STOCK VALUE PRICE PREDICTION - PREDICTING THE VALUE USING ALGORITHMS SVM , RANDOM FOREST AND KNN.

R PREMA[1], ADDANKI RADHIKA[2], SURESH AISWARYA[3]

[1]ASST . Professor & Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya
[2]Department of Computer Science and Engineering & Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya
[3]Department of Computer Science and Engineering & Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya

---

***Abstract*—Stock market prediction has gained monumental prevalence in financial markets globally. The ability to forecast the direction of a stock price or an index is very important for various purposes. The stock market helps to analyze the current market trends and minimize the risk of the investment.**

Prediction of stock value for a company or an entity is the most exhausting work to do After 2000 there have been many advancements where Machine learning and Artificial intelligence concepts got introduced for the prediction of stock value Depending upon the type of data that we have taken there are many machine learning algorithms which we can use as tools for price prediction which among them are KNN, Random forests, Support Vector Machine, ANN LTSM etc. which have the capacity to solve the complex data sets.

In our project here using historical stock data, we developed two models to make short-term predictions for the stock price. The models were refined by including the influence of NASDAQ index. Advanced mathematical techniques were used to formulate these models. Investors can use these models to obtain suggestions and pointers. We have used Random Forest KNN and SVM algorithms for the prediction of the stock market.

***Index Terms*—Machine Learning, Stock Analysis, Support Vector Machine, Random Forest, K-Nearest Neighbours .**

I. INTRODUCTION

Stock market prediction has gained a lot of importance in the universal financial markets . Prediction or forecasting has become an important tool for business as it helps the authority to make wise decisions . In order to reduce the potential risk of loss in the investment done and to save investors from the same ,Stock market prediction has multiple tools which identifies the opportunities for increasing or decreasing the stock value and making profits .

This project deals with advanced math modeling tools and addresses a very complex and popular area of the financial sector. Since my career goals are to go into the field of data science, and
to get involved in the stock analysis, I think this project helped me move forward intellectually. It provided me with a strong academic base and a positive experience regarding the stock market.

We decided to look mostly at NASDAQ, since it contains most of the stocks from the technology sector, the target of our interest. The Stock market encompasses a huge number of companies and is divided into different sectors. We decided to look at the technology sector. The reasoning behind this is that we live in the technological era and our lives are shaped by it every day. It is an area very popular in the society and we also have a personal interest. The technology sector, however, is hard to represent by modeling few stocks. That is why we started working on the project by choosing ten stocks from the Internet Information Providers Industry of the technology sector.
There were a few restrictions (challenges) set from the beginning for picking out the stocks. First, the company stocks should be relatively stable, thus most of the stocks we chose have gone public for several years now. The stock price data for these companies is available for each working day for the past year. Second, the price range of the stocks is above five dollars and below hundred dollars. We chose ten stocks by looking at their prices and putting them in 3 divisions. A price range of (mention the name of the currency) 5 – 20, 20 – 50 and above 50. These stocks with their ticker symbols are: Facebook Inc. (FB), Yahoo! Inc. (YAHOO), Twitter Inc. (TWTR), IAC/InterActiveCorp (IACI), GROUPON Inc. (GRPN), TechTarget Inc. (TTGT), ChinaCache Ltd (CCIH), Blucora Inc. (BCOR), J2 Global Inc. (JCOM), and eBay Inc. (EBAY). We decided to avoid the small market.

II. METHODOLOGY

What is Machine Learning ?

Artificial Intelligence, Data science and Machine learning stands to be the top trending technologies in the present tech world . The trending topics Data Mining ,Bayesian analysis added a little more demand to these technologies. Machine Learning is a branch of artificial intelligence which automates analytical model building. It is a method of data analysis through which systems can learn from data, identify patterns in data and make decisions with a very less human-machine interaction. Here, learning represents recognizing and analyzing the input data and making accurate decisions

depending on the data supplied. Since it is difficult to consider each and every decision depending on the input data given, algorithms were developed which makes it easier. The algorithms build knowledge from the provided and the past available data through principles
like Statistical science, probability, logic reinforcement learning e.t.c.

Python is the most demanding platform used for conducting research and development
in production systems. It's a very versatile language with numerous modules,
libraries and packages that provide us ample ways to achieve a task.

Python has libraries like NumPy, SciPy, SciKit-Learn and Matplotlib which can
be said to be the best tools for Data Analysis. Another major and important
benefit of these libraries is that they are colossally used in creating scalable
machine learning algorithms. Popular machine learning techniques , in particular
classification , regression , clustering and recommendation are implemented using python.

Algorithms Used

Support Vector Machine

Support Vector Machine or SVM in short is an important machine learning algorithm Which is used in problems of classification and regression. Although it can be used for both
classification and regression, in majority of cases, it is used in classification. In the SVM algorithm, the value of each plot is represented by its coordinate and is plotted
as an n-dimensional space. Classification is the type of problem in which the hyper plane segregates the data into two different classes. A support vector is nothing but the coordinate of individual observation.

How does it work ?
The first thing SVM does is separating data for which it uses an imaginary line called Hyper plane . A hyperplane is an n-dimensional flat Euclidean space where the n-1
a dimensional subset divides that space into two different disconnected parts .
The hyper plane divides the data into two parts and makes it easier to train . SVM algorithm takes data as input and outputs the line which separates the two possible classes.

The main function of the algorithm SVM is to find the closest points from the line to the both sides of the classes . These points are called support vectors . Now , we calculate
the distance between the two support vectors which is called Margin . The goal of the algorithm is to maximize the margin . The Optimal Hyperplane is the hyperplane for which the margin is maximum

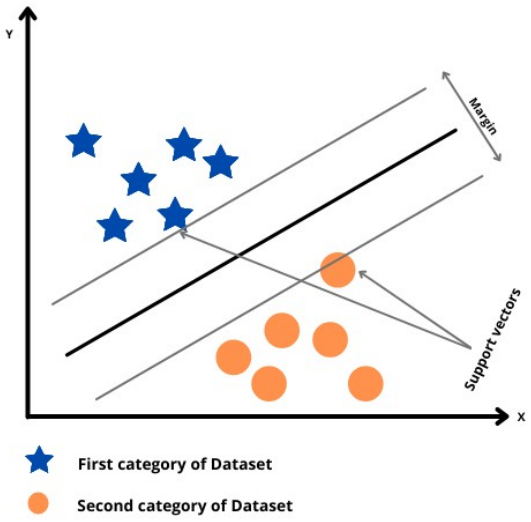


Fig 1 - Hyperplane separating data set

Non-linearly separable data

This is the type of data which cannot be separated linearly with an imaginary line or hyperplane . A straight line cannot be drawn to classify the data . But through high
dimensions , this data can be converted into linearly separable data. An extra dimension called Z axis is added for which the coordinates are governed by the constraint
$z = x^2+y^2$
So, basically z coordinate is the square of distance of the point from origin. Let's plot the data on z-axis.

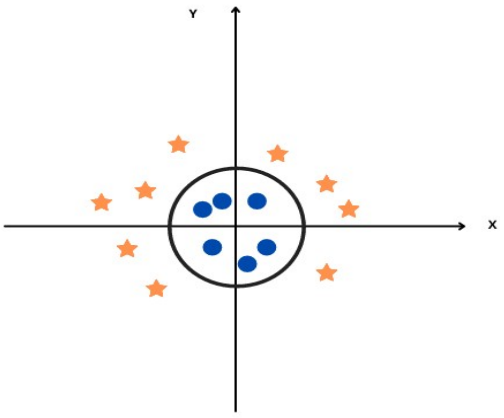


Fig 2 - Non-linearly separable data

Random Forest

Data science provides an ample number of classification algorithms which among them are logistic regression, support vector machines, naive Bayes classifiers, and decision trees. But random forest stands at the top of all other algorithms and is also said as one of the algorithms which gives the perfect and most accurate output as prediction .
Here in the project , we have examined how the basic decision trees work, how multiple decision trees are combined to make a huge random forest, and finally explain why random forests are that extraordinary at what is their prime performance.
Decision tree : For the random forest model , decision trees are the prime elements or the basic blocks . It is a supporting tool in the structure of a tree through which we get predictions as outcomes. Decision trees are very useful aspects of life because they come in use in multiple situations in life.
The Random Forest Classifier : As the name suggests , Random forest contains a lot of decision trees related to the same data which gives multiple predictions . Each individual decision tree spits out different predictions and the output which is given by most of the decision trees will be considered as output (see figure below).

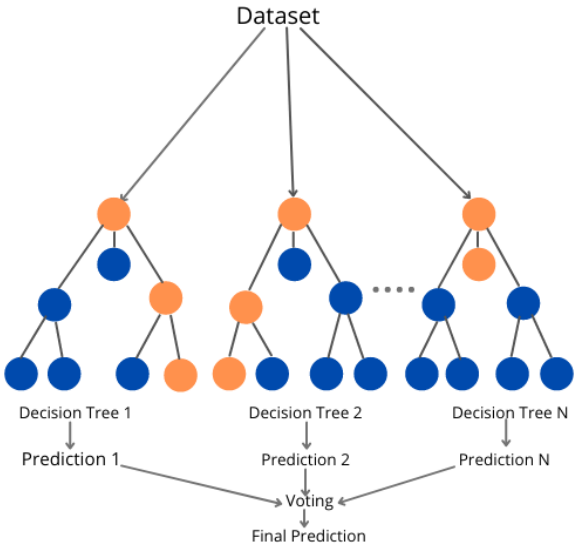


Fig 3 - splitting data into decision trees and getting final prediction

K-Nearest Neighbor

K-Nearest Neighbor is an Instance based algorithm based on supervised learning technique, which is used for classification and regression problem solving, it is mostly used for classification of predictive problems. KNN is a process of recognizing similarity of features between the new data and past labeled data. K in the algorithm KNN represents the number of data samples that we are taking from the neighborhood inorder to predict the class label of our specified point in sample space.

In this type of Algorithms we don't process the samples immediately after getting training samples, instead we store the samples and classify the instance then we will do the training and class label association with the test samples, so this is also called as Lazy Algorithm

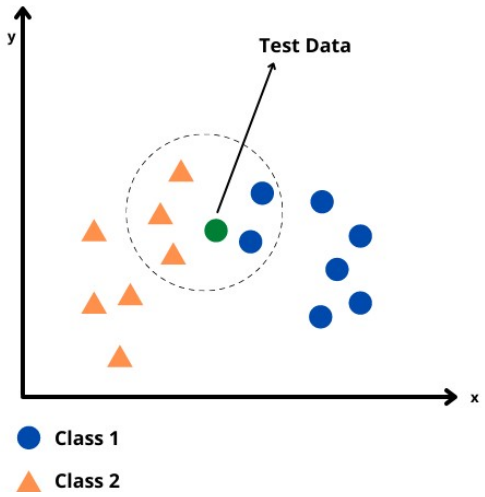


Fig 4 - KNN Classification

Step 1: Choosing the K value

We will choose the value of K which will decide the number of nearest samples that we need to consider to forecast the value of test samples

Step 2: Calculating the distance

Determine the nearest samples with respect to the test samples using Euclidean distance formula. The distance is calculated using the below formula

Step 3: Classifying the nearest points

we will classify the nearest points based on the class or category that belongs to, so going with the majority it concludes the class for that point .

KNN Regression:

KNN in Regression is nearly the same approach as classification, there will be only one difference that we won't take the bulk of class labels for the test sample. Instead, we will take the average of all the class labels and then we will set the value as a class label for the test sample.

III. RESULT

Fig.1

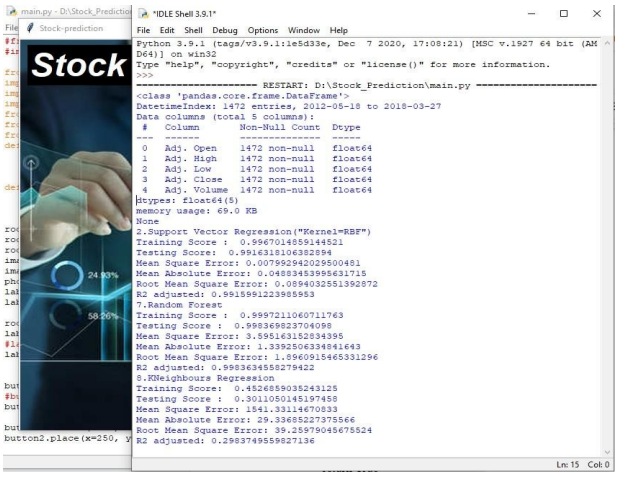


Fig.2

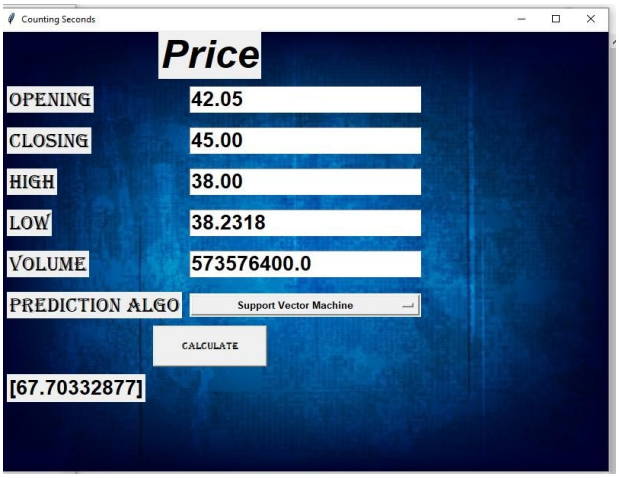
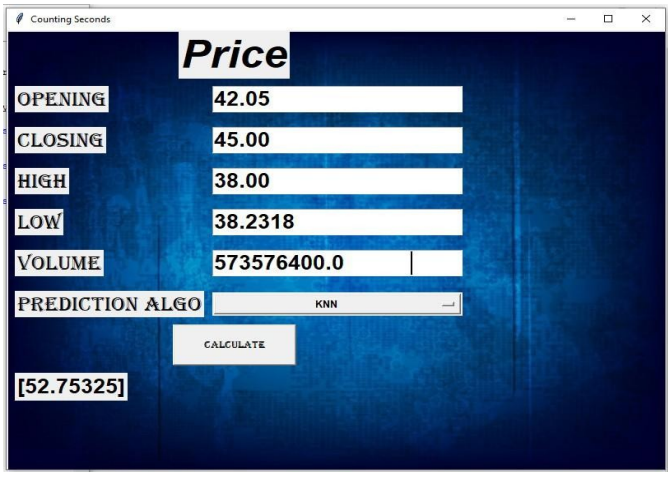


Fig.3

Fig.4

Fig.5



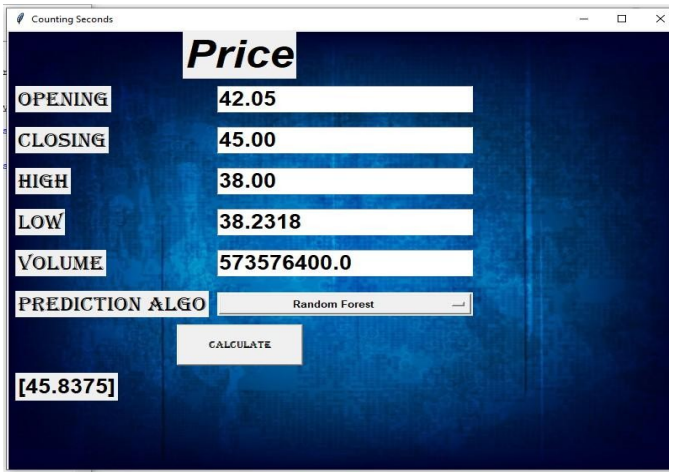
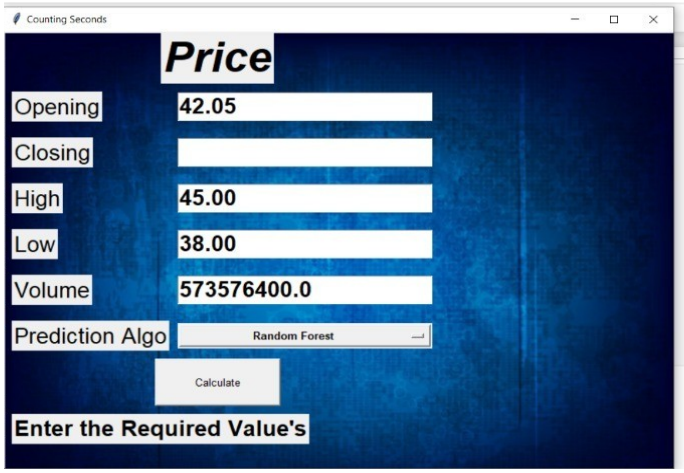


Fig. 6



IV. CONCLUSION

To wrap it all up , In this project we have used Random Forest , Support Vector machine and K-NN algorithms for the stock market prediction. The dataset is preprocessed successfully and gives the future predicted value of the stock with a slight variation of value between the algorithms . We made the project with an aim to make it easier for the stock investors to clear their hesitations and learn stock 0fluctuations practically . Our aim is to put an end to the myths fabricated on the stock market , untie the intricate knots of confusions and make it stocker friendly . Three different algorithms are designed in such a way that helps the stocker in making accurate decisions and make it sure that they don't end up in loss.

V. REFERENCES

[1] D. Y. N. Le, A. Maag and S. Senthilananthan, "Analysing Stock Market Trend Prediction using Machine & Deep Learning Models: A Comprehensive Review," 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA), Sydney, Australia, 2020, pp. 1-10, doi: 10.1109/CITISIA50690.2020.9371852.

[2] J. Rasheed, A. Jamil, A. Ali Hameed, M. Ilyas, A. Özyavaş and N. Ajlouni, "Improving Stock Prediction Accuracy Using CNN and LSTM," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 2020, pp. 1-5, doi: 10.1109/ICDABI51230.2020.9325597.

[3] Z. H. Kilimci and R. Duvar, "An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange (BIST 100)," in IEEE Access, vol. 8, pp. 188186-188198, 2020, doi: 10.1109/ACCESS.2020.3029860.

[4] Desai Mitesh Madhusudan, “Stock Closing Price Prediction using Machine Learning SVM Model ”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 8, doi: 10.22214/ijraset.2020.32154

[5] Shreenidhi Sriram, Sanjana Rangarajan, “Stock Market Prediction using Logistic Regression Analysis”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 8, doi: 10.22214/ijraset.2020.30744

[6] Sahil Bhatia , Srishti Kishore , Siddharth Nager , Mr. Anurag Agarwal, “Stock Price Trend Prediction”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 8, doi: 10.22214/ijraset.2020.6293

[7] Partho Protim Dey, Nadia Nahar, B M Mainul Hossain, "Forecasting Stock Market Trend using Machine Learning Algorithms with Technical Indicators", International Journal of Information Technology and Computer Science(IJITCS), Vol.12, No.3, pp.32-38, 2020. DOI: 10.5815/ijitcs.2020.03.05

[8] Anthony R. Caliñgo, Ariel M. Sison, Bartolome T.Tanguilig III,"Prediction Model of the Stock Market Index Using Twitter Sentiment Analysis", International Journal of Information Technology and Computer Science(IJITCS), Vol.8, No.10, pp.11-21, 2016. DOI: 10.5815/ijitcs.2016.10.02

[9] Isaac Kofi Nti, Adebayo Felix Adekoya, Benjamin Asubam Weyori, “Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana.” Applied Computer Systems , 2020. doi:10.2478/acss-2020-0004.

[10] Mosavi, Amir. “Training Multilayer Perceptron with Genetic Algorithms and Particle Swarm Optimization for Modeling Stock Price Index Prediction.” Entropy, 2020. doi:10.3390/e22111239.

(Data set taken from old/existing data from different companies.