International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST5231035

Stock Trend Prediction Using KNN Algorithm

Anjali Patil[1], Gayatri Padole[2], Akansha Sontakke[3]

[1]Student of M-Tech Electronics Engineering Department in J D college of Engineering & Management, Nagpur

[2,3] Professor of M-Tech Electronics Engineering Department in J D college of Engineering & Management, Nagpur, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 05 April 2023
Published: 03 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
693-699

ABSTRACT

Stock forecasting has always been a difficult task for statisticians and financial analysts. The key strategy used to make this prediction is buying stocks with a high probability of price growth and selling stocks with a high probability of price decline. There are typically two approaches to stock market forecasting. One of them is fundamental analysis, which is dependent on a company's methodology and fundamental data. The performance of the supervised machine learning algorithm KNN (K-Nearest Neighbor) is evaluated by the author in this study. Stock trading is one of the most significant activities in the world of finance. Trying to anticipate the future value of a stock or other financial instrument traded on a financial exchange is known as stock market prediction. Python is the computer language used to make stock market predictions using machine learning. In this article, we present a Machine Learning (ML) approach that will be trained using the stock market data that is currently accessible, gain intelligence, and then use the learned information to make an accurate prediction. This study employs prices with both daily and up-to-the-minute frequencies and a machine learning method known as K-Nearest Neighbor to forecast stock prices for both large and small capitalizations and in the three separate marketplaces.

Keywords: Stock Price Prediction, K-Nearest Neighbors, Probabilistic Method.

I. INTRODUCTION

In the investment world, analysing financial data in securities has been a significant and difficult problem. Due to the conflicting impacts of information rivalry among significant investors and the unfavourable selection costs imposed by their knowledge advantage, stock price efficiency for publicly traded companies is challenging to accomplish. The two primary schools of thinking used to analyse the financial markets are

as follows. The first strategy is referred to as fundamental analysis [1]. Through qualitative and quantitative analysis, the fundamental analysis methodology determines a stock's intrinsic value in order to evaluate it. This method looks at the managerial, market, micro, and macroeconomic elements of an organisation. Technical analysis is the name for the second strategy. Utilising historical market data analysis is the method used in technical analysis to predict price direction. A number of charts are used in technical analysis to predict what is likely to happen. The several types of stock charts include candlestick, line, bar, point-and-figure, OHLC (open-high-low-close), and mountain charts. The charts can be viewed with price and volume in a variety of time intervals. The charts use a variety of indicators, including breakout, trending, momentum, resistance, support, and breakout. There have been a number of different ways to handle this kind of issue, from conventional statistical models to approaches based on artificial intelligence and computer learning. Vanstone and Tan reviewed the research on using soft computing in financial trading and investment. They divided the papers under consideration into the following categories: classification, hybrid approaches, time series, optimisation, and pattern recognition [3].
The survey revealed that the majority of research was being done in the area of technical analysis within the context of financial trading discipline. In order to assess the stock price trends by concentrating on macroeconomic analysis, an integrated fundamental and technical analysis model was looked at. Additionally, it examined how each company behaved in relation to its industry and the overall economy, which gives investors more information with which to make investment decisions. By combining the KNN methodology with technical analysis, a closest neighbour search (NNS)
method obtained the desired result. This model utilised technical analysis to historical price and trade volume data from the stock market. Stop loss, stop gain, and RSI filters were used as technical indicators. The distance function was applied to the gathered data by the KNN algorithm portion. Using the fundamental analysis method, this model was contrasted with the buy-and-hold technique.

II. PROPOSED SYSTEM

Everyday billions of dollars are traded on the exchange, and behind each dollar is an investor hoping to profit in one way or another. Entire companies rise and fall daily based on the behaviour of the market. Should an investor be able to accurately predict market movements, it offers a tantalizing promise of wealth and influence. It is no wonder then that the Stock Market and its associated challenges find their way into the public imagination every time it misbehaves [2]. The 2008 financial crisis was no different, as evidenced by the flood of films and documentaries based on the crash. If there was a common theme among those productions, it was that few people knew how the market worked or reacted. Perhaps a better understanding of stock market prediction might help in the case of similar events in the future. The suggested approach starts by processing the data using data set 2, where each record provides a stock's financial characteristics and the expected results in a structured category format. These records served as inputs for the prediction of stock price movements utilising a hybrid KNN-Probabilistic model is suggested. For KNN, the data points in metric space with a notion of distance are the features in the data set [5]. Each entry in the data set includes a group of vectors and class labels related to every vector. Each class label either reads "Profit" for classes that are positive or "Loss" for classes that are negative. The number of neighbours that can affect classification is determined by the k value. Finding the right k value is the first stage in KNN. The k value is highly reliant on the training set. A low k value indicates that noise will have a greater impact on the outcome, and a high value leads to an overfit model. The highest categorised generalizability was

achieved by the k value, according to the utilisation of k-fold cross-validation. When there are two classes, an odd integer is typically used as the k value so that a choice can be made based on the class value with the most occurrences [4].

III.IMPLEMENTATION

Double click on 'run.bat' file to get below screen

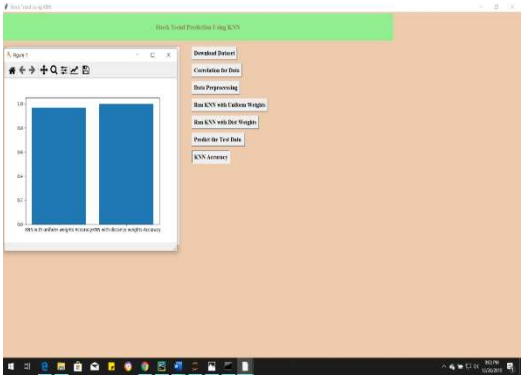

Figure 1: Stock Trend Prediction Using KNN

In above screen click on 'Download Button' download the Apple Stock and competitors' data from Yahoo Finance Dataset

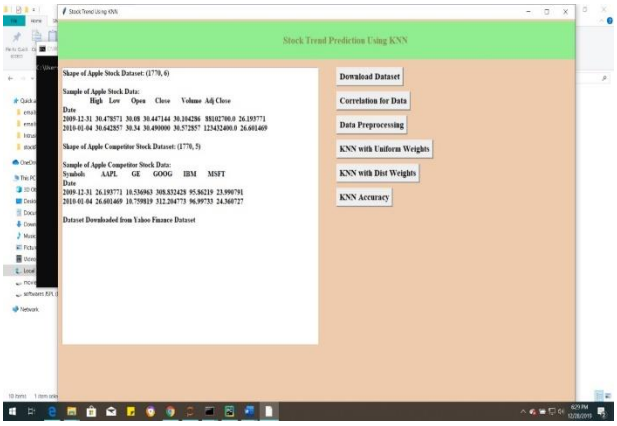

Figure 2: Download Dataset

In above screen Downloading of Apple Stock and Apple competitor Stock Data from Yahoo Finance Dataset [7].

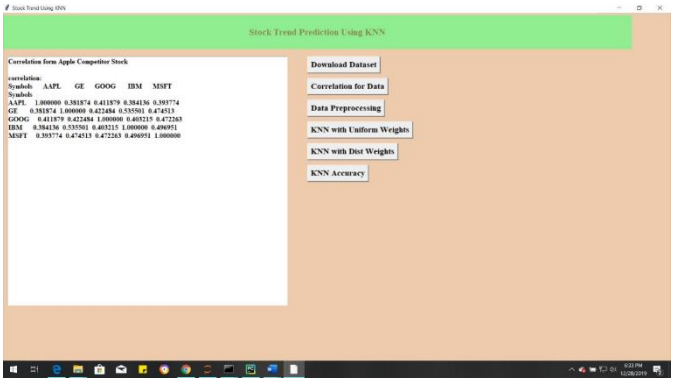

Figure 3: Correlations for Data

After pre-processing all missing values are dropped, Separating the label here, Scaling of X, find Data Series of late X and early X (train) for model generation and evaluation, Separate label and identify it as y and Separation of training and testing of model. In above screen we can see dataset contains total 1752 records and 1226 used for training and 526 used for testing [6]. Now click on 'Run KNN with Uniform Weights' to generate KNN model with uniform weights and calculate its model accuracy

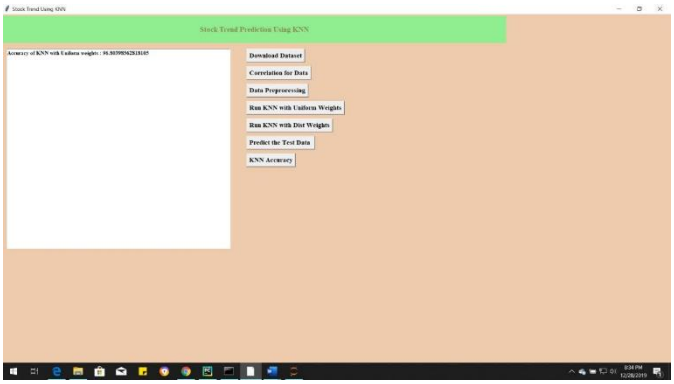

Figure 4: Run KNN with Uniform weight

In above screen we can see with KNN with uniform weights got 96.8% accuracy, now click on 'Run KNN with distance weights' to calculate accuracy

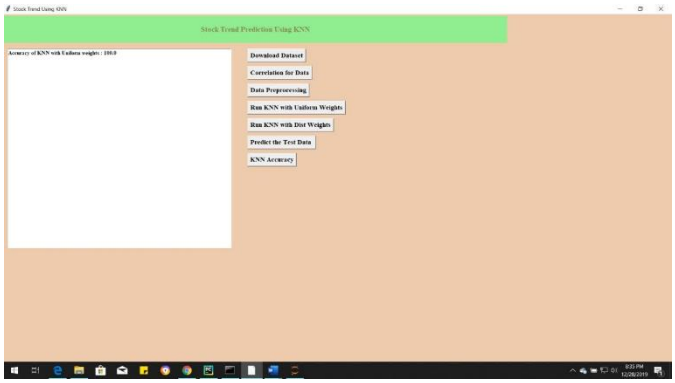

Figure 5: Run KNN with Dist. weight

In above screen we got 100% accuracy, now we will click on 'Predict Test Data 'button to upload test data and to predict whether test data stock market for both models. Accuracy score (>0.95) for most of the models [8].

In above screen we got 100% accuracy, now we will click on 'Predict Test Data 'button to upload test data and to predict whether test data stock market for both models.

Accuracy score (>0.95) for most of the models.

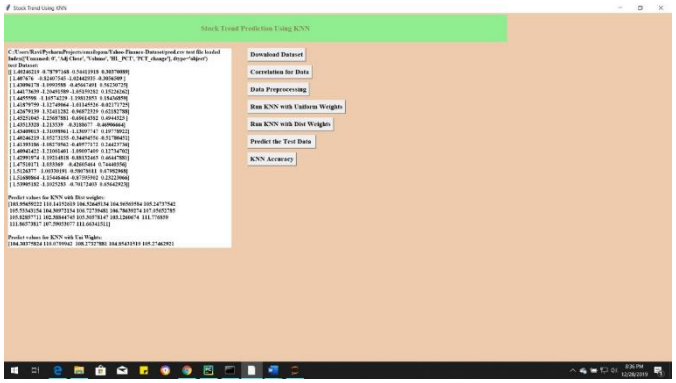

Figure 6: KNN Accuracy

In above screen for each test data, we got forecast values for Apple Stock for each test record. Now click on 'KNN Accuracy' button to save the predicted values for each model save in the local directory and Accuracy comparison to both the models

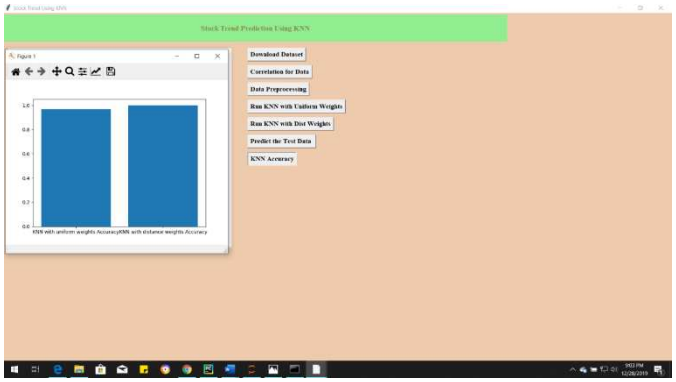

Figure 7: Test Data Results

IV. RESULT

The proposed model was evaluated and contrasted with KNN, Naive Bayes, OneR, and ZeroR, four other common algorithms. The test evaluated the MAE and RMSE as well as the accuracy with which the tested algorithms forecast changes in stock price. The test results are shown in Table 5. We estimate the estimated accuracy of the hybrid KNN-Probabilistic model to be 89.1725%, which is higher than the stand-alone KNN reported accuracy of 86.6667% and the Naive Bayes accuracy of 76.1194%. OneR and ZeroR classifiers have accuracy rates of 71.6418% and 64.1791%, respectively. The MAE and RMSE rates of the KNN-Probabilistic model are significantly lower than those of the other classifiers at 0.0667% and 0.2582%, respectively.

Classifier	Accuracy(%)	MAE	RMSE
KNN-Probabilistic	93.3333	0.0667	0.2582
KNN	86.6667	0.1333	0.3651
Naive Bayes	76.1194	0.1726	0.2824
OneR	71.6418	0.5325	0.6139
ZeroR	64.1791	0.4619	0.4805

Table 1: Prediction Results of Classifiers.

Overall, KNN-Probabilistic model has better accuracy rate and error rates than the other classifiers used for comparisons. The test demonstrated that the hybrid mechanism of KNN and probabilistic method produced significantly improved results, compared with each of the KNN and Naïve Bayes classifiers.

V. CONCLUSION

The purpose of this study is to enhance the statistical fitness of the suggested model to deal with a KNN issue because of its computing strategy. The empirical distribution across the Profit and Loss class values in the k number of nearest neighbours can be computed by the KNN classifier. Due to the lack of data, the result is, however, far from satisfactory. The KNN classifier has a problem with underfitting because it can't handle sparse data generalisation outside of the immediate neighbourhood. On the issue of predicting the trends in stock prices, we have contrasted a hybrid KNN-Probabilistic model with four common algorithms. Our findings demonstrated that the suggested KNN-Probabilistic model outperforms the traditional KNN algorithm and other classification algorithms by a wide margin. The suggested model has a drawback because it uses a binary classification method.

This binary classification model's real output is a prediction score in two class. The score reflects how confident the model is that the provided observation belongs to either the Class of Profit or Class of Loss. The knowledge component for upcoming work entails converting binary classification to multiclass classification. The multiclass classification entails observation and analysis of more statistical class values than the two already present. To provide more detailed information about each class value, additional study will apply the probabilistic model to multiclass data. Five class labels with the names "Sell", "Underperform", "Hold", "Outperform", and "Buy" will be present in the newly created multiclass classification. For mapping purposes, we will translate "Sell" to -2, which denotes strongly unfavourable; "Underperform" to -1, which denotes moderately unfavourable; "Hold" to 0 to denote neutral; "Outperform" to 1 to denote moderately favourable; and "Buy" to 2 to denote strongly favourable.

VI. REFERENCES

[1] H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. M. Selim, and K. Muhammad, ‘‘A local and global event sentiment based efficient stock exchange forecasting using deep learning,’’ Int. J. Inf. Manage., vol. 50, pp. 432–451, Feb. 2020.

[2] W. Long, Z. Lu, and L. Cui, ‘‘Deep learning based feature engineering for stock price movement prediction,’’ Knowl.-Based Syst., vol. s164, pp. 163–173, Jan. 2021.

[3] Kalra S, Prasad JS. Efficacy of News Sentiment for Stock Market Prediction. Proc Int Conf Mach Learn Big Data, Cloud Parallel Comput Trends, Prespectives Prospect Com 2020. Published online 2020:491-496. doi:10.1109/COMITCon.2019.8862265

[4] Menon A, Singh S, Parekh H. A review of stock market prediction using neural networks. 2019 IEEE Int Conf Syst Comput Autom Networking, ICSCAN 2019. Published online 2019

[5] Kian-Ping Lim, Chee-Wooi Hooy, Kwok-Boon Chang, and Robert Brooks, Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity, The North American Journal of Economics and Finance.

[6] Yauheniya Shynkevicha, T.M. McGinnity, Sonya A. Coleman, and Ammar Belatreche, Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, Decision Support Systems, Vol. 85, http://linkinghub.elsevier.com/retrieve/pii/S0167923616300252, 2021, pp. 74–83.

[7] J. B. Duarte, L. H. Talero Sarmiento, and K. J. Sierra Juárez, "Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency," Contaduría y Administración, vol. 62, no. 4, pp. 1361–1376, Oct. 2020.

[8] Charles D. Kirkpatrick II and Julie R. Dahlquist, Technical Analysis: The Complete Resource for Financial Market Technicians (3rd Edition), Pearson Education, Inc., 2021.

Cite this article as :

Anjali Patil, Gayatri Padole, Akansha Sontakke, "Stock Trend Prediction Using KNN Algorithm", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 13-18, May-June 2023. Available at doi : https://doi.org/10.32628/IJSRST5231035
Journal URL : https://ijsrst.com/IJSRST5231035