A KNN BASED APPROACH FOR PREDICTION OF STOCK MARKET TRENDS

[1] Ch. Sindhu Priyanka, B.Tech, M.Tech, Assistant Professor, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

[2] Gudavalli Tarun Sindhuja, M.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

Abstract: Stock market trends is interesting and challenging research topic. Developed countries' economies are measured according to their power economy. Currently, stock markets are considered to be an illustrious trading field because in many cases it gives easy profits with low risk rate of return. Stock market with its huge and dynamic information sources is considered as a suitable environment for data mining and business researchers. In this paper, we applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock prices for a sample of six major companies listed on the Jordanian stock exchange to assist investors, management, decision makers, and users in making correct and informed investments decisions. According to the results, the kNN algorithm is robust with small error ratio; consequently the results were rational and also reasonable. In addition, depending on the actual stock prices data; the prediction results were close and almost parallel to actual stock prices.

1. INTRODUCTION

Recent business research interests concentrated on areas of future predictions of stock prices movements which make it challenging and demanding. Researchers, business communities, and interested users who assume that future occurrence depends on present and past data, are keen to identify the stock price prediction of movements in stock markets. However, financial data is considered as complex data to forecast and or predict. Predicting market prices are seen as problematical, and as explained in the efficient market hypotheses that was put forward by Fama (1990). The efficient market hypotheses is considered as bridging the gap between financial information and the financial market it also affirms that the fluctuations in prices are only a result of newly available information and that all available information reflected in market prices. In addition to purchasing and selling stocks and shares in stock markets, each stock is not only c characterized by its price, but also by other variables such as closing price which represents the most important variable for predicting next day price for a specific stock. There is a relationship and specific behavior exists between all variables that effect stock movements overtime. Different economic factors, such as political stability, and other unforeseeable circumstances are variables that have been considered for stock price predictions. In stock predictions, a set of pure technical data, fundamental data, and derived data are used in prediction of future values of stocks. The pure technical data is based on previous stock data while the fundamental data represents the company's activity and the situation of market. Combining data mining classification approaches in stock prediction yields a future value for each unknown entities of companies stocks values based on historical data. This prediction uses various methods of classification approaches such as neural networks, regression, genetic algorithm, decision tree induction, and k-Nearest Neighbors. In classification approaches, a data set is divided into training dataset and testing set. KNN uses similarity metrics to compare a given test entity with the training data set. Each data entity represents a record with n features. 1 From the above research background, it is clear that each of the algorithms can effectively solve stock prediction problems. However, it is vital to notice that there are specific limitations for each of them. The prediction results not only are affected by the representation of the input data but also depend on the prediction method. Moreover, using only prominent features and identifying them as input data instead of all features can noticeably develop the accuracy of the prediction models Employing tree-bat ensemble methods and deep learning algorithms for predicting the stock and stock market trend is a recent research activity. In light of employing bagging and major. vote methods, Tsai used two different kinds of ensemble classifiers, such as tremendous and homogeneous methods. They also consider macroeconomic features.

2. LITERATURE SURVEY

2.1 Poonam Somani, Shreyas Talele, Suraj Sawant in 2014[1] were proposed the Stock market is the most popular investment scheme promising high returns albeit some risks. An intelligent stock prediction model would thus be desirable. So, the aims at surveying recent literature in the area of Neural Network, Hidden Markov Model and Support Vector Machine used to predict the stock market fluctuation. Neural networks and SVM are identified to be the leading machine learning techniques in stock market prediction area. Also, a model for predicting stock market using HMM is presented. Traditional techniques lack in covering stock price fluctuations and so new approaches have been developed for analysis of stock price variations. Markov Model is one such recent approach promising better results. In this paper a predicting method using Hidden Markov Model is proposed to provide better accuracy and a comparison of the existing techniques is also done.

2.2 Feng Wang, Yongquan Zhang, Hang Xiao, Li Kuang ,Yi Lai in 2015[2] were proposed the focus on the problem of how to design a methodology which can improve the prediction accuracy as well as speed up prediction process for stock market prediction. As market news and stock prices are commonly believed as two important market data sources, we present the design of our stock price prediction model based on those two data sources concurrently. Firstly, in order to get the most significant features of the market news documents, we propose a new feature selection algorithm (NRDC), as well as a new feature weighting algorithm (N-TF-IDF) to help improve the prediction accuracy.

2.3 Meryem Ouahilal ,Mohammed El Mohajir, Mohamed Chahhou, Badr Eddine El Mohajir in 2016[3] were proposed the Predicting stock prices is an important task of financial time series forecasting, which is of great interest to stock investors, stock traders and applied researchers. Many machine learning techniques have been used in recent times to predict the stock price, including regression algorithms which can be useful tools to provide good accuracy of financial time series forecasting. In this paper, we propose a novel hybrid approach which combines Support Vector Regression and Hodrick-Prescott filter in order to optimize the prediction of stock price. To assess the performance of this proposed approach, we have conducted several experiments using Maroc Telecom (IAM) financial time series.

2.4 Pankaj K. Bharne, Sameer S. Prabhune in 2017[4] were proposed the Swarm intelligence (SI) is powerful, newly emerged domain belongs to the field of Artificial Intelligence. The SI is inspired from the behavior of biological entities such as honey bee, fireflies, bat, cuckoo, ant etc. The basic idea of SI is that, the collective behavior of agents with a very limited set of rules. In recent SI is applied in various kind of application including appropriate stock market price movement. They initially describes the details of a stock market, SI and its various types of algorithm. Meghna Misra, Ajay Prakash Yadav, Harkiran Kaur in 2018[5] were proposed Predicting the stock market has been an area of interest not only for traders but also for the computer engineers. Predictions can be performed by mainly two means, one by using previous data available against the stock and the other by analysing the social media information. Predictions based on previous data lack accuracy due to changing patterns in the stock market al.so, some fields might have been missed due to their insignificance in some stocks or unavailability of data. For

example, some models may require `return rate' as a parameter for stock prediction, but the available data might not have it

2.5 Rachna Sable, Shivani Goel, Pradeep Chatterjee in 2019[6] were proposed the Stock market prediction is a crucial and challenging task due to its nonlinear, evolutionary, complex, and dynamic nature. Research on the stock market has been an important issue for researchers in recent years. Companies invest in trading the stock market. Predicting the stock market trend accurately will minimize the risk and bring a maximum amount of profit for all the stakeholders. During the last several years, a lot of studies have been done to predict stock market trends using Traditional, Machine learning and deep learning techniques. Srinath Ravikumar, Prasad Saraf in 2020[7] were proposed The stock market is an interesting industry to study. There are various variations present in it. Many experts have been studying and researching on the various trends that the stock market goes through. One of the major studies has been the attempt to predict the stock prices of various companies based on historical data. Prediction of stock prices will greatly help people to understand where and how to invest so that the risk of losing money is minimized. This application can also be used by companies during their Initial Public Offering (IPO) to know what value to target for and how many shares they should release

2.6 Ajinkya Yelne, Dipti Theng in 2021[8] were proposed Using Supervised Machine learning, our project is to analyzed and predict the stock value. As due to pandemic situation stock market trading is the most learned and become important activities to earn money as a second source of income in the people of India. The concept of predicting a stock's future worth is known as stock trading or stock prediction. So, we need a better tool to get out of this contemptuous situation and help the common man to make profit. In this research, we discuss a Machine Learning strategy that will be taught using publicly released stock data to build information, then using that information to make a valid prediction. For accuracy and prediction of stock Classification and Regression Algorithms are used with Kaggle dataset a machine learning technique comes under supervised learning that are Random Forest, Decision Tree, and Logistic Regression to predict stock prices for the given company previous year data, employing prices with daily trading prices.

3. EXISTING SYSTEM
The genetic algorithm had been adopted by Shin et al.

(2005); the number of trading rules was generated for Korea Stock Price Index 200 (KOSPI 200), in Sweden Hellestrom and Homlstrom (1998) used a statistical analysis based on a modified kNN to determine where correlated areas fall in the input space to improve the performance of prediction for the period 1987-1996. Both models mentioned were provided in the Zimbabwe stock exchange to predict the stock prices which included Weightless Neural Network (WNN) model and single exponential smoothing (SES) model Mpofu (2004). Clustering stocks approach was provided by Gavrilov et al. (2004) to group 500 stocks from the Standard & Poor. The data represented a series of 252 numbers including the opening stock price.

DISADVANTAGES:

• Stock prices prediction is interesting and challenging research topic. Developed countries' economies are measured according to their power economy. Currently, stock markets are considered to be an illustrious trading field because in many cases it gives easy profits with low risk rate of return

• Existing algorithms doesn't give better results

4. PROPOSED SYSTEM

The Proposed System is based on the evaluating performance of various machine learning algorithms like KNN Algorithm to predict stock prices using stock dataset. The dataset using in normal values (continuous) and binary data means convert stock values to binary data by using indicators. Here we compare machine learning algorithms after all the comparisons random forest and decision tree algorithms give better results. The programming language used to predict the stock market using machine learning is Python.

ADVANTAGES

• Accuracy is Very high

• High Efficiency
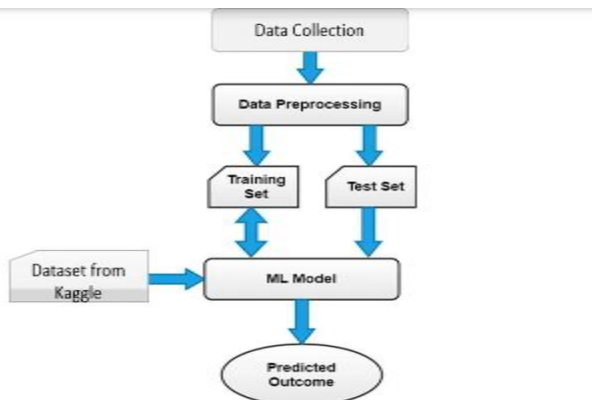
SYSTEM ARCHITECTURE

Fig 1: System Architecture

5. UML DIAGRAMS

1. CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. It is also known as a structural diagram. Class diagram contains • Classes • Interfaces • Dependency, generalization and association.
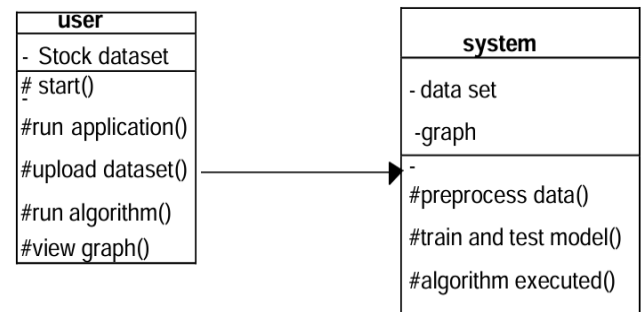

Fig 5.1 shows the class diagram of the project

2. USECASE DIAGRAM:

Use Case Diagrams are used to depict the functionality of a system or a part of a system. They are widely used to illustrate the functional requirements of the system and its interaction with external agents (actors). In brief, the purposes of use case diagrams can be said to be as follows

• Used to gather the requirements of a system.

• Used to get an outside view of a system.

• Identify the external and internal factors influencing the system.

Use case diagrams commonly contains

• Use cases

• Actors

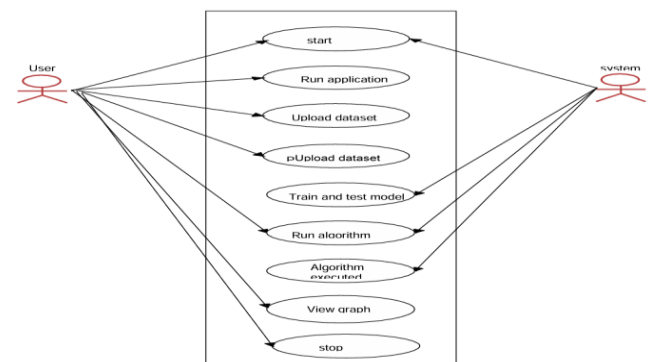• Dependency, generalization and association relationships.

Fig 5.2 Shows the Use case Diagram

3. SEQUENCE DIAGRAM:

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. Sequence diagrams are used to formalize the behaviour of the system and to visualize the communication among objects. These are useful for identifying additional objects that participate in the use cases. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.
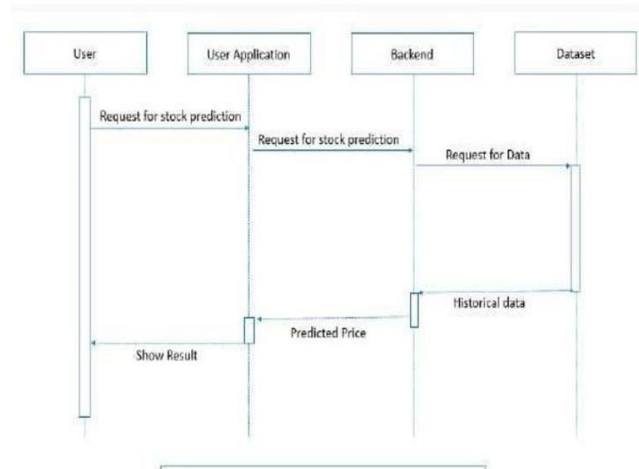

Fig 5.3 Shows the Sequence Diagram

6. RESULTS

Output Screens

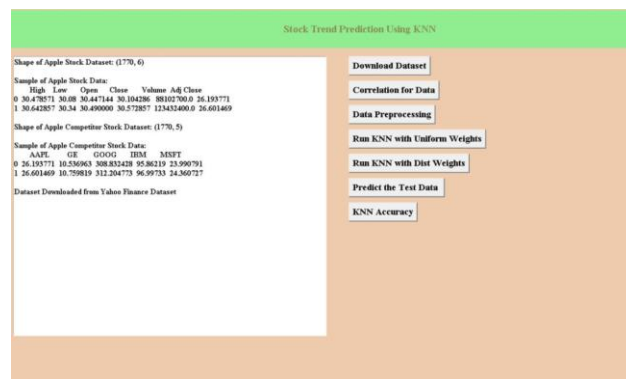To run project double click on 'run.bat' file to get below screen

Fig6.1 Upload stock Dataset

In above screen click on 'Upload Stock Dataset' button to load dataset Below3 datasets we are using to train and test all algorithms
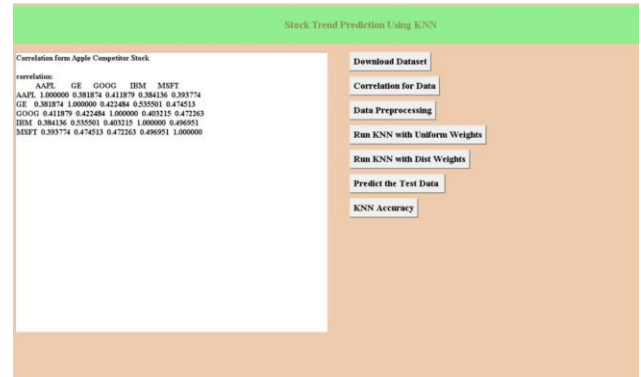

Fig6.2 Correlation for Data

In above screen click on correlation for data button to get the correlation results and then click on Preprocessing button to get below screen Below screen showing preprocessing results.
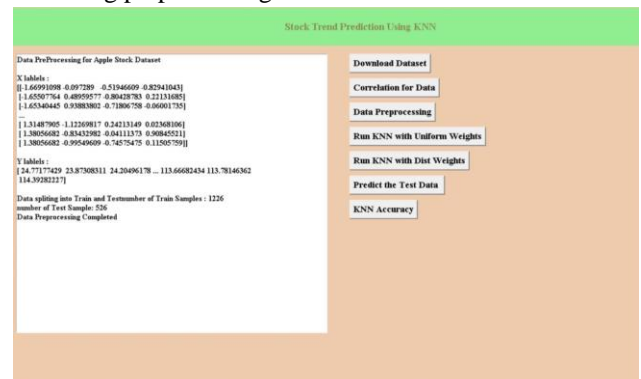

Fig6.3 Pre-processing the dataset

In above screen shows the preprocessing results and dataset contains some missing values so to remove missing values and to split dataset into train and test part so click on 'Run KNN Algorithm' button to get below screen

Fig6.4 Run KNN Algorithm with Uniform Weight

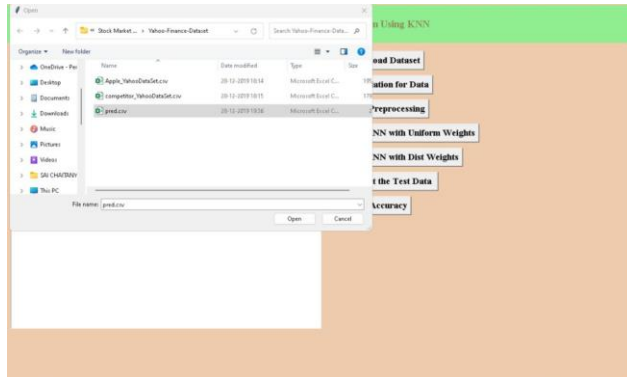In the above screen shows the KNN Algorithm Accuracy

Fig6.5 Upload Test Dataset

In above screen upload the test dataset to predict the stock market trends

Fig6.6 Prediction Result

Fig6.6 Accuracy Comparison Graph

7. CONCLUSION

By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the random forest algorithm. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

FUTURE SCOPE

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole

8. REFERENCES

[1] Poonam Somani, Shreyas Talele, Suraj Sawant. "Stock market prediction using Hidden Markov Model" In 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference

[2] Feng Wang, Yongquan Zhang, Hang Xiao, Li Kuang, Yi Lai. "Enhancing Stock Price Prediction with a Hybrid Approach Based Extreme Learning Machine" In 2015 IEEE International Conference on Data Mining Workshop (ICDMW)

[3] Meryem Ouahilal, Mohammed El Mohajir, Mohamed Chahhou, Badr Eddine El Mohajir."Optimizing stock market price prediction using a hybrid approach based on HP filter and support vector regression" In 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)

[4] Pankaj K. Bharne, Sameer S. Prabhune." Survey on combined swarm intelligence and ANN for optimized daily stock market price" In 2017 International Conference on Soft Computing and its Engineering Applications (icSoftComp)

[5] Meghna Misra, Ajay Prakash Yadav, Harkiran Kaur. " Stock Market Prediction using Machine Learning Algorithms:A Classification Study" In 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE)

[6] Rachna Sable, Shivani Goel, Pradeep Chatterjee. "Empirical Study on Stock Market Prediction Using Machine Learning" In 2019 International Conference on Advances in Computing, Communication and Control (ICAC3)

[7] Srinath Ravikumar, Prasad Saraf. "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms" in 2020 International Conference for Emerging Technology (INCET)

[8] Ajinkya Yelne, Dipti Theng ."Stock Prediction and analysis Using Supervised Machine Learning Algorithms" In 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)

[9] L Mathan prasad, M Gunasekaran."Analysing the Trend of Stock Market and Evaluate the performance of Market Prediction using Machine Learning Approach" In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)

[10] Vidushi Tiwari, Bhanu Prakash Lohani, Ajay Rana, Upendra Pratap Pandey, Bhavdeep Dhariwal. "Stock Market Prediction using different Machine Learning Algorithms" In 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)