STOCK PRICE PREDICTION USING K-NEAREST NEIGHBOUR(KNN)

# FINAL REPORT

## GROUP-12

## ANIRUDH MANDAVA

## LAKSHIT RAJPUT

## M V B PRANEETH REDDY

**EXECUTIVE SUMMARY:**

Here we applied k-nearest neighbor algorithm in order to predict stock prices for a sample of seven major companies listed on the NASDAQ stock market to assist investors, management, decision makers, and users in making correct and inform/ed investments decisions. For distance metric, Euclidian distance will be used. The average of the K nearest neighbour’s outcome will be using for prediction of the stocks. According to the results, the kNN algorithm is mildly robust with a good accuracy; consequently, the results were rational and reasonable.

**INTRODUCTION:**

Now a days many researching platforms are concentrating on prediction of stock prices. Researchers, business communities, and interested users who assume that future occurrence depends on present and past data, are more likely to predict the accurate values for stock price. However financial data is considered as complex data to forecast or predict. In addition to purchasing and selling stocks and shares in stock markets, each stock is not only 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. 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 companies’ 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 Nearest Neighbours(kNN). In classification approaches, a data set is divided into training data set and testing set. kNN uses similarity metrics to compare a given test entity with the training data set.

**PROJECT MOTIVATION:**

Presently used methods like EMH are problematical and are not so efficient. EMH affirms that the fluctuations in prices are only a result of newly available information, and that all available information reflected in market prices. Furthermore, it has been affirmed that stock prices do not pursue a random walk and stock prediction needs more evidence. So, we are using data mining techniques to predict values more accurately so that it would be user friendly.

**DATA DESCRIPTION:**

The sample data was extracted from the NASDAQ stock exchange from Kaggle data sets. The study sample included stock data of four elected companies listed on the NASDAQ stock exchange. Each of these companies has 5 attributes including date, high, low, opening and closing price. In kNN algorithm closing price is important factor to predict future stock price. This is the most common fluctuating data that has to be predicted.

**k- Nearest Neighbour:**

The KNN algorithm it is a popular method for classifying objects based on closest training examples in the feature space. though not popular it provides some basis for prediction with efficiency of over 50%. This algorithm starts with splitting the dataset which in training and testing dataset which is preferably 70:30. Then deciding for the value of K which is the number of clusters in which the training set is divided. The value of k is taken as sqrt(no. of data points in training set). Now the training is done using training dataset using the fit method and then extrapolate using the predict method. Then accuracy of the system is calculated using accuracy formula which compares the exact values and provides the output. KNN algorithm is applied on the data for a period

of 5 years to predict the stock price and compare it with the original movements for a period of 5 years. It turns out that the efficiency of the test data is around 65-70% if the data is not largely skewed else the accuracy is around 48 – 53%.

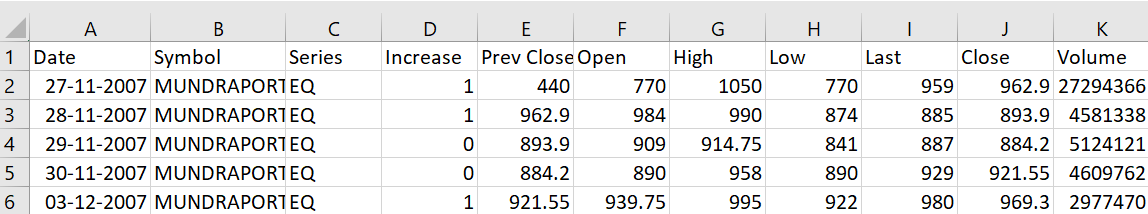
Experimental results of KNN algorithm on Yes

Bank stock are as follows:

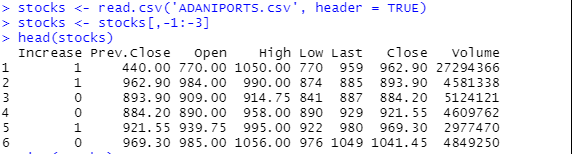
The accuracy for ADANIPORTS is around 48%. It turns out that there is a need for improvement as this prediction is highly volatile and can’t be thoroughly trusted by the users.

**Steps we used to perform kNN are**:

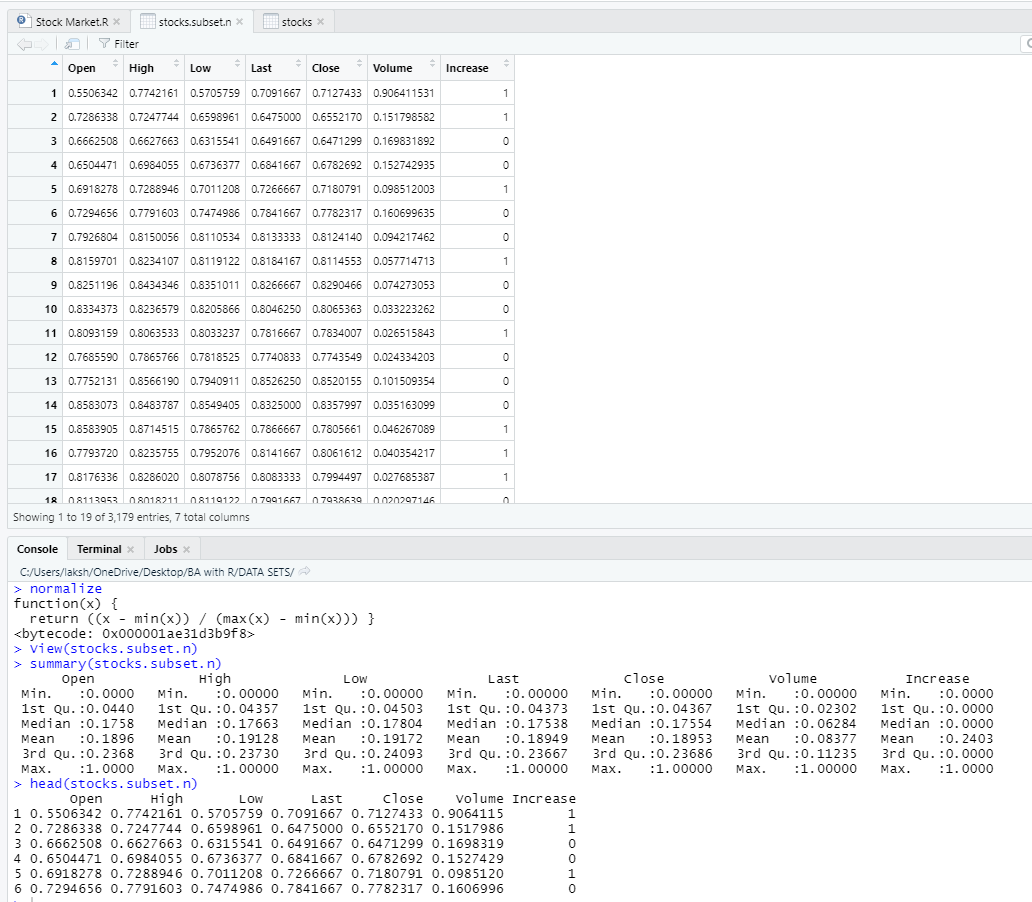
* Handle data: Get data using yahoo finance and save and the load the dataset from CSV and split into test/train datasets.
* Similarity: Calculate the distance between two data instances
* Neighbors: Locate k most similar data instances
* Response: Use a majority vote for the class labels of k nearest neighbors and generate a response from a set of data instances
* Accuracy: Summarize the accuracy of predictions
* Main: Tie it all together
* **SAMPLE DATA :**



* **SAMPLE DATA BEFORE NORMALIZE:**



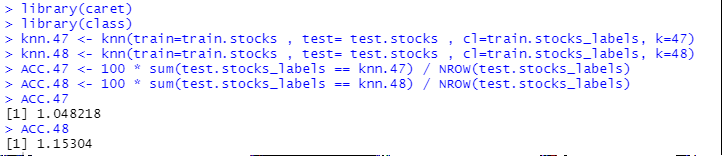
* **SAMPLE DATA AFTER NORMALIZE OF THE DATA:**



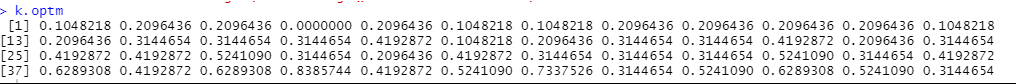
**N number of observations or number of rows= 2225**

So the sqrt of the no of rows or number of observation give us the nearest knn integer which is near 47 and 48.

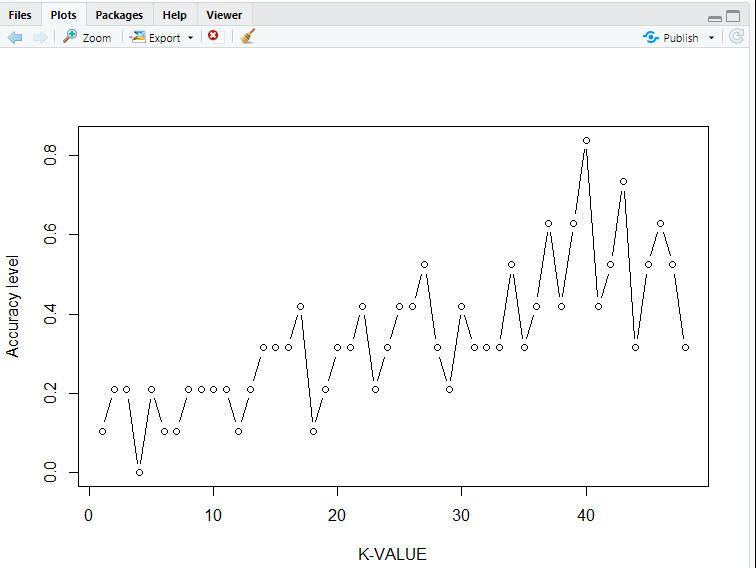
* **ACCURACY:**



* Sample data after iteration of I loop means for every value of knn what is the accuracy:



* **Bar Graph:**



**SAMPLE DATA ACCURACY OF DIFFERENT COMPANY STOCKS:**

|  |  |
| --- | --- |
| **SAMPLE DATA** | **ACCURACY%** |
| ADANI PORTS | 48% |
| HDFC | 55% |
| ICICI | 60% |
| BAJAJ AUTO | 61.5% |

**CONCLUSION:**

Investors can make a lot of money by analyzing the last five company’s stock prices. A stock represents equity which has a very higher rate of return so that the investor can think of investing in a company with a profound knowledge compared to bond or investing in land etc. Stocks are volatile which gives authority to the investor to sell them off at any point. So, the knn algorithm is useful to provide a complete accurate prediction which will be easy for the user to understand.

References:

1. R. Lakshman Naik, D. Ramesh , B. Manjula , Dr. A. Govardhan, “Prediction of Stock Market Index Using Genetic Algorithm”, International journal of Computer Engineering and Intelligent Systems www.iiste.org ISSN 2222-1719, Vol 3, No.7, 2012 https://iiste.org/Journals/index.php/CEIS/article/viewFile/2177/2189

2. Sadegh Bafandeh Imandoust, Mohammad Bolandraftar,“Application of KNearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background”, Int. Journal of Engineering Research and Applications ,Vol. 3, Issue 5, Sep-Oct 2013, pp.605-610 https://www.researchgate.net/publication/3048 26093\_Application\_of\_Knearest\_neighbor\_KNN\_approach\_for\_predict ing\_economic\_events\_theoretical\_background-

3. Phayung Meesad, Islam Rasel, “Predicting stock market using support vector Regreesion”, 2013 International Conference on Informatics, Electronics and Vision (ICIEV) <https://ieeexplore.ieee.org/document/6572570>

4. Ganesh Bonde and Rasheed Khaled,”Stock price prediction using genetic algorithms and evolution strategies”, semantic scholar, 2012 https://www.semanticscholar.org/paper/Stockprice-prediction-using-genetic-algorithms-andBonde/29fa7a9131fd975cbb81801311fbb75b4c41 bf0e-

5. Siddhaling Urolagin, Text Mining of Tweet for Sentiment Classification and Association with Stock Prices, 2017 International Conference on Computer and Applications(ICCA), https://ieeexplore.ieee.org/stamp/stamp.jsp?arnum ber=8079788&tag=1

6 .https://ieeexplore.ieee.org/stamp/stamp.jsp? arnumber=8079788

7. https://www.investopedia.com/terms/s/sharp eratio.asp

8. https://www.sciencedirect.com/science/article/ pii/S2405918818300060