# pure logo copy

# Janardan Bhagat Shikshan Prasarak Sanstha’s

# CHANGU KANA THAKUR ARTS, COMMERCE, SCIENCE COLLEGE

# Re-accredited ‘A+’ Grade by NAAC

# College with Potential for Excellence Status Awarded by

# University Grants Commission

# ‘Best College Award’ by University of Mumbai

# New Panvel (W)

PROJECT PROPOSAL ON

**“Stock Market Prediction using Data Mining Techniques”**

## By

## MR. RAHUL BALIRAM PATIL

(M.Sc.-II (SEM-III) COMPUTER SCIENCE)

Under the Guidance of

Prof. Ms.

Department of Computer Science

**CERTIFICATE**

This is to certify that the Project Proposal entitled

**“Stock Market Prediction using Data Mining Techniques”**

Is successfully completed by **Rahul Baliram Patil**, Examination Seat Number under the guidance of **Prof. Ms.,** during the academic period of 10th June, 2017 to 28th Dec, 2017as per the Syllabus, and the fulfilment for the completion of the M.Sc.-II (Semester-III) in the Computer Science of **University of Mumbai**. It is also to certify that this is original work of the candidate done during academic year 2017-2018.

**Place:**

**Date:**

**Internal Examiner Head of Department**

**External Examiner**

**ACKNOWLEDGEMENT**

No project is ever completed without the guidance of those expert have already traded this past before and hence become and master of it and as a result, our leader. So I would like to take this opportunity to take all those individuals how have helped me in visualizing the project.

It is indeed a matter of great pleasure and proud privilege to be able to present this project proposal on “**Stock Market Prediction using Data Mining Techniques**”.

The completion of the project work is a milestone in student life and its execution is inevitable in the hands of guide. I express my deep gratitude to my project guide **Prof.**  for providing timely assist to my query and guidance that they give their experience in this field past many year. They had indeed been a lighthouse for us in this journey.

I would like to tender our sincere thanks to the H.O.D. **Prof. Mrs. P. M. Jadhav** and all the teachers for their co-operation.

I would also like to express our deep regards and gratitude to the Principal **Dr. .**

I would wish to thank the non – teaching staff and my friends who have helped me all the time in one way or the other.

Really it is highly impossible to repay the depth of all the people who have directly or indirectly helped me for performing the project.

**“STOCK MARKET PREDICTION USING DATA MINING TECHNIQUES”**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Title** | **Page No.** |
| 1 | Introduction | 6-8 |
| 2 | Related Work | 9-10 |
| 3 | Objective | 11 |
| 4 | Methodology | 12 |
|  | 1. ARIMA (Auto Regressive Integrated Moving Average ) |
|  | 1. ANN ( Artificial Neural Networks) | 13 |
|  | 1. Clustering | 13 |
| * 1. K-Means | 14 |
|  | 1. Classification |  |
|  | 1. Dataset | 15 |
|  | 1. System Configuration | 16 |
| 5 | References | 17-18 |

**INTRODUCTION**

Today we live and breathe data. Forecasting the stock data is an important financial subject which involves an assumption that the fundamental information publicly available to the past predictive relationship for future stock return. Stock market prediction contain uncovering market trends, planning investment tactics, identifying the best time to purchase the stock and which stock to purchase. A stock exchange or business sector is a non-direct, non-parametric framework that difficult to model. It is mix speculators which need purchase or offer of hold a share at a specific time. Prediction will continue to be an exciting locale of research, making scientists in the analytics field always desiring to enhance the existing forecasting models. The motivation is that companies and individuals are empowered to make investment decision to develop viable system about their future endeavours.

Stock trend forecasting is considered as one of the most difficult tasks to achieve in money related gauging because of the difficulty to achieve in money related gauging because of the difficulty in the multifaceted world of stock market. Many investor in the stock market are finding a technique that could guarantee easy profiting by forecasting the stock trends and minimize the risk of investing. This motivates the researchers in the domain field to delve and develop new forecasting models. Time series data analysis techniques use verifiable information as the premise for evaluating future results. Time series data can be defined as numerical data collected in a particular sequence over a period of time at regular intervals. The time series data can include the values collected at the end of every week, month, quarter, or year etc. The intention is to find if there is any link between the data collected so far and in what way does the data changes. In order to reduce the risk of investment, exchange the securities between the seller and buyer are facilitated by the stock exchanges. A stock exchange is an organization or a place where the stock traders or investors can deal with stock. Some example of stock market organization S&P, NSE, BSE, NASDAQ etc.

R is a programming language and environment for statistical processing and graphics. The   
R is generally utilized among analysts the data and data excavators for statistical programming and data analysis. R language was created by Ross Ihka and Robbert Gentleman at the University of Auckland, New Zeland and is presently maintained by the R development core team. Amid the most recent decade, the energy originating from both scholarly world and industry has lifted the R programming dialect to turn into the most essential tool for computational insights, perception and data science. Around the world, a large number of analysts and information researchers use R language to take care of their most difficult issues in the fields going from computational science to quantitative promoting.

**RELATED WORK**

The objective of forecasting the stock returns, fundamental analysis and technical analysis are used by number of researchers. The stock trading rules are developed based on the information related to economics, ancient data of stock trading.

The most efficient way to forecast the future is to understand the present scenarios. The author Banerjee D [ 1] tried to develop an appropriate model that helps to forecast the unseen values of the Indian stock market, based on the information collected on the monthly closing stock indices. Based on the ARIMA model they predict the future stock indices which have the strong performance of the Indian economy.

The authors Li Bing [2] have extracted the ambiguous text through MLP techniques to get the real stock price movements and public sentiments. It has been said that public emotions may be co-related that has shown through Twitter. Data miming algorithms to mine Twitter data in order to forecast the stock trends using sentimental analysis which comes under fundamental analysis.

To forecast stock price trend the author Tao Xing [3 ] have introduced a method based on Hidden Markov Model. Hidden Markov Model proposed by Baum and Egon, which is a kind of Markov chain and is used for the pattern recognition technique. This paper finds the hidden relationship existing between the Hidden Markov Model and stock prices.

It is tedious task for the stock market financial specialists to guesstimate the pattern of the stock exchange costs as effectively as could be allowed to settle on the best exchanging choices. The Vishwanath [4] have proposed a system called APST, which performs the pre-processing of verifiable stock time arrangement in information produce the grouping of approximated values by utilizing multi-scale segment mean methodology. To locate the closet neighbour object they utilize the Euclidian separation way to recognize the comparative arrangement of articles.

The Ayodele [5] have used the ARIMA model to develop an extensive process of building stock price predictive model by obtaining data from NYSE and NSE. Artificial Neural Networks (ANNs) model is very popular due to its ability to learn pattern from data and infer solution from unknown data. Hybrid approaches also engaged to improve stock price predictive models by exploiting the unique strength of each them. The result obtained from real life data demonstrated the potential strength of ARIMA models to provide investor short- term prediction decision making process.

Stock exchange markets facilitate saving and investment that are beneficial to increase the effectiveness of national economy. The Li Zhe [7] has used the method of technical analysis which trading rules were established based on the ancient data of stock trading price and volume. Technical analysis uses various methods that aim to predict future stock price movements based on the assumption can be determine historical stock prices.

**OBJECTIVE**

1. To take an overview of the Indian stock market and encapsulate various investment avenues available.
2. To know various options available in the capital market to invest.
3. To study investment behaviour of investors and the factors that affects their investment decisions.
4. To know the satisfaction of investor regarding return of different investment avenues.
5. To know investor’s perception regarding investment of stock market.
6. To organize stock in a fair, transparent and competitive way.
7. To learn about trading of stock in stock exchanges.

**METHODOLOGY**

1. **ARIMA (Auto Regressive Integrated Moving Average )**

ARIMA is a forecasting technique that project future values of a series based inertia. The first step applying ARIMA methodology is to check for stationary series remain fairly constant level over time. If trend exists as most economic or business application, then data is not stationary. The data should be constant variance in fluctuations over time. To process the data we use ARIMA (p, d, q) model.

p = order of the auto regressive part

d = degree of first moving differencing involved

q = order of the moving average part

In auto regressive and moving averaged model forecast stock trends.

* 1. **Auto Regressive Technique**

Auto regression technique estimates the future values based on the previous values. The first order of auto regression process is represented.

X(t) = A(1) \* X(t-1) + E(t)

Where X(t) = time series under investigation

A(1) = the auto regressive parameter of order 1

X(t-1) = the time series lagged 1 period

E(t) = the error term of the model

This simply mean any given value X(t) can be explained by some function of previous value X(t-1).

* 1. **Moving Average Technique**

A moving average is a technique to find overall idea with in a data set. It finds the future trends based on the previous values.

1. **ANN Algorithm ( Artificial Neural Networks)**

ANN computational model is capable of estimation and pattern reorganization. It is robust with respect to noisy and erroneous data, and is able to learn adapt to the environment Neural Network is applicable to the problem that algorithm is indefinable or exhaustive search is infeasible.

Neural Network is multi- layer perception and use back propagation algorithm is its learning.

1. **CLUSTERING**

Clustering is the grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities. Regarding to data mining, this methodology partitions the data implementing a specific join algorithm, most suitable for the desired information analysis of Stock market values.

* 1. **K-means clustering**

k-means is an iterative clustering algorithm in which items are moved among set of cluster until the desired set is reached. This technique is used to classify the data which have no previous knowledge about the data or training set. The parameter k denotes the amount of cluster required to partition the data. The idea of this clustering technique is, give k number of cluster we can define k centres, one for each cluster based on all sample belonging to a cluster.

In our Stock Market Prediction calculate the historic return and volatility and the proceed the K-Means clustering algorithm to divide the stock into distinct groups based upon return volatilities

1. **CLASSIFICATION**

**4.1. Decision tree (D- Tree)**

Decision tree is a classification method which yields output as flowchart – like tree structure. The result from D-Tree is highly interpretable, but the outcome must be represented in categorical data. D- Tree algorithm called “J48” is applied to classify future stock market.

**4.2. K –Nearest Neighbors (KNN)**

KNN is a non-parametric lazy learning algorithm that predicts class of the object based on the k closest training example in the feature space. An object classified by a majority vote of its neighbors.

1. **DATASET:**

The dataset used for this work can make us understand the algorithms more clearly.

The effects algorithm can be understood in much depth

1. **SYSTEM CONFIGURATION**

**H/W System Configuration:**

* System - Core-2-Due 2.4Ghz
* Speed - 2.4Ghz
* RAM - 4GB
* Hard Disk - 500GB
* Keyboard - Standard Windows Keyboard(Neosoft)
* Mouse - Standard Intex
* Monitor - 19” LED

**S/W System Configuration:**

* Operating system - Win 10
* Database - MySql/Excel
* Language - R/Weka/Python

**REFERENCES**

**Journal Reference**

[1] Banerjee, D., "Forecasting of Indian stock market using time-series ARIMA model", 2nd IEEE International Conference on Business and Information Management (ICBIM), January 2014, pp. 131-135.

[2] Li Bing, Chan, K. C. C., C. Ou, "Public sentiment analysis in Twitter data for prediction of a company's stock price movements", 11th IEEE International Conference on e-Business Engineering (ICEBE), November 2014, pp. 232-239.”

[3] Tao Xing, Yuan Sun, Qian Wang, Guo Yu, “The analysis and prediction of stock prices”, IEEE International Conference on Granular Computing (GrC), December 2013, pp. 368–373.

[4] Vishwanath R. Ha, Leena Sa, Srikantaiah K. Ca, K. Shreekrishna Kumar b., P. Deepa Shenoya, Venugopal K. Ra, S. S. Iyengarc, L. M. Patnaik, "Forecasting stock time-series using data approximation and pattern sequence similarity", International Journal of Information Processing (IJIP), September 2013, pp. 90-100.

[5] Ayodele A. Adebiyi, Aderemi O. Adewumi, Charles K. Ayo, "Stock price prediction using the ARIMA model", 16th IEEE International Conference on Computer Modelling and Simulation (UKSim), March 2014, pp. 106 -112.

[6] Qasem A. Al-radaideh, Adel Abu Asaf, Eman Alnagi, “Predicting stock prices using data mining techniques”, The International Arab Conference on Information Technology 2013.

[7] Li Zhe; “Research on China's stock exchange markets: problems and improvements”, International Conference on Education and Management Technology, 2010. pp 465-469.

* **URL Reference**