**SRS On**

**Stock Candle**

Submitted to SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE

OF

**BACHELOR OF ENGINEERING**

**IN**

**INFORMATION TECHNOLOGY**

**BY**

**Utkarsha M Abhang - T150318501**

**Piyush S Adhalikar - T150318503**

**Atharva M Mulay – T150318609**

**Anushka M Mulay – T150318608**

**UNDER THE GUIDANCE OF**

**PROF . YOGITA FATANGARE**



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**PES’S MODERN COLLEGE OF ENGINEERING,**

**SHIVAJINAGAR, PUNE 05**

**2021-22**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| Chapter No. | Name of Chapter | Page Number |
| 1. | Introduction to Project   * 1. Introduction to Project   2. Motivation behind project topic   3. Aim and Objective(s) of the work   4. Introduction to ARIMA   5. References | 5 |
| 2. | Overall Description  2.1 Product Perspective  2.2 User roles  2.3 User Documents  2.4 Design and implementation constraints  2.5 Assumptions and dependencies | 10 |
| 3. | System Features  3.1 Description  3.2 Response sequences  3.3 Functional Requirement | 13 |
| 4. | External Interface Requirement  4.1 User Interfaces  4.2 Software Interfaces  4.3 Hardware Interfaces | 15 |
| 5. | Non-functional Requirements  5.1 Performance Requirement  5.2 Safety Requirement  5.3 Security  5.4 Software quality attributes | 17 |
| 6. | Other requirements  6.1 Appendix A  6.2 Appendix B | 19 |
|  |  |  |

**1. INTRODUCTION**

* 1. **Introduction to Project**
* This project will provide a way to investors, stock markets to invest money in order to potentially earn a share of the company’s profits (knowing that the risk of losses exists too).
* Active investors and traders can easily buy and sell their securities due to the abundant liquidity in most major stock markets.
* Stock markets promote investment. The raising of capital allows companies to grow their businesses, expand operations and create jobs in the economy. This investment is a key driver for economic trade, growth and prosperity.
  1. **Motivation behind project topic**

•The fluctuation of stock market is violent and there are many complicated financial indicators. However, the advancement in technology, provides an opportunity to gain steady fortune from stock market

•It can help experts to find out the most informative indicators to make better prediction.

•The prediction of the market value should help investors in maximizing the profit of stock option purchase while keeping the risk low

* 1. **Aim and Objective(s) of the work**
* Aim of this stock market prediction project is to create software that analyzes previous stock data of certain companies.
* To predict and analyze the stock price of the company with the help of historical data and present data using ARIMA.

Objective of this project is

* To the study of past stock prices to predict the trend of prices in future.
* To help investors in the direction of movement of the share prices.
* To save investors from suffering a big loss in future.
* To determine the future value of a company stock or other financial instrument traded on an exchange.
  1. **Introduction to ARIMA**

Before working with non-stationary data, the Autoregressive Integrated Moving Average (ARIMA) Model converts it to stationary data. One of the most widely used models for predicting linear time series data is this one.

The ARIMA model has been widely utilized in banking and economics since it is recognized to be reliable, efficient, and capable of predicting short-term share market movements.

Consider you have a certain value A that is influenced by another value B. Then you must determine the link between data points A and B in order to determine linear regression. These two values, A and B (A’s previous value), are now linked in such a way that A’s current value is predicated on A’s previous value. As a result, any future value for A will be determined by its current value. The algorithm determines the linear regression of (Present fitted values) vs. (Past fitted values).

* For data acquisition - Use **of Yahoo finance API** to fetch stock data for past 6 months.
* For data processing - Scale data using sklearn preprocessing library

Since it is a time series data and we will be creating sequences out of the data fetched, it is better to build **ARIMA** model.

**1.5 Literature survey**

The study, “Stock Chart Pattern recognition with Deep Learning” [1] evaluated the performances of CNN and LSTM for recognizing common charts patterns in a stock historical data. It presents two common patterns, the method used to build the training set, the neural networks architectures and the accuracies obtained.

In paper [2], A “window” of various durations is run, breaking it into frames that scale in duration to a single size and to 1 in amplitude. The received frames are converted into 2D matrices and fed for analysis to a 2D convolutional NN, which determines the probability of frames belonging to the classes of patterns. The accuracy of the detector is about 98.6 % with a convolutional NN response speed of about 0.65 seconds per 1000 data samples, which corresponds to an analysis of the closing prices of trades on the exchange for more than 2.5 years.

A lot of studies provide strong evidence that traditional predictive regression models face significant challenges in out-of sample predictability tests due to model uncertainty and parameter instability. Recent studies introduce particular strategies that overcome these problems. Support Vector Machine (SVM) is a relatively new learning algorithm that has the desirable characteristics of the control of the decision function, the use of the kernel method, and the sparsely of the solution. In this paper, we present a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Firstly, four company-specific and six macroeconomic factors that may influence the stock trend are selected for further stock multivariate analysis. Secondly, Support Vector Machine is used in analyzing the relationship of these factors and predicting the stock performance. Our results suggest that SVM is a powerful predictive tool for stock predictions in the financial market. [3]

The prediction of the stock market can generate an actual financial loss or gain, so it is practically important to enhance the predictability of models. Consequently, many studies have been trying to model and predict financial time series, using statistical or soft computational skills that are capable of examining the complex and chaotic financial market. In recent years, deep learning techniques have been actively applied based on their excellent achievements in various classification problems. In this study, we constructed a stock price prediction model based on RNN using LSTM units, which is one of the typical methodologies of deep learning. We integrated GA and LSTM network to consider the temporal properties of the stock market, and utilized the customized architectural factors of a model. The LSTM network used in this study is composed with two hidden layers, which is a deep architecture for expressing nonlinear and complex features of the stock market more effectively. GA was employed to search the optimal or near-optimal value for the size of the time window and number of LSTM units in an LSTM network. [4]

Sentiment analysis has proven to be effective while analyzing people's attitudes by examining large social data. In this, a novel approach is designed to extract people's opinions on specific topics by relying on social media content. 70% of dataset is used for training whereas remaining dataset is used during testing. It has been determined that the proposed classifier, classifies the positive, negative and neutral sentiments with an accuracy of 98.32% From the experiment, it has been analyzed that the detection accuracy of sentiments has been increased by 8.99 % from the existing work. The main advantage of this work is that a stop word panel is added in to the GUI, so that a user can add or remove the stop words as per the need. [5]

A system is called offline when system takes images as inputs and tries to predict characters from given images by applying machine learning methods. We have worked on offline character recognition problem in this project. Many machine learning methods have been proposed over the years for solving this problem. In this paper we implemented 6 most popular machine learning methods to solve offline handwriting character recognition problem and compare the performance results to decide which method gives best accuracy results under pre-defined conditions. We have selected 92255 images from NIST Special 19 Database and used them as input images during the training phase of the selected machine learning methods. These methods are SVM, Decision Tree, Bag of Trees, Artificial Neural Networks (ANN), Deep learning network with autoencoders and Convolutional Neural Networks (CNN). We implemented all of these methods and compare the performance of the results according to accuracy metric. The results obtained from the comparison is going to help in deciding which ML method should be used to solve Offline Handwriting Character Recognition problem. [6]

It is seen that, regression architectures are capable of capturing dynamics and are able to make predictions. We trained the model using the data of stock and was able to predict stock price of stock. This shows that, the proposed system is capable of identifying some inter relation with in the data. Also, it is evident from the results that, SVR, RFR & DTR model is capable of identifying the changes in trends. For the proposed methodology DTR is identified as the best model. It uses the information given at a particular instant for prediction. Even though the other two models are used in many other time dependent data analysis, it is not out performing the DTR model in this case. This is due to the sudden changes that occurs in stock markets. The changes occuring in the stock market may not always be in a regular pattern or may not always follow the same cycle. Based on the companies and the sectors, the existence of the trends and the period of their existence will differ. The analysis of these type of trends and cycles will give more profit for the investors. To analyze such information we must use networks like DTR as they rely on the current information. [7]

Based on the results obtained, we conclude that the two companies under consideration have less correlation. The stock value change does not depend on the stock exchange index. It is dependent on the sentiments of social media. The prediction using machine learning algorithms do not give accurate results since the correlation between them is less. Results are not accurate as the dependency is less than 50% for all variables. But the graph trends between TCS and Infosys show similar variation except at some points where it was opposite. Combination of natural language processing techniques for analysis and summarization of text can help in handling such cases. Apart from the parameters which are considered in this paper there can be other parameters which can affect the stock shares such as Inflation, Deflation, International currency and gold rates and International economic policies, etc. Other techniques that can be used are Momentum, Mean Reversion and Martingales. [8]

This paper presents a survey of different techniques such as machine learning techniques, hidden Markov model, ARIMA model and also deep learning techniques. It is observed that selection of the right parameters for the dataset used for prediction plays important role good prediction accuracy. Various machine learning models as well as hybrid and ensemble model give higher rate of accuracy. To get even better accuracy fundamental analysis can be used which uses sentiment analysis and feature selection along with machine learning and deep learning techniques. [9]

Stock price prediction is an important issue in the financial world, as it contributes to the development of effective strategies for stock exchange transactions. In this paper, we propose a generic framework employing Long Short-Term Memory (LSTM) and convolutional neural network (CNN) for adversarial training to forecast high-frequency stock market. This model takes the publicly available index provided by trading software as input to avoid complex financial theory research and difficult technical analysis, which provides the convenience for the ordinary trader of nonfinancial specialty. Our study simulates the trading mode of the actual trader and uses the method of rolling partition training set and testing set to analyze the effect of the model update cycle on the prediction performance. Extensive experiments show that our proposed approach can effectively improve stock price direction prediction accuracy and reduce forecast error. [10]

**2. Overall Description**

**2.1 Product perspective**

* This project will provide a way to investors, stock markets to invest money in order to potentially earn a company's share.
* Stock markets promote investment. The raising of capital allows companies to grow their businesses, expand operations and create jobs in the economy. This investment is a key driver for economic trade, growth and prosperity.

**2.2 User roles and rights**

* When user will enter the web application, he/she will see the stock related news on the home page.
* Everyday top stocks will appear on the home page. Also there will be a search bar, if a user searches for the stock, he/she will see the stock information of the company.
* User can also see the predictions of that stock.

**2.3 User documents**

Email-id

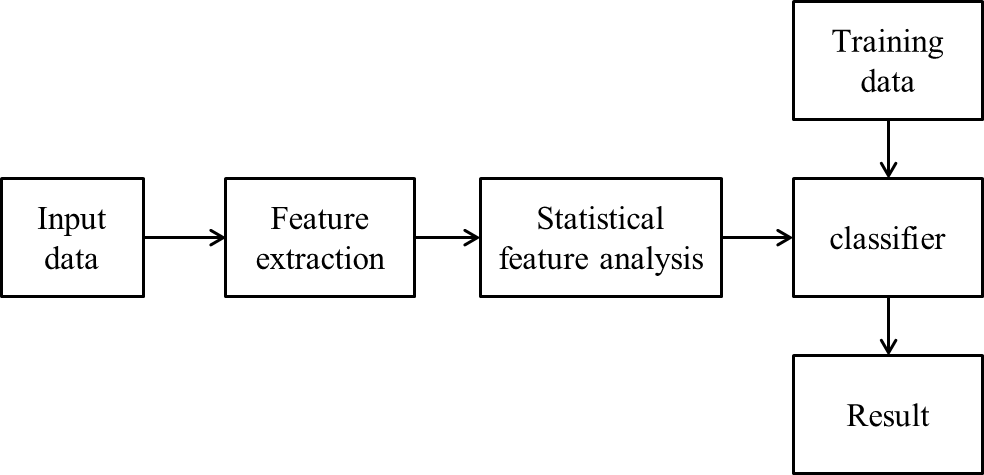
Phone no

Demat Account details

Birthdate

**2.4 Design and implementation constraints**

Stock price of particular commodity and stock value in previous years is acts as input to training. Statistical features are extracted and analyzed from given data and fed to classifier for comparison.it forms stock chart as training data. Current stock values of commodity are given as input to system. Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process



**Implementation plan**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Activity** | **I**  **week** | **II**  **week** | **III**  **week** | | **IV**  **week** | **V**  **Week** | **VI**  **week** | | **VII**  **week** | **VIII**  **week** | **IX**  **week** |
|  | **July 15** | **July 31** | **Aug 20** | | **Sept 25** | **Oct 12** | **Oct 26** | | **Nov 22** | **Dec 10** | **Dec 28** |
| **Initiate the project** |  |  |  | |  |  |  | |  |  |  |
| Communication |  |  |  | |  |  |  | |  |  |  |
| Literature survey |  |  |  | |  |  |  | |  |  |  |
| Define scope |  |  |  |  |  |  |  | |  |  |  |
| Develop SRS |  |  |  |  |  |  |  | |  |  |  |
| **Plan the project** |  |  |  | |  |  |  | |  |  |  |
| Design mathematical model |  |  |  | |  |  |  | |  |  |  |
| Feasibility Analysis |  |  |  | |  |  |  | |  |  |  |
| Develop work breakdown structure |  |  |  | |  |  |  |  |  |  |  |
| Planning project schedule |  |  |  | |  |  |  |  |  |  |  |
| Design UML and other diagrams |  |  |  | |  |  |  | |  |  |  |
| Design test plan |  |  |  | |  |  |  | |  |  |  |
| Design risk management plan |  |  |  | |  |  |  | |  |  |  |

**2.5 Assumptions and dependencies**

It is assume perfect working conditions with min 2 GB RAM and above 1GHz speed

**3. System Features**

**3.1 Description**

User can see predictions about bank nifty.

Latest news about IPOs and FPOs.

Daily news about stocks.

We are using LSTM ,CNN , ARIMA algorithm for stock pattern and prediction.

* 1. **Response Sequences**

When user will enter the web application, he/she will see the stock related news on the home page. Everyday top stocks will appear on the home page. Also there will be a search bar, if a user searches for the stock, he/she will see the stock information of the company. User can also see the predictions of that stock.

Admin takes the stock data from Yahoo finanace API. Admin will train the data using ARIMA model. The data set from Yahoo finance API contains information like previous closing, opening, high, low, volume of the stock of that company

Using various best algorithms to improve the accuracy of stock prediction we proposes a stock price method based on ARIMA for the stock closing price of the next days. We can adopt CNN to efficiently extract features from the data which are the items of the previous 6 months. And then, we will adopt ARIMA to predict the stock price with the extracted feature data

* 1. **Functional requirement**
* **User Module**
  + When user will enter the web application, he/she will see the stock related news on the home page.
  + User can also see the predictions of that stock.
* **Admin**
  + Admin takes the stock data from Yahoo finanace API.
  + Admin will train the data using ARIMA model.
  + Store data of different values of particular stock
* **Prediction**
  + Using various best algorithms to improve the accuracy of stock prediction we proposes a stock price method based on CNN-LSTM for the stock closing price of the next days.
  + We can adopt CNN to efficiently extract features from the data which are the items of the previous 6 months.

1. **External Interface Requirement**

**4.1 User Interfaces**

* The system shall use standard user interface controls such as buttons, text boxes, radio buttons, check boxes, labels, list boxes, spin boxes, combo boxes, sliders, scroll bars, tabs, tool tips, progress bars, and file selection dialogs .
* The user interfaces shall be presented as web pages and shall be displayed by web browsers.
* The system shall provide a visual display such as a progress bar icon while any function is being performed.
* Front-end software : Python
* Back-end software: MYSQL

**4.2 Hardware Interfaces**

* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor – SVGA

**4.3 Software Interfaces**

* Programming language: Python
* Database: MySQL
* ARIMA model
  1. **System Requirement**

**Database requirement**

MySQL

**Software requirement**

1. Operating System - Windows 7/8/10
2. Front End - Python
3. IDE - Python Idle
4. Algorithm- ARIMA

**Hardware Requirement**

1. Processor - Intel i3/i5/i7
2. Speed - 1.1 GHz
3. RAM - 2 GB(min)
4. Hard Disk - 40 GB
5. Floppy Drive - 1.44 MB
6. Key Board - Standard Windows Keyboard
7. Mouse - Two or Three Button Mouse
8. Monitor – SVGA
9. **Other Non-functional Requirement**

**5.1 Performance Requirements**

* RAM should be empty enough to run program smoothly.

## 5.2 Safety Requirements

* The data safety must be ensured by arranging for a secure and reliable transmission media.
* The source and destination information must be entered correctly to avoid any misuse or malfunctioning.
* Safety requirements against the natural disaster and accidents.
* Failures due to technical issues.

**5.3 Security Requirements**

* The access to the system is given only to valid users. We need a specific ID and password to get access to the system.
* Communication needs to be restricted when the application is validating the user or license.
* By incorporating a robust and proven SQL into the system, reliable performance and integrity of data is ensured. There must be a power backup for server system.

**5.4 Software Quality Attributes**

* **Reliability**

In order to ensure reliability, this system is being designed using software that is established to be stable and easy to use.

* **Availability**

This system is designed to run 24/7 and be readily available to the user.

* **Software Quality Assurance**

The quality of software and the security patches will be provided by the team of concerning developers, if the user finds any loopholes he/she may contact the concerned team and the team will provide them with suitable patches.

* **Maintainability**

The parts of the system coded in Python.

* **Portability**

The user interface shall run on Microsoft Internet Explorer 8.0, Mozilla Firefox 3.5, Google Chrome 3.0, and Apple Safari 4.0.

* 1. **Appendix A : Definition/References Links**

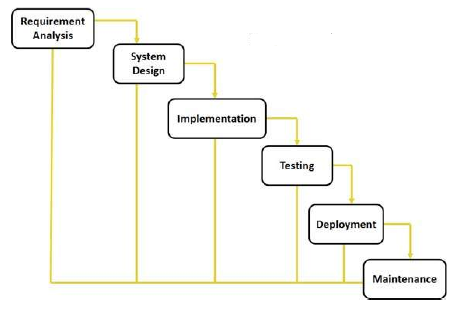
1. Marc Velay and Fabrice Daniel, “Stock Chart Pattern recognition with Deep Learning”, Researchgate, June 2018
2. Victor Skuratov, Konstantin Kuzmin, Igor Nelin, Mikhail Sedankin, “Application of a convolutional neural network to create a detector of technical analysis figures on exchange quotes charts”, (2019), «EUREKA: Physics and Engineering» Number 6 DOI: 10.21303/2461-4262.2019.001055
3. International Journal of Current Trends in Engineering & Technology Volume: 02, Issue: 01 (JAN-FAB, 2016) 18 Stock Market Prediction Using Support Vector Machine Mr. Sachin Sampat Patil, Prof. Kailash Patidar, Assistant Prof. Megha Jain
4. Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction Hyejung Chung and Kyung-shik Shin Sustainability 2018, 10, 3765; doi:10.3390/su10103765
5. An optimized CNN based robust sentiment analysis system on big social data using text polarity feature Komalpreet Kaur, Chitender Kaur, Tarandeep Kaur Bhatia International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6, April 2019
6. Performance Comparison of Machine Learning Methods for Solving Handwriting Character Recognition Problem Ş.G.KIVANÇ1, A.E. BAKTIR2 and B.ŞEN3 International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES’18), May 11-13, 2018 Safranbolu, Turkey
7. JSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 07, 2018 | ISSN (online): 2321-0613 All rights reserved by www.ijsrd.com 408 Stock Market Prediction using RFR, DTR & SVR Ravikant1 Suman Kumar Swarnkar2L. P. Bhaiya
8. International Journal of Computer Sciences and Engineering Open Access Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693 Stock Market Analysis and Prediction using Hadoop and Machine Learning Piyush Jain, Kaustubh Bhat, HarshalKesharwani3 , Pritesh Bhate, Khushboo P Khurana
9. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072 © 2019, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 4842 Stock Market Forecasting Techniques: A Survey Rashmi Sutkatti1, Dr. D. A. Torse
10. Xingyu Zhou, Zhisong Pan, Guyu Hu, Siqi Tang, Cheng Zhao “Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets”, Mathematical Problems in Engineering Volume 2018, Article ID 4907423, 11 pages

**6. 2 Appendix B : Analysis Model: SDLC model Implementation**

The waterfall model is a sequential design process, used in software development processes, in which progress is seen as flowing steadily downwards (like a waterfall) through the phases of conception, initiation, analysis, design, construction, testing, production/Implementation and maintenance. Waterfall approach was first SDLC Model to be used widely in Software Engineering to ensure success of the project. In the waterfall approach, the whole process of software development is divided into separate phases. In Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially. Following is a diagrammatic representation of different phases of waterfall model.

**1. Requirement Gathering and analysis**: All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification doc.

**2. System Design:** The requirement specifications from first phase are studied in this phase and system design is prepared. System Design helps in specifying hardware and system requirements and also helps in defining overall system architecture.



Waterfall model

**3. Implementation:** With inputs from system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality which is referred to as Unit Testing.

**4. Integration and Testing:** All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.

**5. Deployment of system:** Once the functional and non-functional testing is done, the product is deployed in the customer environment or released into the market.

**6. Maintenance:** There are some issues which come up in the client environment. To fix those issues patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment. All these phases are cascaded to each other in which progress is seen as flowing steadily downwards (like a waterfall) through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name “Waterfall Model”. In this model phases do not overlap.

**DFD**

