International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

STOCK PRICE PREDICTION USING TWITTER DATASET

Y. SHIVA RAO[1], M N V SAI SHREYAS[2], SYDU VAMSHI KRISHNA SAI[3], PILLARISETTY NYMISHA[4], GUNIGANTI ANIRUDH REDDY[5]

[1]Assistant professor, Department of CSE, Malla Reddy College of Engineering Hyderabad, TS, India.
[2,3,4,5] UG students, Department of CSE, Malla Reddy College of Engineering Hyderabad, TS, India.

ABSTRACT:

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. Stock market with its huge and dynamic information sources is considered as a suitable environment for data mining and business researchers. In this project, we applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock prices for a company's stock data to assist investors, management, decision makers, and users in making correct and informed investments decisions. This algorithm uses the open, close, high, low values of a stock in a day and volumes of that stock to train the module. Then for testing, an opening value of the stock is taken from the user and given as test variable for the module. The module will return the predicted closing value of that stock. The differences between actual and predicted closing values of the stock can be interpreted using visualization graph plotted between those two data. 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.

Key words: KNN, testing stock, stock price data.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

I 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 (Kim, 2003). 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 (EMH) that was put forward by Fama (1990). The EMH 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. The EMH assert that stocks are at all times in equilibrium and are difficult for inventors to speculate. Furthermore, it has been affirmed that stock prices do not pursue a random walk and stock prediction needs more evidence. Data mining technology is used in analyzing large volume of business and financial data, and it is applied in order to determine stock movements. Mining temporal stock markets is required to provide additional capabilities required in cases where the existing data and their interactions need to be observed through time dimension. 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 neural networks, regression, genetic algorithm, decision tree induction, and k-Nearest Neighbors (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. Each data entity represents a record with n features. In order to predict a class label for unknown record, kNN selects k recodes of training data set that are closest to the unknown records.

2. RELATED STUDY

The literature attempting to prove or disprove the efficient market hypothesis can be classified in three strands, according to choice of variables and techniques of estimation and forecasting. The first strand consists of studies using simple regression techniques on cross sectional data. The second strand of the literature has used time series models and techniques to forecast stock returns following economic tools like Autoregressive Integrated Moving Average (ARIMA), Granger Causality Test, Autoregressive Distributed Lag (ARDL) and Quantile Regression to forecast stock prices. The third strand includes work using machine learning tools for prediction of stock returns.

The major drawback of the existing propositions in literature for stock price prediction is their inability to predict stock price movement in a short-term interval. The current work attempts to address this shortcoming by exploiting the power of deep neural networks in stock price movement modeling and prediction.

3 METHODOLOGY

We have decided to provide the interface for the users where they can manually select the stock data of the company whose market price is to be predicted. Then they can generate the vector the data items in that data set

using Generate Vector option. After generating vector, the training of the stock data is done. User can give the opening value whose closing value is to be predicted. Then algorithm takes that input as testing variable and gives to the trained module. Using the kNN algorithm on the given data set and user input the machine learning module will predict the closing value. Then the predicted value is displayed for the user. A visualization graph is used to report the effectiveness of the algorithm used.

How Does kNN Algorithm Works:

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.
- It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

classification, it performs an action on the dataset.

- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

4 RESULTS EXPLANATION

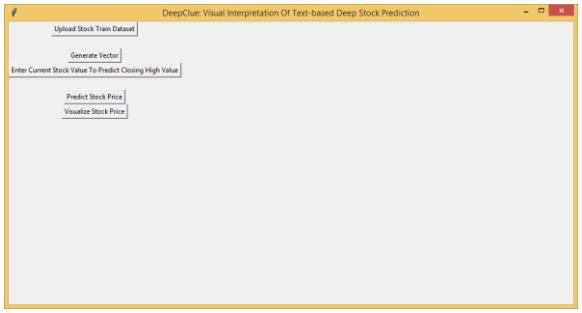

Fig.4.1. Admin page.

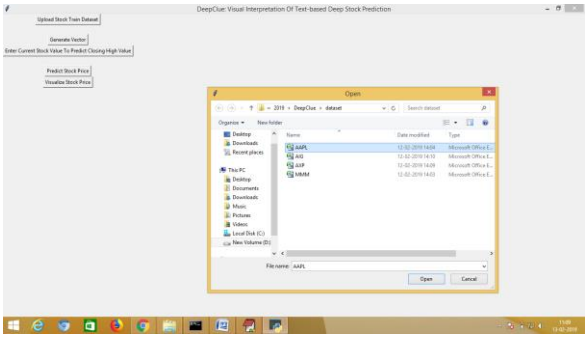

Fig.4.2. Uploading Data Set.

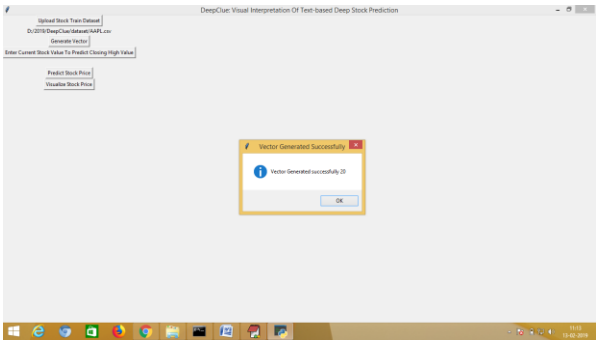

Fig.4.3. Generate vector.

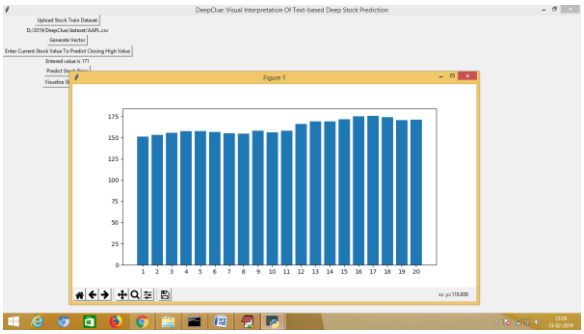

Fig.4.4. Data Visualization.

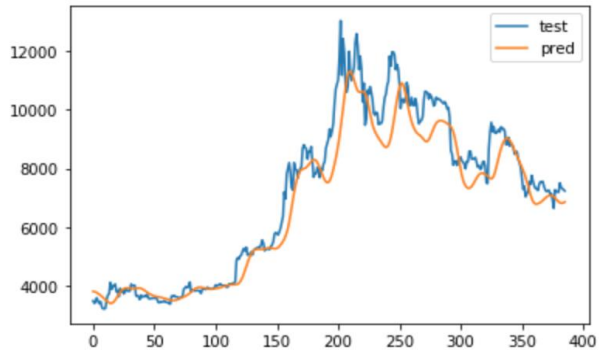

Fig.4.5. Actual vs Predicted Visualization.

CONCLUSION

Stock market prediction is a very difficult task because of the volatile nature of the movement of financial

share data for sectors and companies in the stock market. The ideal solution for achieving efficient and accurate forecasting is employing artificial intelligence in applying machine learning techniques. The kNN-algorithm used in this project gave the effective results. According to the results, kNN algorithm was stable and robust with small error ratio, so the results were rational and reasonable. In addition, depending on the actual stock prices data; the prediction results were close to actual prices. Having such rational results for predictions in specific, and for using data mining techniques in real life; this presents a good indication that the use of data mining techniques could help decision makers at various levels when using kNN for data analysis. So, we consider that employing this prediction model, kNN is real and viable for stock predictions.

REFERANCES

[1] Khedr, Ayman E., and Nagwa Yaseen. "Predicting stock market behavior using data mining technique and news sentiment analysis." International Journal of Intelligent Systems and Applications 9.7, pp. 22. (2017).

[2] Chittineni, Suresh, et al. "A Comparative Study of CSO and PSO Trained Artificial Neural Network for Stock Market Prediction." International Conference on Computational Science, Engineering and Information Technology. Springer, Berlin, Heidelberg,pp. 186-195,2011.

[3] M. Kumari and S. Godara, “Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction”, IJCST ISSN: 2229- 4333, vol. 2, no.2, (2011).

[4] Khalid, Balar, and Naji Abdelwahab. "Big Data and Predictive Analytics: Application in Public

Health Field.", International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol6, No.5, 2016.

[5] S.Archana and Dr. K.Elangovan, "Survey of Classification Techniques in Data Mining", International Journal of Computer Science and Mobile Applications, Vol. 2 Issue. 2, February 2014.

[6] Nyce, Charles. "Predictive Analytics White Paper, sl: American Institute for Chartered Property Casualty Underwriters." Insurance Institute of America, p.1, (2007).

[7] Shah, Dev, Haruna Isah, and Farhana Zulkernine. "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques." International Journal of Financial Studies, 7.2, pp. 26 (2019).

[8] Alkhatib, Khalid, et al. "Stock price prediction using k-nearest neighbor (kNN) algorithm." International Journal of Business, Humanities and Technology 3.3, pp. 32-44, (2013).

[9] Farshchian, Maryam, and Majid Vafaei Jahan. "Stock market prediction with hidden markov model." 2015 International Congress on Technology, Communication and Knowledge (ICTCK). IEEE, 2015.

[10] BhaveshPatankar and Dr. Vijay Chavda, "A Comparative Study of Decision Tree, Naive Bayesian and k-nn Classifiers in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 12, December 2014.