take place. Once the trading strategy is assigned to the user and the trading decision is made by the Decision-Making module, the stock orders are placed in the Trade Execution module. For this purpose, a paper trading platform is used which is a simulated trade that allows people to perform buying and selling actions without the involvement of real money.

C. System Operation

The fundamental purpose of the proposed system is to make stock market trading easy for users. The interaction of the user with the system is explained in Figure 3. It is an automated system which only requires the users to set the limits for performing trading actions in order to assign a trading strategy based on their requirement. It allows them to view their portfolio, start trading or reset their input to change the strategy. By viewing the portfolio, a user can either check their current worth based on their transactions or ask the system to run a prediction to check their future portfolio. Prediction module helps the user to get an estimated profit made using the current trading strategy. By starting the trading, the user gives a signal to the system to start the execution of trades and by resetting the input, users can redefine the limits and parameters.

IV. Results and Discussion

This research aims to fulfill the objectives mentioned earlier. To achieve those objectives, a system consisting of different modules is put together and following results with the respect to the mentioned objectives were achieved.

A. Sentiment Analysis on News Articles

News articles play an important role in predicting the stock prices and performing buy and sell actions in trading. By using the Vader analyzer, the compound scores of all tickers were calculated. The compound scores for three Indian stocks: HDFC Bank, Infosys and Tata Motors were compared based on their news articles for a period of four days.

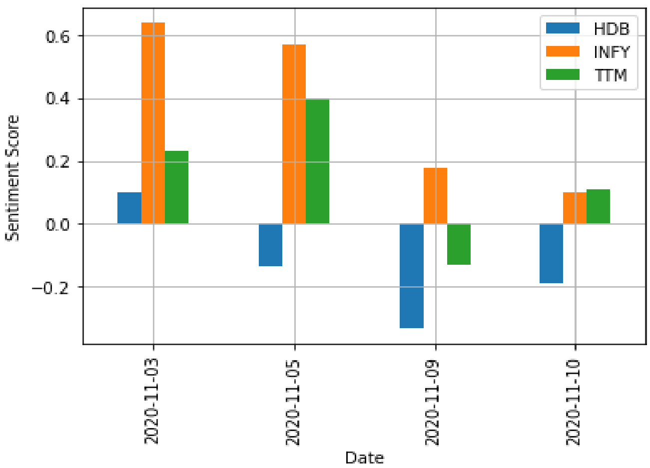


Fig. 4: Comparison of Tata Motors, HDFC Bank and Infosys Stocks

It is observed from Figure 4 that Infosys had a very high score on November 3, 2020, this means that there were several positive news articles about that company. Hence, it can be inferred that the stock price of INFY would be high based on its demand. Similarly, it can also be viewed that Tata Motors and HDFC Bank has negative score on November 9, 2020. This shows that there were negative sentimental articles on those companies which contributed to the low prices of their stocks.

B. Prediction of stock prices

Future values of the user portfolio was calculated using LSTM neural net model. Figure 5 represents the user's portfolio over time. Blue line indicates the training data for the model. Yellow and orange lines indicate the training and validation over the period of 2 months. As seen in Figure 6, it can be clearly inferred that the model is accurate in predicting results in the future and has a train loss of 0.001 and validation loss of 0.02. The model predicted that the user's portfolio will increase by 12% over the next week.

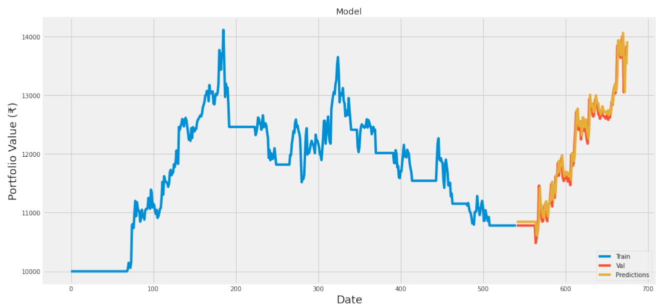


Fig. 5: User Portfolio Prediction

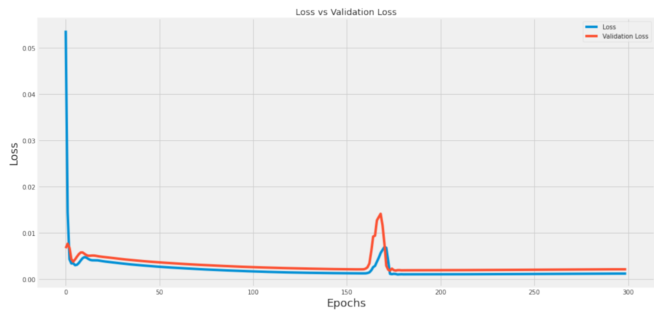


Fig. 6: Loss and Validation Loss

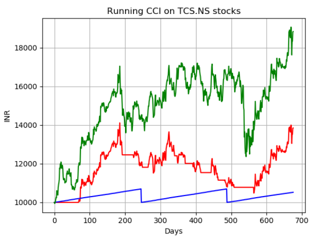
C. User specific Trading Strategies

One of the primary objectives was to implement different trading strategies and find the best fit according to the risk associated with those strategies. Bollinger bands, CCI, Moving momentum and various moving averages based Trading Strategies were implemented in the system. Backtesting was done on these trading strategies by initialising the amount at hand as Rs 10,000 and for a period of 2 years for stock TCS. Figure 7 compares backtesting of various trading strategies. This comparison shows how the strategies will react to the change in price. Green line indicates the company's stock price, red line indicates the user's portfolio and the blue line is the risk free return which is set to 7% as the standard market rate.

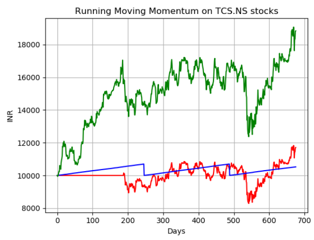
TABLE I: PROFIT GENERATED AFTER BACKTESTING

Strategy Name	Average Profit	Total Profit	Backtesting Period
CCI	5.5	3851	700
Momentum	2.4	1710	700
SMA	4.7	3350	700
EMA	7.5	5252	700
DEMA	7.9	5573	700
TEMA	8.6	6068	700

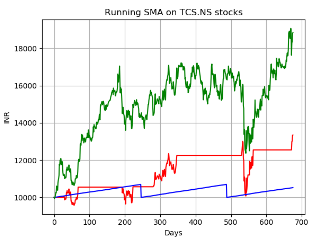
As seen in Table I, TEMA yields the best result for the period of time taken into consideration. This results might alter with change in duration and other influencing factors that affect the stock market.



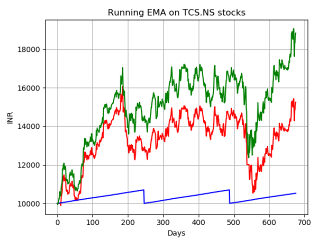
(a) CCI



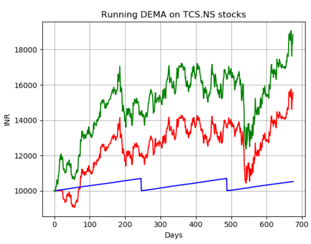
(b) Momentum



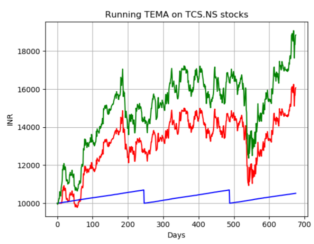
(c) SMA



(d) EMA



(e) DEMA



(f) TEMA

Fig. 7: Evaluation graphs for Trading Strategies

V. Conclusion

Stock market is an important aspect in the financial industry. A lot of research has been carried out in the past to automate the trading system. With recent advancements in the technology sector, there is an increasing impetus to optimise automated system and make it user adaptive. The proposed system uses both stock market data and news articles related to them to perform trading. From this, using RoCE filter, the stocks that are non-profitable and unnecessary are removed. The sentiment analysis of stocks is also performed using NLP to generate the trending stocks. Trending stocks are the ones whose news articles have high positive sentiment values. One of the several trading strategies is assigned to a user using a Fuzzy Logic Controller by taking into consideration the Sharpe ratio, Sortino ratio and the Return on Investment, and the user defined limits. Using the combined filtered data, the system performs trading system using the user-specific trading strategy. With the help of this, the solution can help to perform trading easily and efficiently. In addition, the system also predicts future value of users' portfolios using LSTM. The validation loss of the model was 0.02, hence, it fit well with the system. These novel features make the proposed system robust, user-friendly and one of a kind application which is available for all masses.

References

[1] P. Gagdani, “Stock markets in 2019: How mobile apps, internet growth may sharply shore up number of equity investors,” The Financial Express, 31-Dec-2018. [Online]. Available: https://www.financialexpress.com/market/stock-markets-in-2019-how-mobile-apps-internet-growth-may-sharply-shore-up-number-of-equity-investors/1429416/. [Accessed: 6-Jan-2021].

[2] Amir, Eli & Kama, Itay. (2005). The Market Reaction to ROCE and ROCE Components. SSRN Electronic Journal. 10.2139/ssrn.623161.

[3] Joshi, Kalyani & N, Bharathi & Rao, Jyothi. (2016). Stock Trend Prediction Using News Sentiment Analysis. International Journal of Computer Science and Information Technology. 8. 67-76. 10.5121/ijcsit.2016.8306.

[4] Ding, Guangyu & Qin, Liangxi. (2019). Study on the prediction of stock price based on the associated network model of LSTM. International Journal of Machine Learning and Cybernetics. 10.1007/s13042-019-01041-1.

[5] Prabhu, Ruchi. (2019). Risk & return analysis of nifty stock in Indian capital market. 10.13140/RG.2.2.26421.01767.

[6] F. Yang and J. Zhang, ”Bullish-bearish-based neural network stock trading decision supportano its application in Hong Kong stock market,” 2015 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Guangzhou, 2015, pp. 179-184, doi: 10.1109/ICWAPR.2015.7295947.

[7] D. Th. Vezeris and C. J. Schinas, ”Performance Comparison of Three Automated Trading Systems (MACD, PIVOT and SMA) by Means of the d-Backtest PS Implementation,” International Journal of Trade, Economics and Finance vol.9, no.4, pp. 170-173, 2018.

[8] Rundo, Francesco & Battiato, Sebastiano & Trenta, Francesca & Stallo, Agatino. (2019). Grid Trading System Robot (GTSbot): A Novel Mathematical Algorithm for trading FX Market. Applied Sciences. 9. 1796. 10.3390/app9091796.

[9] L. Chen and Q. Gao, ”Application of Deep Reinforcement Learning on Automated Stock Trading,” 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2019, pp. 29-33, doi: 10.1109/ ICSESS47205.2019. 9040728.

[10] Azhikodan A.R., Bhat A.G.K., Jadhav M.V. (2019) Stock Trading Bot Using Deep Reinforcement Learning. In: Saini H., Sayal R., Govardhan A., Buyya R. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 32. Springer, Singapore.

[11] A. B. Prasetijo, T. A. Saputro, I. P. Windasari and Y. E. Windarto, ”Buy/sell signal detection in stock trading with bollinger bands and parabolic SAR: With web application for proofing trading strategy,” 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, 2017, pp. 41-44, doi: 10.1109/ICITACEE.2017.8257672.

[12] C. N. W. Tan, ”Trading a NYSE-stock with a simple artificial neural network-based financial trading system,” Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, Dunedin, New Zealand, 1993, pp. 294-295, doi: 10.1109/ANNES.1993.323022.

Bibliography

[1] Amir, Eli & Kama, Itay. (2005). The Market Reaction to ROCE and ROCE Components. SSRN Electronic Journal. 10.2139/ssrn.623161.

[2] Joshi, Kalyani & N, Bharathi & Rao, Jyothi. (2016). Stock Trend Prediction Using News Sentiment Analysis. International Journal of Computer Science and Information Technology. 8. 67-76. 10.5121/ijcsit.2016.8306.

[3] Ding, Guangyu & Qin, Liangxi. (2019). Study on the prediction of stock price based on the associated network model of LSTM. International Journal of Machine Learning and Cybernetics. 10.1007/s13042-019-01041-1.

[4] Prabhu, Ruchi. (2019). Risk & return analysis of nifty stock in Indian capital market. 10.13140/RG.2.2.26421.01767.

[5] F. Yang and J. Zhang, "Bullish-bearish-based neural network stock trading decision supportano its application in Hong Kong stock market," 2015 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Guangzhou, 2015, pp. 179-184, doi: 10.1109/ICWAPR.2015.7295947.

[6] D. Th. Vezeris and C. J. Schinas, "Performance Comparison of Three Automated Trading Systems (MACD, PIVOT and SMA) by Means of the d-Backtest PS Implementation," International Journal of Trade, Economics and Finance vol.9, no.4, pp. 170-173, 2018.

[7] Rundo, Francesco & Battiato, Sebastiano & Trenta, Francesca & Stallo, Agatino. (2019). Grid Trading System Robot (GTSbot): A Novel Mathematical Algorithm for trading FX Market. Applied Sciences. 9. 1796. 10.3390/app9091796.

[8] L. Chen and Q. Gao, "Application of Deep Reinforcement Learning on Automated Stock Trading," 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2019, pp. 29-33, doi: 10.1109/ ICSESS47205.2019. 9040728.

[9] Azhikodan A.R., Bhat A.G.K., Jadhav M.V. (2019) Stock Trading Bot Using Deep Reinforcement Learning. In: Saini H., Sayal R., Govardhan A., Buyya R. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 32. Springer, Singapore.

[10] A. B. Prasetijo, T. A. Saputro, I. P. Windasari and Y. E. Windarto, "Buy/sell signal detection in stock trading with bollinger bands and parabolic SAR: With web application for proofing trading strategy," 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, 2017, pp. 41-44, doi: 10.1109/ICITACEE.2017.8257672

[11] C. N. W. Tan, "Trading a NYSE-stock with a simple artificial neural network-based financial trading system," Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, Dunedin, New Zealand, 1993, pp. 294-295, doi: 10.1109/ANNES.1993.323022.

Curiginal

Document Information

Analyzed document	Group4_Research_Paper.pdf (D111209428)
Submitted	8/12/2021 2:54:00 PM
Submitted by	
Submitter email	arya.kothari@spit.ac.in
Similarity	6%
Analysis address	lukesh.kadu.sakec@analysis.urkund.com

Sources included in the report

W	URL: https://www.semanticscholar.org/paper/Bullish-bearish-based-neural-network-stock-trading-Yang-Zhang/acdeb406bcbc3d3961b61b80658541d5a948edb1 Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://www.investopedia.com/ask/answers/122314/what-exponential-moving-average-ema-formula-and-how-ema-calculated.asp Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://www.investopedia.com/trading/using-bollinger-bands-to-gauge-trends/ Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://www.financialexpress.com/market/stock-markets-in-2019-how-mobile-apps-internet-growth-may-sharply-shore-up-number-of-equity-investors/1429416/ Fetched: 8/12/2021 2:55:00 PM	2
W	URL: https://www.semanticscholar.org/paper/Trading-strategy-design-in-financial-investment-a-Li-Deng/a885522734c4ecc26a929e5673f80c29dc083436 Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://thekoala.finance/papers/TheKoalaFinanceWhitepaperV1.pdf Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1724&context=etd_projects Fetched: 6/22/2020 7:14:00 PM	1
W	URL: https://www.ijeat.org/wp-content/uploads/papers/v10i2/B20341210220.pdf Fetched: 1/15/2021 8:37:24 PM	1
W	URL: https://www.researchgate.net/publication/322515519_Buysell_signal_detection_in_stock_trading_with_bollinger_bands_and_parabolic_SAR_With_web_application_for_proofing_trading_strategy Fetched: 5/7/2021 9:48:02 AM	1
W	URL: https://ieeexplore.ieee.org/document/323022 Fetched: 8/12/2021 2:55:00 PM	1

Curiginal

Document Information

Analyzed document	Group4_Black_Book.pdf (D111209427)
Submitted	8/12/2021 2:54:00 PM
Submitted by	
Submitter email	arya.kothari@spit.ac.in
Similarity	4%
Analysis address	lukesh.kadu.sakec@analysis.urkund.com

Sources included in the report

W	URL: https://www.semanticscholar.org/paper/Bullish-bearish-based-neural-network-stock-trading-Yang-Zhang/acdeb406bcbc3d3961b61b80658541d5a948edb1 Fetched: 8/12/2021 2:55:00 PM	4
W	URL: https://c.mql5.com/forextsd/forum/224/Automated%20Stock%20Market%20Trading%20System%20using%20Machine%20Learning.pdf Fetched: 10/10/2019 8:12:33 AM	1
W	URL: https://www.investopedia.com/ask/answers/122314/how-do-i-use-exponential-moving-average-ema-create-forex-trading-strategy.asp Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://core.ac.uk/download/pdf/322327703.pdf Fetched: 3/24/2021 11:31:44 AM	1
W	URL: https://www.semanticscholar.org/paper/Trading-strategy-design-in-financial-investment-a-Li-Deng/a885522734c4ecc26a929e5673f80c29dc083436 Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://arxiv.org/pdf/2011.09607 Fetched: 8/12/2021 2:55:00 PM	1
W	URL: https://www.ijeat.org/wp-content/uploads/papers/v10i2/B20341210220.pdf Fetched: 1/15/2021 8:37:24 PM	1
W	URL: https://www.researchgate.net/publication/322515519_Buysell_signal_detection_in_stock_trading_with_bollinger_bands_and_parabolic_SAR_With_web_application_for_proofing_trading_strategy Fetched: 8/12/2021 2:55:00 PM	1