Stock Market Prediction using Machine Learning

Dipali Dube . 1) Bhagawat Omkar Vitthal 2) Rahane Divya Ashok 3)Gunjal Sujata Nanasheb 4) shirsath Aarti Dnyandev

VIDYA NIKETAN COLLEGE OF ENGINEERING BOTA

Abstract−In this report we analyze existing and new methods of stock market prediction. We take three different approaches at the problem: Fundamental analysis, Technical Analysis, and the application of Machine Learning. We find evidence in support of the weak formof the Efficient Market Hypothesis, that the historic price does not contain useful information but out of sample data may be predictive. We show that Fundamental Analysisand Machine Learning could be used to guide an investor's decisions. We demonstratea common flaw in Technical Analysis methodology and show that it produces limited useful information. Based on our findings, algorithmic trading programs are developed and simulated using Quantopian

Keywords: Stock Prediction, Data Analysis, Natural Language Processing, Machine Learning.

♦ INTRODUCTION

Investment in Stock is one of the most rated businesses for making money for middle class investors. After that it is actual trading business of high-class investors and traders. Company's share price is most important point for investor which always fluctuates up and downwards. Eyes always need on live price of share market and instant decision making is necessary to prevent loss of money and eventually to gain money. For this you have to make study of the company's financial history and future agenda. Dependent on overall study related to market and company you can decide to invest. But you have limits to study because one cannot be sure about that study and analysis is correct. Company's market history, tendency of maintaining business in any period or slack, policies and announcements are the key points of Stock Rate. It is difficult field of work and need lot of experience to be a successful investor.

♦ MOTIVATION

We get motivated by disadvantages of existing system. The purpose of system is to build so that it will analyze by itself and express predictions.

Stock price prediction is a classic and important problem. With a fruitful model for stock forecast, we can pick up knowledge about market conduct after some time, spotting patterns that would somehow or another not have been taken note. With the inexorably computational intensity of the PC, AI will be an efficient technique to take care of this issue. Nonetheless, people in general stock dataset are unreasonably restricted for some, AI calculations to work with, while requesting more highlights may cost a huge number of dollars regular. In this paper, we will present a structure in which we coordinate client expectations into the present AI calculation utilizing open verifiable information to improve our outcomes. The spurred thought is that, in the event that we know all data about today's stock exchanging (of all specific merchants), the value is predictable. Thus, if we can obtain

just a partial information, we can hope to improve the present expectation parcel. With the development of the Internet, informal organizations, and online social associations, getting day by day client forecasts is practical job1. In this way, our inspiration is to structure an open help consolidating chronicled information and client's forecasts to make a more grounded model that will benefit everybody.

PROBLEM DEFINITION

Financial analysts investing in stock market usually are not aware of the stock market behavior. They are facing the problem of trading as they do not properly understand which stocks to buy or which stocks to sell in order to get more profits. In today's world, all the information pertaining to stock market is available. Analyzing all this information individually or manually is tremendously difficult. As such, automation of the process is required.

SOFTWARE REQUIREMENT

Purpose and Scope of Document
A software requirements specification (SRS) is a document that is created when a de- tailed description of all aspects of the software to be built must be specified before the project is to commence. It is important to note that a formal SRS is not always written. In fact, there are many instances in which effort expended on an SRS might be better spent in other software engineering activities.

Overview of responsibilities of Developer
1.To have understanding of the problem statement.
2.To know what are the hardware and software requirements of proposed system.
3.To have understanding of proposed system.
4.To do planning various activities with the help of planner.
5.Designing, programming, testing etc.
RAM : 8 GB
As we are using Machine Learning Algorithm and Various High Level Libraries Laptop RAM minimum required is 8 GB.
Hard Disk : 40 GB
Data Set of CT Scan images is to be used hence minimum 40 GB Hard Disk memory is required.
Processor : Intel i5 Processor
PyCharm and Spyder IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required

IDE : VsCode
Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that makes typing feasible and fast
Coding Language : Python Version 3.5
Highly specified Programming Language for Machine Learning because of availability of High-Performance Libraries
Operating System : Windows 10
Latest Operating System that supports all type of installation and development Environment.

♦ RELATED WORK

The most well-known publication in this area is by Bollen. They investigated whether the collective mood states of public (Happy, calm, Anxiety) derived from twitter feeds are correlated to the value of the Dow Jones Industrial Index. They used a Fuzzy neural network for their prediction. Their results show that public mood states in twitter are strongly correlated with Dow Jones Industrial Index. Chen.

and Lazer derived investment strategies by observing and classifying the twitter feeds. Bing et al. studied the tweets and concluded the predictability of stock prices based on the type of industry like Finance, IT etc. Zhang found out a high negative correlation between mood states like hope, fear and worry in tweets with the Dow Jones Average Index. Recently, Brian et al. investigated the correlation of sentiments of public with stock increase and decreases using Pearson correlation coefficient for stocks. In this paper, we took a novel approach of predicting rise and fall in stock prices based on the sentiments extracted from twitter to find the correlation. The core contribution of our work is the development of a sentiment analyzer which works better than the one in Brian's work and a novel approach to find the correlation. Sentiment analyzer is used to classify the sentiments in tweets extracted. The human annotated dataset in our work is also exhaustive. We have shown that a strong correlation exists between twitter sentiments and the next day stock prices in the results section. We did so by considering the tweets and stock opening and closing prices of Microsoft over a year.

❖ SDLC Model

SDLC Models stands for Software Development Life Cycle Models. In this article, we explore the most widely used SDLC methodologies such as Agile ... Each software development life cycle model starts with the analysis, in which the Also, here are defined the technologies used in the project, team load.
One of the basic notions of the software development process is SDLC models which stands for Software Development Life Cycle models. SDLC – is a continuous process, which starts from the moment, when it's made a decision to launch the project, and it ends at the moment of its full remove from the exploitation. There is no one single SDLC model. They are divided into main groups, each with its features and weaknesses.

❖ Mathematical Model
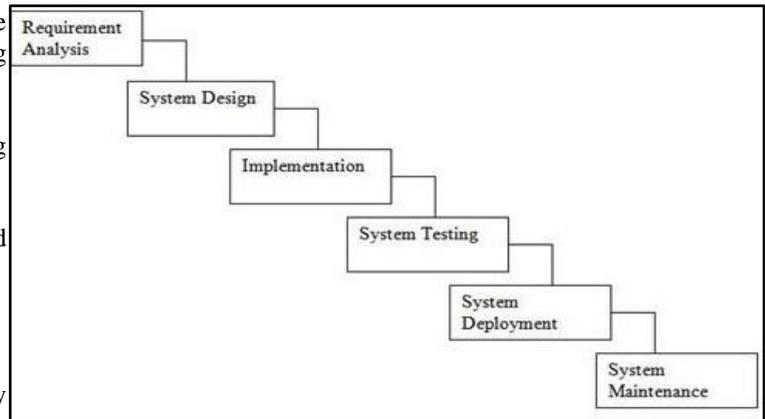
Let S is the Whole System Consist of S = I, P, O
Where,
I=CURLOC, S ELOC, LOG, RE, PRO
LOG = user login into system
SECO = Select Company
PRO = S tock Movements
R = S hare price P = Process
Step1 : user will login
Step2 : User will select Company
Step3 : User will Analyze Data
Step3 : User will Apply Mathematical Algorithmic Method
Step4 : System will give Prediction about further Stock movement
OUTPUT : predict stock market price
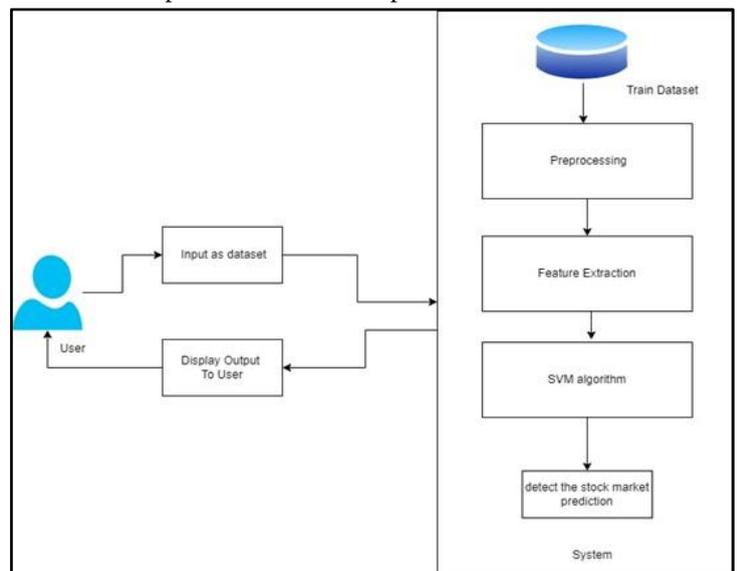
Fig - System Architecture

Overview of Project Modules

In this chapter we are going to have an overview about how much time does it took to complete each task like- Preliminary Survey Introduction and Problem Statement, Literature Survey, Project Statement, Software Requirement and Specification, System Design, Partial Report Submission, Architecture Design, Implementation, Deployment, Testing, Paper Publish,

Report Submission. This chapter also gives focus on stakeholder list which gives information about project type, customer of the proposed system, user and project member who developed the system.

Tools and Technologies Used

Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant white space. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. Python was created in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system with reference counting.

Python 3.0, released in 2008, was a major revision of the language that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3.

The Python 2 language was officially discontinued in 2020 (first planned for 2015), and "Python 2.7.18 is the last Python 2.7 release and therefore the last Python 2 release."[30] No more security patches or other improvements will be released for it. With Python 2's end-of-life, only Python 3.6.x and later are supported.

Python interpreters are available for many operating systems. A global community of programmers develops and maintains C, Python, a free and open-source reference implementation. A non-profit organization, the Python Software Foundation, manages and directs resources for Python and C, Python development.

Python was conceived in the late 1980s by Guido van Rossum at Centrum Wiskunde Informatica (CWI) in the Netherlands as a successor to the ABC language (itself in- spired by SETL), capable of exception handling and interfacing with the Amoeba operating system. Its implementation began in December 1989. Van Rossum shouldered sole responsibility for the project, as the lead developer, until 12 July 2018, when he announced his "permanent vacation" from his responsibilities as Python's Benevolent Dictator For Life, a title the Python community bestowed upon him to reflect his long- term commitment as the project's chief decision-maker. He now shares his leadership as a member of a five-person steering council. In January 2019, active Python core developers elected Brett Cannon, Nick Coghlan, Barry Warsaw, Carol Willing and Van Rossum to a five-member "Steering Council" to lead the project.

❖ Algorithm

Support Vector Machine:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

SVM algorithm can be used for Face detection, image classification, text categorization, etc.

Types of SVM

SVM can be of two types:

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non- linear data and classifier used is called as Non-linear SVM classifier.

♦ OUTPUT
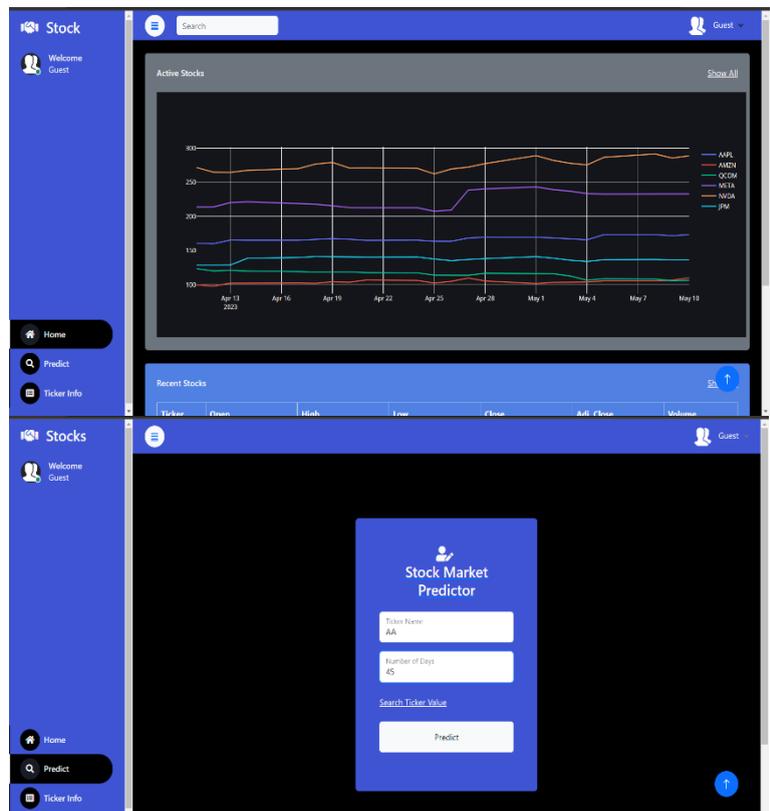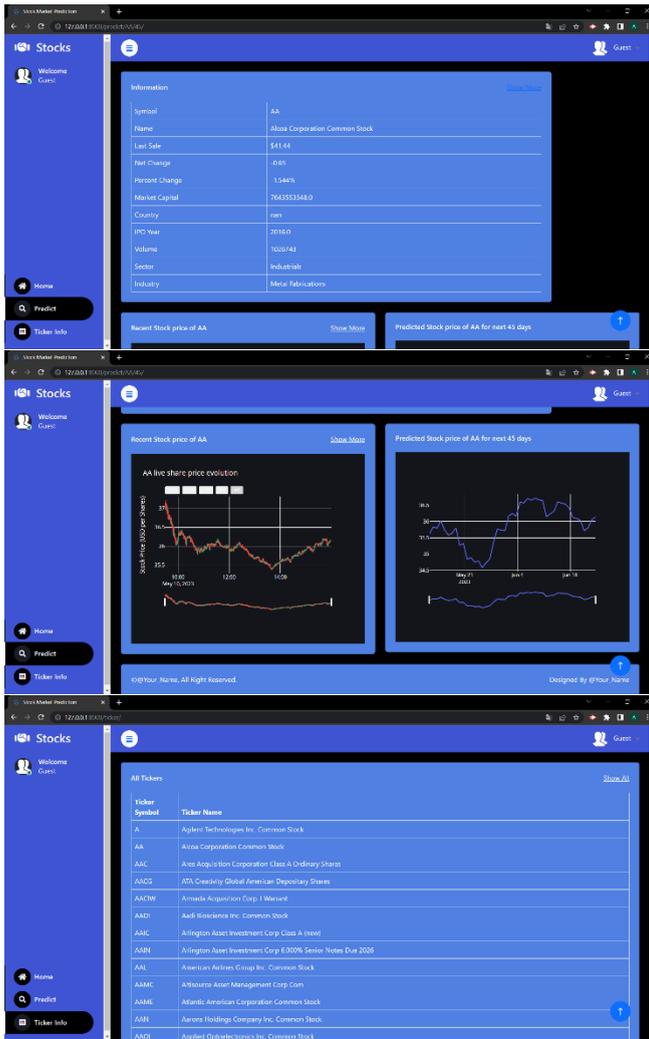
can be tracked for early prediction.

♦ **REFERENCES**

1) H. Isah, "Social Data Mining for Crime Intelligence: Contributions to Social Data Quality Assessment and Prediction Methods," University of Bradford, 2017.

2) P. Wei and N. Wang, "Wikipedia and stock return: Wikipedia usage pattern helps to predict the individual stock movement," in Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 591-594: International World Wide Web Conferences Steering Committee.

3) E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," Expert Systems with Applications, vol. 83, pp. 187- 205, 2017.

4) J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system," Expert Systems with Applications, vol. 97, pp. 60-69, 2018.

♦ **CONCLUSION**

Here , we found for stock market prediction we can use machine learning technology. A person cannot read and learn deeply about graph of any company's stock price. In practical we need to analyze data on vast level with multiple companies. Hence we can take help of Machine Learning algorithms and can have much better prediction. We can use CNN algorithm. So, we can have much more accuracy in prediction

♦ **FUTURE SCOPE**

We are here covering multiple aspects in terms of getting better and updated results from our research-based system. In future we can add more features. Multiple Country data can be tracked for international investments and multinational banking. Mutual Funds also can be tracked and can be given suggestion regarding mutual fund investments. Every activity , National Level effective Decisions and Decision makers