A PROJECT REPORT ON

**STOCK MARKET PERFORMANCE ANALYSIS USING SENTIMENT ANALYSIS AND DEEP LEARNING**

SUBMITTED BY

**HATIM M. TAMBUWALA (402079)**

**HUSSAIN BOHRA (402037)**

**SHRADDHA SINGH (403019)**

**KUSHAL KARKAREY (402064)**

UNDER THE GUIDANCE OF

**Prof. S. S. BOBDE**



**Department Of Computer Engineering**

**MAEER‘s MAHARASHTRA INSTITUTE OF TECHNOLOGY**

**Kothrud, Pune 411 038**

**2017-2018**

****

**MAHARASHTRA ACADEMY OF ENGINEERING AND EDUCATIONAL RESEARCH‘S**

**MAHARASHTRA INSTITUTE OF TECHNOLOGY**

**PUNE**

**DEPARTMENT OF COMPUTER ENGINEERING**

**C E R T I F I C A T E**

This is to certify that

**HATIM M. TAMBUWALA (402064)**

**HUSSAIN BOHRA (402037)**

**SHRADDHA SINGH (403019)**

**KUSHAL KARKAREY (402064)**

Of B. E. Computer successfully completed B.E project in

**STOCK MARKET PERFORMANCE ANALYSIS USING SENTIMENT ANALYSIS AND DEEP LEARNING**

to my satisfaction and submitted the same during the academic year 2017-2018 towards the partial fulfilment of degree of Bachelor of Engineering in Computer Engineering of Savitri Bai Phule Pune University under the Department of Computer Engineering , Maharashtra Institute of Technology, Pune.

Prof. S. S. Bobde Prof. (Dr.) V. Y. Kulkarni

(Seminar Guide) (Head of Computer Engineering Department)

Place: Pune Date:

ABSTRACT

The prediction of stock price performance is a difficult and complex problem. Multivariate analytical techniques using both quantitative and qualitative variables have repeatedly been used to help form the basis of investor stock price expectations and influence investment decision making. However, the performance of multivariate analytical techniques is often less than conclusive and needs to be improved to more accurately forecast stock price performance. A neural network method has demonstrated its capability of addressing complex problems. A neural network method may be able to enhance an investor's forecasting ability. The purpose of this project is to examine the capability of a neural network method. Another part of our project is Sentiment Analysis. There are several news articles on stock market on the Internet and investors have to understand them immediately to invest in a stock market. Through this project, we propose an automatically dictionary construction approach and sentiment analysis of stock market news using the dictionary. To discuss our method we compare polarities determined by a financial expert with polarities determined with our proposed method. Hence, we confirm that the proposed method can make an appropriate dictionary and produced a reliable source of results. Results has shown that CNN can be better than RNN on catching semantics from texts and RNN is better on catching the context information and modelling complex temporal characteristics for stock market forecasting. The proposed method shows some improvement in accuracy when compared with similar previous studies.

i

ACKNOWLEDGEMENT

We take this opportunity to express my sincere appreciation for the cooperation given by Prof. (Dr.) V. Y. Kulkarni, HOD (Department of Computer Engineering) and need a special mention for all the motivation and support.

We are deeply indebted to our guide Prof. S. S. Bobde, for completion of this project for which she has guided and helped us going out of the way.

For all eﬀorts behind the Project report, we would also like to express our sincere appreciation to the staﬀ of the department of Computer Engineering, Maharashtra Institute of Technology Pune, for their extended help and suggestions at every stage.

HATIM M. TAMBUWALA (Exam Seat No. : B120024246)

HUSSAIN ALI BOHRA (Exam Seat No. : B120024247)

SHRADDHA A. SINGH (Exam Seat No. : B120024330)

KUSHAL K. KARKAREY (Exam Seat No. : B120024257)

ii

Contents

1 PROBLEM STATEMENT 1

2 PROBLEM DEFINITION 2

3 SCOPE OF THE PROBLEM STATEMENT 3

4 LITERATURE SURVEY 4

4.1 Predicting Stock Price Performance: A Neural Network Approach ….. 4

4.2 Stock Market’s Price Movement Prediction with LSTM Neural

Networks……………………………………………………………….. 5

4.3 Deep learning for stock market prediction from financial news

Articles………………………………………………………………..... 5

4.4 Sentiment-Aware Stock Market Prediction: A Deep Learning

Method………………………………………………………………….. 6

4.5 Sentiment Analysis of Stock Market News with Semi-supervised

Learning………………………………………………………………… 6

4.6 Stock market prediction using hybrid approach……………………….. 7

4.7 Trading on Twitter: Financial Information Content of Emotion in

Social media……………………………………………………………. 7

4.8 Sentiment Analysis to Predict Bombay Stock Exchange Using

Artificial Neural Network………………………………………………. 8

5 HARDWARE AND SOFTWARE REQUIREMENTS 9

5.1 Hardware Requirements………………………………………………... 9

5.2 Software Requirements……………………………………………….... 9

iii

6 FEASIBILITY STUDY 10

6.1 Technological Feasibility………………………………………….……10

6.2 Economic Feasibility…………………………………………….…….. 10

6.3 Time Feasibility………………………………………………………... 10

6.4 Satisfiable Feasibility…………………………………………………... 11

7 DESIGN 12

7.1 Use-Case Diagram……………………………………………………....12

7.2 Class Diagram…………………………………………………………...13

7.3 Activity Diagram………………………………………………...……...14

7.4 Sequence Diagram………………………………………………………16

7.5 Deployment Diagram……………………………………………………17

7.6 Architecture……………………………………………………………. 18

7.6.1 User-Interface (UI) Layer………………………………...…… 18

7.6.2 Data Pre-Processing Layer……………………………………..19

7.6.3 Intelligence Layer…………………………………………....... 19

7.6.4 Data Source Layer………………………………………………20

7.6.5 Low-Level Design………………………………………………21

8 TIME ANALYSIS 24

9 CONCLUSION 25

10 REFERENCES 26

iv

List of Figures

7.1 Use-Case Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

7.2 Class Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

7.3 Activity Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

7.4 Sequence Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

7.5 Deployment Diagram . . . . . . . . . . . . . . . . . . . . . . . . . 17

7.6 Architecture . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

7.7 Keyword Extraction Module . . . . . . . . . . . . . . . . . . . 21

7.8 Sentiment Analysis Module. . . . . . . . . . . . . . . . . . . . 22

7.9 Deep Learning Module . . . . . . . . . . . . . . . . . . . . . . . 23

8.1 Timeline and Deadlines . . . . . . . . . . . . . . . . . . . . . . . 24

v

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 1

PROBLEM STATEMENT

The aim of this project is to create a prototype that can take in data of the events of the stock market through news feeds and previous years’ stock price data. Then using the concepts of Sentiment Analysis and Deep Learning calculate the stock market trend and suggest the end user the most profitable stock market trade options. The dataset consists of stock market opening and closing rates, highest and lowest rates, percentage loss or gain of daily trades and volume of stocks available in the stock market and live feeds of stock market news from a reliable website (www.moneycontrol.com).

MIT, Pune . 1 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 2

PROBLEM DEFINITION

The prediction of stock price performance is a difficult and complex problem. Multivariate analytical techniques using both quantitative and qualitative variables have repeatedly been used to help form the basis of investor stock price expectations and influence investment decision making. However, the performance of multivariate analytical techniques is often less than conclusive and needs to be improved to more accurately forecast stock price performance. A neural network method has demonstrated its capability of addressing complex problems. A neural network method may be able to enhance an investor's forecasting ability. This work uses deep learning methods for intraday directional movements’ prediction of BSE Sensex’s 31 company’s index using financial news titles and a set of technical indicators as input. Deep learning methods can detect and analyze complex patterns and interactions in the data automatically allowing speed up the trading process. We intend to focus on architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have had good results in traditional NLP tasks. Results has shown that CNN can be better than RNN on catching semantic from texts and RNN is better on catching the context information and modelling complex temporal characteristics for stock market forecasting. The proposed method shows some improvement when compared with similar previous studies.

Another part of our project is Sentiment Analysis. There are many news on stock market on the Internet and investors have to understand them immediately to invest in a stock market. In this project we determine sentimental polarities of the stock market news using a polarity dictionary, which consists of terms and their polarities. To achieve our aim we have to construct the polarity dictionary automatically because of decrease of human efforts. In construction the dictionary we use a semi-supervised learning approach. In the semi-supervised approach at first we make a small polarity dictionary, which a word polarity is determined manually, and using many stock market news, which polarities are not known, new words are added in the polarity dictionary. Through this project, we propose an automatically dictionary construction approach and sentiment analysis of stock market news using the dictionary. To discuss our proposed method we compare polarities determined by a financial expert with polarities determined with our proposed method. Hence, we confirm that the proposed method can make an appropriate dictionary.

MIT, Pune . 2 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 3

SCOPE OF THE PROBLEM STATEMENT

The scope of this project ranges from first time stock exchange users to veteran stock market wiz that can use this application software. STOCK MARKET PERFORMANCE ANALYSIS is also helpful to non-native people who do not have any idea about the past performance of various stock trades and how to invest their money in the national stock exchanges.

Traders and normal users can use this application to track and record their investments and stock options. Multi-National Corporations can use it to monitor their stock portfolios. This software can be deployed as an android mobile application or as a web service. Additionally, it can be integrated in various wearable devices such as Smart-Watches.

MIT, Pune . 3 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 4

LITERATURE SURVEY

Information and data has been gathered from the following publications:

4.1 Predicting Stock Price Performance: A Neural Network Approach [1]

By: Youngohc Yoon, G. Swales in the year 1991

Significance: It describes the different parameters and fundamentals of stock market performance prediction using neural networks. The purpose of this paper is to examine the capability of a neural network method and compares its predictive power with that of multiple discriminant analysis methods.

Advantages:

1) All input parameters to the prediction model are named and explained

2) The complete application process of the Neural Network and comparison with an MDA is explained.

3) Ultimately the Neural Network approach is more efficient.

Disadvantages:

1) It is difficult to explain the characteristics of each group and the significance each parameter.

2) Hidden units are employed that improve the efficiency but cannot be used to find the relation between input and output parameters.

MIT, Pune . 4 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

4.2 Stock Market’s Price Movement Prediction with LSTM Neural Networks [2]

By: David Nelson, Adriano Pereira, Renato Oliveira in the year 2017

Significance: This article studies the usage of LSTM networks on that scenario, to predict future trends of stock prices based on the price history, alongside with technical analysis indicators. For that goal, a prediction model was built, and a series of experiments were executed and theirs results analysed against a number of metrics to assess if this type of algorithm presents and improvements when compared to other Machine Learning methods and investment strategies.

Advantages:

1. The results that were obtained are promising, getting up to an average of 55.9% of accuracy when predicting if the price of a particular stock is going to go up or not in the near future.
2. High return ratio per operation, meaning that it had more success on detecting high variations.

Disadvantages:

When it comes to the financial results it’s important to note that it was able to keep it positive for all stocks, even though it didn’t necessarily had the best results when compared to the baselines.

4.3 Deep learning for stock market prediction from financial news articles [3]

By: M. Vargas, B. Lima, A. Evsukoff in the year 2017

Significance: This work uses deep learning methods for intraday directional movements prediction of Standard & Poor's 500 index using financial news titles and a set of technical indicators as input. Deep learning methods can detect and analyze complex patterns and interactions in the data automatically allowing speed up the trading process. This paper focus on architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have had good results in traditional NLP tasks.

Advantages:

The results yielded had an accuracy of 60%.

Disadvantages:

1. It can only be used to predict stock changes for intraday trading.
2. It can predict only the movement in direction (+ve or -ve) of the stock price.

MIT, Pune . 5 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

4.4 Sentiment-Aware Stock Market Prediction: A Deep Learning Method [4]

By: J. Li, H. Bu, J Wu in the year 2017

Significance: This study proposes a new method for stock market prediction, which adopts the Long Short-Term Memory (LSTM) neural network and incorporates investor sentiment and market factors to improve forecasting performance. By extracting investor sentiment from forum posts using Naïve Bayes, this paper makes it possible to analyze the irrational component of stock price. This empirical study reveals evidence that helps to better understand investor sentiment and stock behaviours.

Advantages:

Precision was quite high.

Disadvantages;

Precision in prediction of stock closing price was not as accurate as expected.

* 1. Sentiment Analysis of Stock Market News with Semi-Supervised Learning [5]

By: K. Mizumoto, H. Yanaginimoto, M. Yoshioka in the year 2012

Significance: In this paper an automatic dictionary construction approach was proposed and sentiment analysis of stock market news using the dictionary was carried out. To discuss this proposed method polarities determined by a financial expert were compared with polarities determined with the proposed method. It confirms that the proposed method can make an appropriate dictionary.

Advantages:

This method adds many words in a polarity dictionary and can determine correct polarities of 45% of all news.

Disadvantages:

Experiment negation and adversative conjunction decline the accuracy.

MIT, Pune . 6 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* 1. Stock market prediction using hybrid approach [6]

By: V. Rajput, S.S. Bobde in the year 2016

Significance: The objective of this paper is to construct a model to predict stock value movement using the opinion mining and clustering method to predict National Stock Exchange (NSE). Topics and related opinion of shareholders are automatically extracted from the writings in a message board by utilizing our proposed strategy alongside isolating clusters of comparable sort of stocks from others using clustering algorithms. Proposed methodology will give us two output set i.e. one from sentiment analysis and another from clustering based prediction with respect to some specialized parameters of stock exchange. By examining both the results an efficient prediction is produced. In this paper stocks with maximum capitalization within all the important sectors are taken into consideration for empirical analysis.

Advantages:

Using hybrid approach adds to the efficiency of stock price prediction.

Disadvantages:

Sentiment analysis on social media is difficult. The text is usually fake, there are many misspellings and incorrect grammar constructions.  Sometimes sentiments can have weak or strong predictive capabilities.

4.7 Trading on Twitter: Financial Information Content of Emotion in Social media [7]

By: H.K. Sul, A. Dennis, l. Yuan in the year 2014

Significance: Data was collected from Twitter posts about firms in the S&P 500 and their cumulative emotional valence was analyzed (i.e. whether the posts contained an overall positive or negative emotional sentiment). The results show that the cumulative emotional valence (positive or negative) of Twitter tweets about a specific firm was significantly related to that firm's stock returns. The emotional valence of tweets from users with many followers (more than the median) had a stronger impact on same day returns, as emotion was quickly disseminated and incorporated into stock prices. In contrast, the emotional valence of tweets from users with few followers had a stronger impact on future stock returns (10-day returns).

Advantages:

Social media postings is associated with same day abnormal returns and also has future abnormal return predictability.

Disadvantages:

Technical parameters are not taken into consideration.

MIT, Pune . 7 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

4.8 Sentiment Analysis to Predict Bombay Stock Exchange Using Artificial Neural Network [8]

By: S. Khatri, H. Singhal, P. Johri in the year 2014

Significance: In this research work, sentiment analysis was formulated on data from social media which is classified using classification algorithm of machine learning. The classified data is analysed to calculate the net mood of the comments. These comments are classified into four classes' namely happy, hope, sad, disappointing. The net relative mood of all the classes per day is used as input for artificial neural network (ANN) to be trained for data of n days and their respective change in index value on each day. This trained network is finally used to predict the vector of Bombay Stock Exchange index value for (n+1) days.

Advantages:

1. Uses classification and machine learning.
2. Differentiates between moods of the users using social media interface.

Disadvantages:

Generation of dataset is a big task.

MIT, Pune . 8 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 5

HARDWARE AND SOFTWARE REQUIREMENTS

* 1. Hardware Requirements
  2. Memory : 2GB or More
  3. Chipset : Intel i3 or above / AMD Athlon 64 or above
  4. GPU Support : NVIDIA GEFORCE 860m or above
  5. Software Requirements
  6. Platform : Python Virtual Machine for Python 2.7 / 3, Anaconda, Android Studio
  7. Operating System : Ubuntu 14.04 LTS or above
  8. Programming Language : Python, Java, Android
  9. Dependency Packages :

- Make Utility

- Scipy

- Numpy

- Sci-Kit Imaging (skimage)

- GCC Compiler

- Caffe

- ngrams

MIT, Pune . 9 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 6

FEASIBILITY STUDY

6.1 Technological Feasibility

The components and dependencies chosen in software are widely used and have proven to be robust and reliable. The design phase also conveys the feasibility of the project. Since Case has some pre-trained CNN packages for stock market performance analysis, there is no issue in the section of compatibility of various technologies that will together in the system. Python has proved itself to be one of the most efficient, meaningful and widely supported programming languages. Hence, gathering of dependencies is not an issue in this scenario. Other peripheral technologies like package manager for python (pip), GCC compiler, Ipython Notebook (IDE), etc. are well known and have an extensive community support. Therefore, considering all the pieces of the puzzle, the feasibility of the project is very realistic and well documented.

6.2 Economic Feasibility

The project is economically feasible. Although it is very early for us to determine the exact cost of this project.

|  |  |
| --- | --- |
|  |  |

6.3 Time Feasibility

This project has been estimated to be completed in 7-8 months. The project is in the planning and documentation phase and is yet to begin with the implementation phase. Although, a pre-trained model has been implement and executed successfully for better understanding of the core technologies, we plan on to model and train our own CNN for this project.

MIT, Pune . 10 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

6.4 Satisfiable Feasibility

Our project is divided into three parts, the training/learning, sentiment analysis and the performance prediction.

* Training/ learning part is NP-complete since it can be shown with the help of the satisfiability problem that an NP-complete problem can be reduced to an instance of a learning problem for neural networks in polynomial time. Thus, no algorithm is known which could solve the learning problem in polynomial time depending on the number of unknown variables. Moreover, it is very improbable that such an algorithm exists.
* The sentiment analysis part is mere usage of the learnt words in the news articles available on the authentic trade website. It effectively decides whether the stock options will result in positive growth or negative growth. That can be done to produce the output in polynomial time. So, it is NP class problem.
* Performance prediction part is a NP-class problem as it can be done in polynomial time.

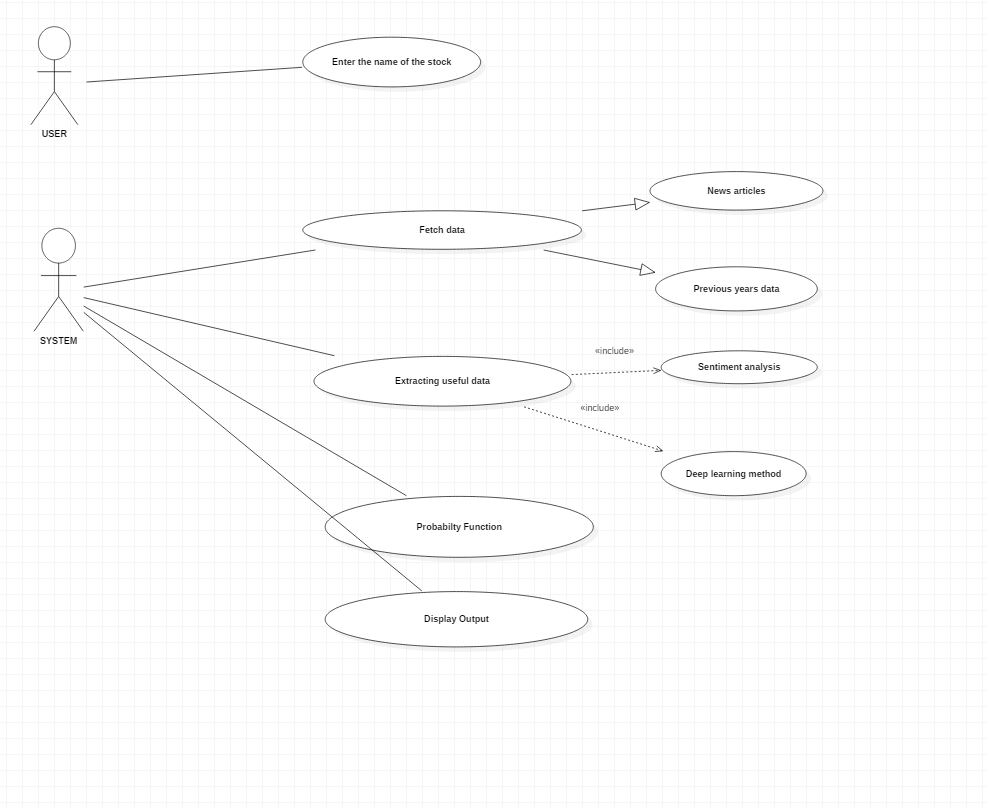
MIT, Pune . 11 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 7

DESIGN

* 1. Use-Case Diagram



**Figure 7.1:** Use-Case Diagram

MIT, Pune . 12 Dept. of Computer Engg.

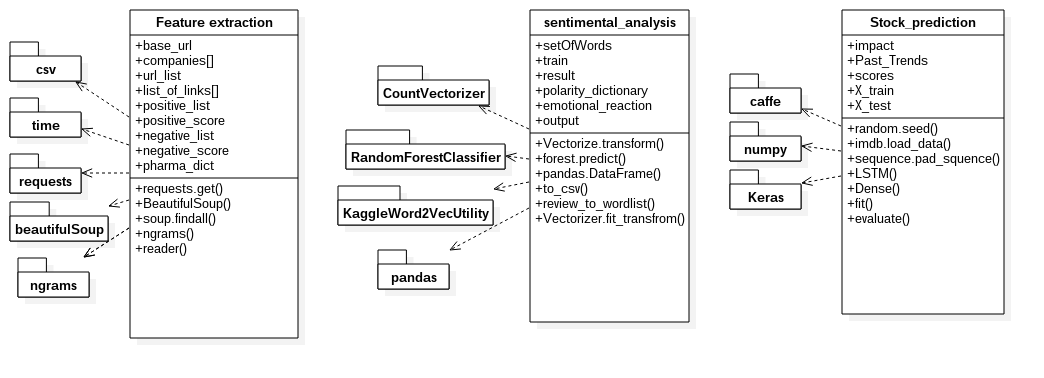
STOCK MARKET PERFORMANCE ANALYSIS

A use case diagram at its simplest is a representation of a user’s interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.

A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. While a use case itself might drill into a lot of detail about every possibility, a use-case diagram can help provide a higher-level view of the system. It has been said before that "Use case diagrams are the blueprints for your system".

* 1. Class Diagram

A class diagram in the United Modelling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the systems classes, their attributes, operations (or methods), and the relationships among objects.



**Figure 7.2:** Class Diagram

MIT, Pune . 13 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

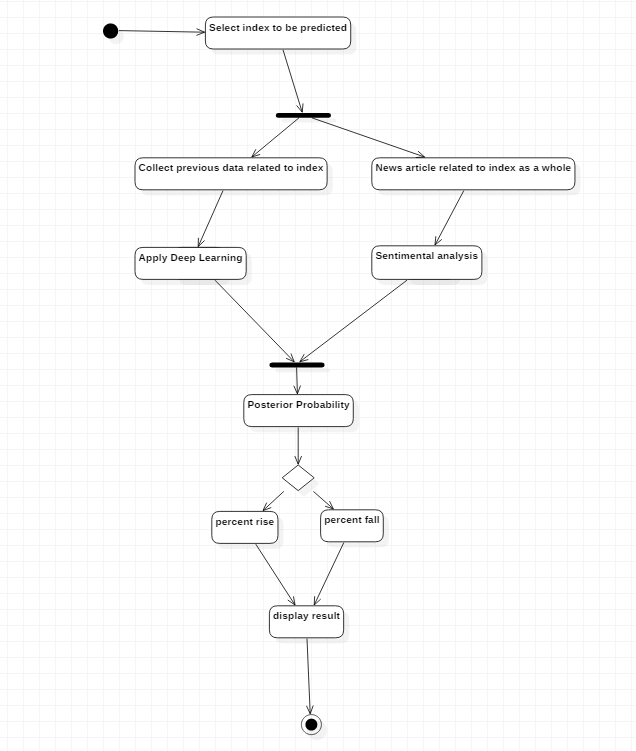
The class diagram is the main building block of object-oriented modelling. It is used both for general conceptual modelling of the systematics of the application, and for detailed modelling translating the models into programming code. Class diagrams can also be used for data modelling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.

* 1. Activity Diagram

Activity diagram is basically allow chart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. So the control ow is drawn from one operation to another. The basic purposes of activity diagrams are similar to other four diagrams. It captures the dynamic behaviour of the system. Other four diagrams are used to show the message ow from one object to another but activity diagram is used to show message ow from one activity to an- other. Activity is a particular operation of the system. Activity diagrams are not only used for visualizing dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. Activity diagram is some time considered as the flow chart. Although the diagrams looks like a flow chart but it is not. It shows different ow like parallel, branched, and concurrent and single ow systems.

MIT, Pune . 14 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

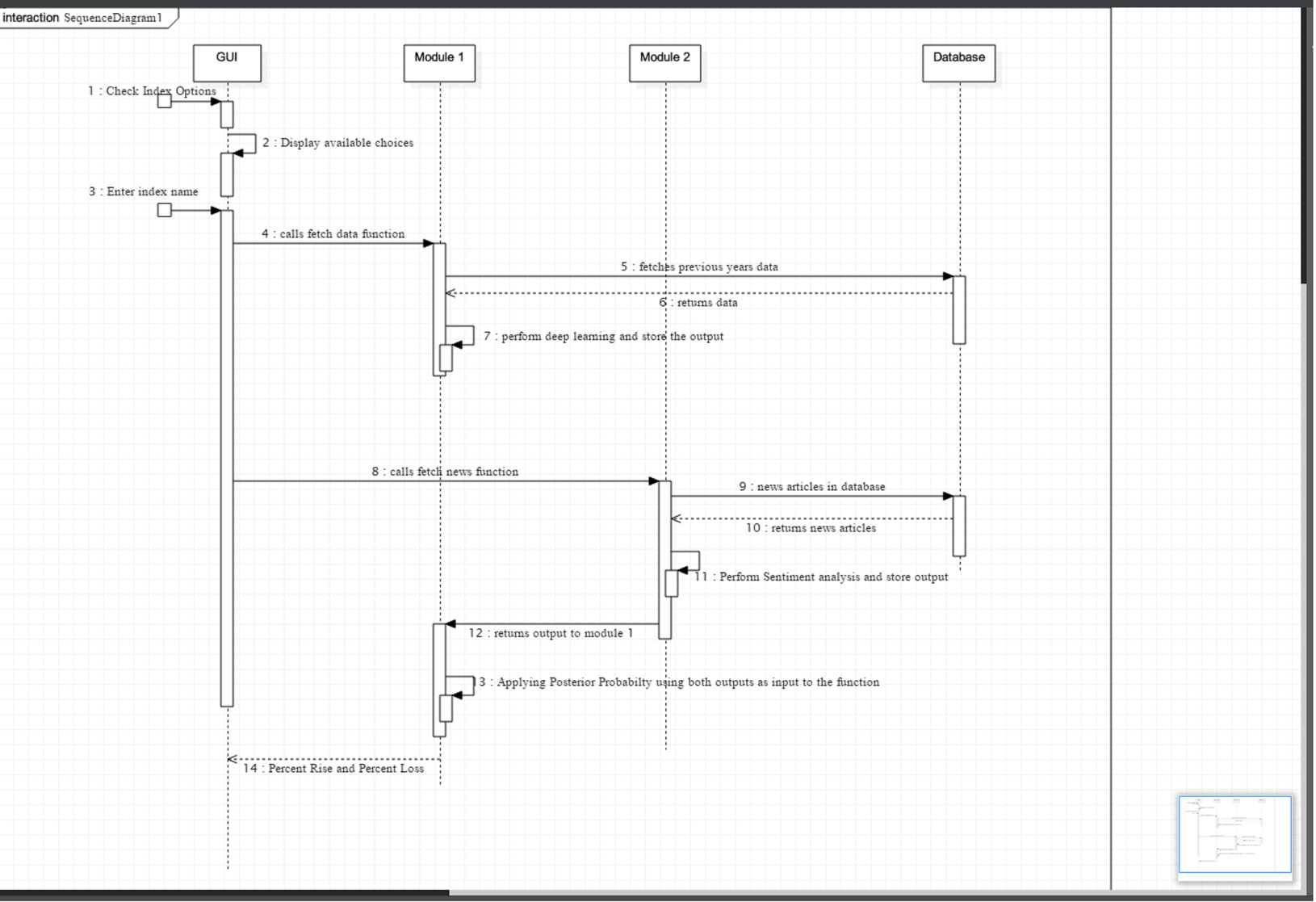
****

**Figure 7.3:** Activity Diagram

MIT, Pune . 15 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* 1. Sequence Diagram



**Figure 7.4:** Sequence Diagram

A sequence diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart.

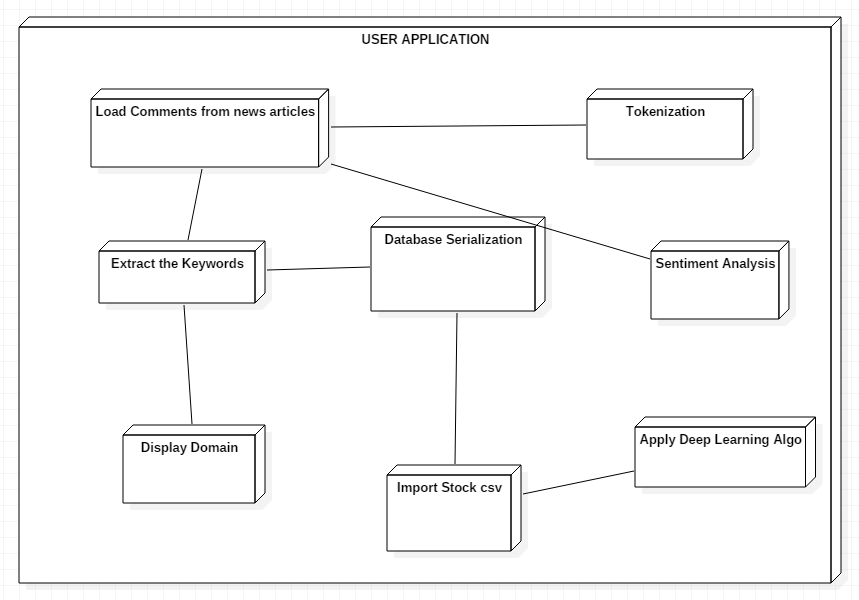
A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios

MIT, Pune . 16 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* 1. Deployment Diagram

A deployment diagram in the Unified Modelling Language models the physical deployment of artefacts on nodes. To describe a web site, for example, a deployment diagram would show what hardware components ("nodes") exist, what software components ("artefacts") run on each node and how the different pieces are connected.



**Figure 7.5:** Deployment Diagram

The nodes appear as boxes, and the artefacts allocated to each node appear as rectangles within the boxes. Nodes may have sub-nodes, which appear as nested boxes. A single node in a deployment diagram may conceptually represent multiple physical nodes, such as a cluster of database servers

MIT, Pune . 17 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* 1. Architecture

**User Interface (UI) layer**

In java and android

1-List of companies

(In .csv format)

2-Previous year’s performance on daily basis.

(Graph based)

**Intelligence layer**

**Sentimental analysis:** Analyzing the live feed from the news article and thus predicting positive or negative trend

**Deep learning:** Predicting the actual change in the stock price (using python)

**Data pre-processing layer**

Extraction of keywords from the website <http://www.moneycontrol.com> using NGram (python)

**Data source layer**

Live feed from the trading website <http://www.moneycontrol.com>

**Figure 7.6:** SYSTEM ARCHITECTURE

* + 1. User Interface (UI) Layer

The user interface (UI), in the industrial design field of human–computer interaction, is the space where interactions between humans and machines occur. The goal of this interaction is to allow effective operation and control of the machine from the human end, whilst the machine simultaneously feeds back information that aids the operators' decision-making process. Examples of this broad concept of user interfaces include the interactive aspects of computer operating systems, hand tools, heavy machinery operator controls, and process controls. The design considerations applicable when creating user interfaces are related to or involve such disciplines as ergonomics and psychology. Basically it will provide a list of companies whose data has been analysed by our project.

MIT, Pune . 18 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* + 1. Data Pre-processing Layer

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: −100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis.

If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, Instance selection, normalization, transformation, feature extraction and selection, etc. The product of data pre-processing is the final training set. Kotsiantis et al. (2006) present a well-known algorithm for each step of data pre-processing.

* + 1. Intelligence Layer
       1. Sentiment Analysis

Sentiment analysis (sometimes known as opinion mining or emotion AI) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation (see appraisal theory), affective state (that is to say, the emotional state of the author or speaker), or the intended emotional communication (that is to say, the emotional effect intended by the author or interlocutor).

MIT, Pune . 19 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* + - 1. Deep Learning

Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, partially supervised or unsupervised.

Some representations are loosely based on interpretation of information processing and communication patterns in a biological nervous system, such as neural coding that attempts to define a relationship between various stimuli and associated neuronal responses in the brain. Research attempts to create efficient systems to learn these representations from large-scale, unlabelled data sets.

Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics and drug design, where they produced results comparable to and in some cases superior to human experts.

* + 1. Data Source Layer

For a financial portal born in late 1999, just when bursting of the dotcom bubble was about to nearly bring down both financial markets and the fledgling worldwide web, we couldn't have chosen a more difficult time to launch. But it was really passion and belief that saw us through. A single-minded passion to become the country's greatest resource for financial information on the Internet. And the belief, that through it, we would be able to make a difference to people's financial lives.

Since 1999 through to today we have been there to chronicle the growth of the Indian economy. And along with it we have grown as well ; starting off as a financial portal that began by offering end-of-day stock prices to today arguably becoming India's biggest store of news (text and videos), analysis, data and tools on investing (across diverse asset classes), personal finance, the business sector and the economy.

moneycontrol.com today gets over 17 million visitors every month across all its platforms-web, mobile and tablets that makes it the largest online financial platform in India. But while we've radically changed and evolved, the belief and passion to be the best and the most insightful hasn't. That, we hope, keeps us ticking. This we believe is merely the start of our journey.

MIT, Pune . 20 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

* + 1. Low Level Design

Module 1: Keyword Extraction

News feed from the website.

**Ngrams and Beautiful Soup**

Extracted keywords

**Figure 7.7:** Keyword Extraction Module

C = { I, O, F, Sc, Fc }

I = Input = {News feed from the website http://[www.moneycontrol.com](http://www.moneycontrol.com) }

O=Output = {keywords or features are extracted from the website}

Description: The output thus obtained is useful for determining the nature of the article

F = Functions = {Word\_extracter()}

* Word\_Extracter() = {Takes text data as input using Ngrams (python library), cleans the data using Beautiful Soup (python library), and produces only important keywords as output}

Sc = Success = {proper internet connection with high speed}

Fc = Failure = {inactive or fluctuating internet connection,}

MIT, Pune . 21 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Module 2: Sentiment Analysis

Extracted keywords

Impact

**Artificial Neural Networks**

**(Self-Written)**

**Figure 7.8:** Sentiment Analysis Module

A = {I, O, F, Sc, Fc}

I = Input = {the output from the module 1 i.e. extracted keywords acts as an input to the next module}

O = Output = {positive nature or negative nature of the article}

Description: Based on keywords extracted from the article sentimental analysis is performed using artificial neural network which yields the impact of the website whether it’s positive or negative.

Functions = {Sentiment\_Analyser()}

Sentiment\_Analyser() = {It compares the collected keywords with the predefined dictionary and hence calculates the net positive or negative sentiment}

Fc = Failure = {inactive or fluctuating internet connection}

Sc = Success = {proper internet connection with high speed}

MIT, Pune . 22 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Module 3: Deep Learning

Impact and past trends

c

**Combine results of ANN and Sentiment Analysis**

Yes or no

**Figure 7.8:** Deep Learning Module

L = {I, O, F, Sc, Fc}

I = input = {output form the module 2 i.e. the sentiment analysis done from live news feeds and previous years data}

O = output = {whether the user should buy stock or not}

Funcitons = {Buy\_stock()}

Buy\_stock() = {Here the combination calculation is performed that analyses past data and sentiment analysis in order to conclude if a stock option will perform positively thereby suggesting the user to buy stocks}

Fc = Failure = {inactive or fluctuating internet connection}

Sc = Success = {proper internet connection with high speed}

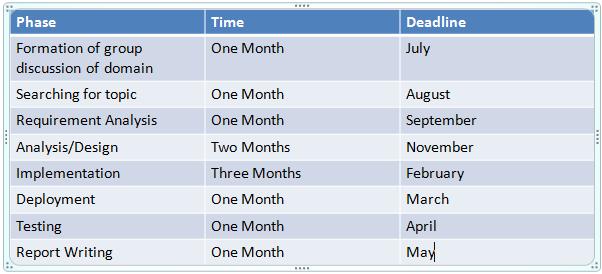
MIT, Pune . 23 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 8

Time Analysis

The figure in the following page gives a brief description of the time line of this project. These time-lines are designed by keeping in mind all the factors team members might face during the course time. Hence, following it would help in well managed work ow.



**Figure 8.1:** Timeline and Deadlines

MIT, Pune . 24 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 9

Conclusion

This report covers an extensive list of the initial requirements and analysis of the project and provides a blueprint of the functionality and productivity delivered by the system. With proper implementation of Sentiment Analysis and CNNs and sound knowledge of the working of Stock Price performance analysis using deep learning, the project believes to deliver as a trading recommendation system that would recognize help in tracking your investments smartly.

MIT, Pune . 25 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

Chapter 10

References

[1]Youngohc Yoon, G. Swales, “Predicting Stock Price Performance: A Neural Network Approach”, System Sciences, 1991. Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences.

[2] David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira, “Stock market's price movement prediction with LSTM neural networks”, 2017 International Joint Conference on Neural Networks (IJCNN).

[3] Manuel R. Vargas, Beatriz S. L. P. de Lima, Alexandre G. Evsukoff, “Deep learning for stock market prediction from financial news articles”, 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA).

[4] Jiahong Li, Hui Bu, Junjie Wu, “Sentiment-aware stock market prediction: A deep learning method”, 2017 International Conference on Service Systems and Service Management.

[5] Keisuke Mizumoto, Hidekazu Yanagimoto, Michifumi Yoshioka, “Sentiment Analysis of Stock Market News with Semi-supervised Learning”, 2012 IEEE/ACIS 11th International Conference on Computer and Information Science.

[6] Vivek Rajput, Sarika Bobde, “Stock market prediction using hybrid approach”, 2016 International Conference on Computing, Communication and Automation (ICCCA).

[7] Hong Keel Sul, Alan R. Dennis, Lingyao Ivy Yuan, “Trading on Twitter: The Financial Information Content of Emotion in Social Media”, 2014 47th Hawaii International Conference on System Sciences.

MIT, Pune . 26 Dept. of Computer Engg.

STOCK MARKET PERFORMANCE ANALYSIS

[8] Sunil Kumar Khatri, Himanshu Singhal, Prashant Johri, “Sentiment Analysis to Predict Bombay Stock Exchange Using Artificial Neural Network”, Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization.

MIT, Pune . 27 Dept. of Computer Engg.