

Stock Price Prediction using KNN and Linear Regression



Poornima S P, Priyanka C N, Reshma P, Suraj Kr Jaiswal and Surendra Babu K N

Abstract: Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. With the ever increasing amounts of data becoming available there is good reason to believe that smart data analysis will become even more pervasive as a necessary ingredient for technological progress. It is an important challenge for the people who invest their money to forecast the daily stock prices, which helps them to put money into stock market with credence by taking risks and also variations into considerations. In this paper, we are going to apply KNN method and linear regression for predicting the stocks. The performance of linear Regression model on the selected data set is better when compared to KNN algorithm technique. The stock holders can invest confidently based on the results obtained from the model.

I. INTRODUCTION

Machine learning is a branch of Artificial Intelligence that aims at enabling machines to perform their jobs skilfully by using intelligent software. The statistical learning methods constitute as a backbone of intelligence software that is used to develop machine intelligence. Because ML algorithms require data to acquire, the discipline must have connection with the discipline of database. Prediction of the stock prices is one of the most crucial topic in finance and business, which is used to forecast the daily stocks. We use two categories of supervised learning concepts that is Classification and Regression. By using the KNN and linear Regression, we are going to forecast the price of the daily stocks and compare which gives the optimum accuracy.

II. RELATED WORK

Forecasting the stock prices is very challenging and complicated process because movement of price just behaves like unusual and time variation. In recent years different researchers have used Machine Learning technique in stock market for trading decisions. Here, we will present a brief review of some researches.

[1].Stock market classification model using sentiment analysis on twitter based on hybrid Naive Bayes classifier. This study proposes Hybrid Naïve Bayer classifier (HNBC's) as a ML method for stock price classification. The outcome is instrumental for investors and companies whereby it will enable them to formulate their plans according to the sentimental of people. The proposed method has produced a significant corollary, it has attained with desired accuracy. There are some five steps of the model were presented that is data collection, filtration, determination of the polarity according to sentiments of people, classification by enhancing NBC's and ends with the performance and evaluation stage. [2]Stock Market Analysis and Prediction using Hadoop and ML. The big data technique by performing sentiment analysis of tweets and twitter and finding the correlation. Also, machine learning techniques are applied on the data of companies to predict the stock price of next day. Python code is used to perform the task and text editor used is Pycharm. Map-reduce technique is used to combine individual sentiment results and find a sentiment for each day. The big data technique is used to handle the large amount of data. All the four model, KNN, Random forest, neural network, and linear regression are used to predict the value of daily exchange. The predicted values showed that name of the four models were able to predict the values accurately. However the closest values were predicted by linear regression. The conclusion of their project was that the stock value change does not depend on the stock exchange index. It is dependent on the sentiments of social media. [3]Stock market prediction using a linear regression. The aim of the research study is to help the stock broken and investors for investing money in the stock market. Python 2.7 version is used for implementing linear, polynomial and RBF. The python scientific computing library numpy was used along with the data analysis library pandas in order to convert these CSV files into panda data frames. The stock price was predicted based on linear regression, polynomial and RBF regression using five variable. The linear regression gives the best result compare to another method stock exchange index. It is dependent on the sentiments of social media. [4]Machine learning in prediction of stock market indicators based on historical data and data from twitter sentimental analysis. This project use neural networks and support vector machine for forecasting future price for a long time. For, twitter sentimental analysis Naïve Bayes algorithm and dictionaries approach is used. Tweets are downloaded using Twitter API which allows to download 14500 tweets in one hour. For the stock market data the yahoo finance website, which provides opening and closing historical prices, as well as the volume for any given trading day. The code is written in javascript.

Manuscript published on 30 May 2019.

* Correspondence Author (s)

Mr. SurendraBabu K N, REVA University, Bangalore, India
Priyanka C N, C&IT, REVA University, Bangalore, India.
Suraj Kr Jaiswal C&IT, REVA University, Bangalore, India
Reshma P, C&IT, REVA University, Bangalore, India.
Poornima S P, C&IT, REVA University, Bangalore, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

The results shows that twitter analysis add some information and cannot say that quality of forecast changes significantly and also information from twitter donot allow significantly increase accuracy. The average rate 64.10% was achieved using support vector algorithm and neural networks for prediction. [5]Yarn price prediction using advanced analytics model. This project uses ARIMA and KNN models that are used to predict future prices. The conclusion of this is predicting the yarn price using the KNN model gives the maximum accuracy(97%) because the three important factors like historical data, most influencing attributes and predicted values of most influencing attributes are taken into considerations for yarn price prediction. And KNN values are predicted using the values predicted by seasonality and ARIMA model so that this becomes an integrated model. [6]Short term short price production using deep learning. This project uses MRP, LSTM and neural networks for prediction. The outcome of this project shows that, MRP has performed better than RSTM model in estimating short term prices. Neural networks has confirmed to be a good tool to predict a chaotic framework like stock market. [7]L1 regularized logistic regression for event driven stock market prediction. The algorithms that are used are L1 regularized logistic regression, decision tree and random forest for the prediction of future stocks. The anticipated results within a week after the event occurs shows superlative, which are frequently used to predict the short term trend of the stock. Result in short term stocks price forecast more accurate. The logical regression has the unsurpassed results. The decision tree works well on the training set but the performance on back test set is weak, which stipulates that this mode has poor robustness on datasets. Finally, random forest gives errors for both the training and the testing which are worsted than the L1. [8]Stock Market Prediction. In this project it uses support vector machine and principal component analysis and the result shows that, there are different methods can envisagesignals through ML algorithms and numerical method. PCASVM was applied to both to eradicate the wrong predictions and to determine what features are essential. Comparing to the uncomplicated techniques from SVM and developing to GASVM and PCASVM, the solution to the main problem and sub-issues is cost effective and shows promising results for a real prediction using recent data sets. [9]Stock Market Prediction using data mining techniques. This project proposes the use of machine learning techniques for evaluating data concerned to the stock market using a novel method for prediction of stock prices to minimize the risk of investment in a stock market. Natural language processing is used to derive content from a used articles and other relevant sources along with an ensemble learning model called random forest model and support vector machine. The forecasting problem of stock price is treated as a classification problem to make better decisions. The random forest model using a 1-gram model for text analysis produced an accuracy of 84.3% and on using a 2-gram model produced an accuracy of 86.2%. the linear support vector machine using 1-gram model and 2-gram model for text analysis produced predictions with an accuracy of 82.2% and 84.6%, while the nonlinear support vector machine produced predictions with an accuracy of 85.1% for both 1-gram and 2-gram models. They observed that the random forest model out performs the support vector machine by using the given data set. [10] A Machine Learning model for stock price prediction. The project

propounded is a machine learning model that combines particle swarm optimization(PSO) algorithm and LS-SVM for stock price prediction using financial technical indicators. The LS-SVM-PSO model convergence to the global minimum. Also, it has the ability to get over the overfitting problem which found in ANN, especially in case of fluctuations in stock sector. PSO-LS-SVM algorithm parameters can be adjusted easily. The presentation of the proposed model is superior to LS-SVM and compared algorithms. LS-SVM-PSO achieves the lowest error value followed by single LS-SVM, while ANN-BP algorithm is the worst one.

III. DATA PREPROCESSING

Data pre-processing provides clean and relevant data sets which then can be used in further steps like Classification and Regression. Processing data is mathematically transforming data from one form into another with the goal of amplifying pertinent information for traders. Using pre-processing, the unwanted data will be removed i.e, NaNs and also used scaling to normalize the data. We have taken 5 company datasets i.eTCS, Amazon, Yes bank, Bharat Electronics Ltd and HDFC bank from quandl.com. There are nearly 4566 datasets available for each company. Based on the Close price the prediction is done. The parameters i.e, open, high, low are known for input variable, whereas close is the output variable.

Table[1] Sample of Dataset (YES Bank)

Date	Open	High	Low	Close
02-04-2019	277.10	282.00	274.80	280.40
03-04-2019	283.20	285.90	272.30	273.75
04-04-2019	274.95	274.95	265.40	268.15
05-04-2019	270.80	274.15	264.00	267.10
08-04-2019	267.20	268.00	258.25	260.00

IV. METHODOLOGY

A. KNN algorithm

k-nearestneighbourtechnique is a supervised algorithm that is considered as easy to implement. KNN is also known to be the laziest learning that does not construct a model or functionality previously, but gives the closest k values of the training dataset that have the topmost equivalence to the test data. Then, a majority vote is achieved among the selected k values is to direct the class label and then assigned it to the test data.

The following steps for predicting of stock prices is calculated using KNN are

1. Here, the no of k values are taken as k=15
2. Then, 70% of training data and 30% of testing data will be taken.
3. We will calculate the Euclidean distance between the training and the testing dataset.

- Classify all training samples according to the distance values.
- Take a majority vote for the class label of KNN and assign it as a predicted value of the testing data.

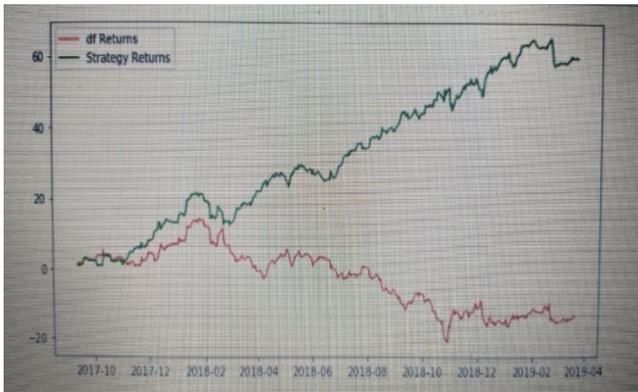


Fig.1 Result of KNN Algorithm

B. Linear Regression

Linear Regression deals with the relationship between a dependent variable y and one or more independent variables that are denoted by x.

$$Y = a + bx$$

Y is the estimated value

B is the slope of the line

X is the co-efficient or independent variable

a is the y-intercept

Regression performs operations on dataset. Model Selection is used to train and test the given data. First, we divide the data into 2 types of training and testing. And we apply the training section for starting analysis and defining the model 70% data used for training and 30% data used for testing purpose. prediction is made based on the accuracy of the test dataset.



Fig.2 Result of Linear Regression

V. RESULTS

The outcome of these 2 techniques have been compared on the bases of the Confidence value. By using R² (Coefficient of determination) we are finding the accuracy. KNN-Algorithm shows 63% of accuracy, whereas in Linear regression 98% of accuracy has been shown for daily stock prices.

VI. CONCLUSION

The aim of our project is to help the stock brokers and investors for investing money or stocks. The prediction plays

a very important role in stock business which is complicated and challenging process due to dynamic nature of stock market. As per the discussed works above, predictions of the stock prices based on KNN algorithm and linear regression. We compare these two methods on bases of confidence value and analyzed that linear regression provide best result compare to KNN model.

REFERENCES

- R. Iacomini, "Stock Market Prediction," 2015 19th International Conference on System Theory, Control and Computing (ICSTCCO), October 14-16, Cheile Gradistei, Romania.
- S., Kamath, "Stock Market Analysis", Master's Projects, pp. 326, 2012.
- P. Domingos, "A Few Useful Things to Know about Machine Learning," Communications of the ACM, Vol. 55 No. 10, October 2012.
- Gharehchopogh, F.S., & Khaze, S.R (2012), Data Mining Application for Cyber Space Users Tendency in Blog Writing: A Case Study. International Journal of Computer Applications, 47(18), 40-46.
- Berry, M. J., & Linoff, G. S. (2004). Data Mining Techniques: for marketing, sales, and customer relationship management. Wiley.com.
- S Abdulsalam Sulaiman Olaniyi, Adewole, Kayode S., Jimoh, R. G, "Stock Trend Prediction Using Regression Analysis- A Data Mining Approach", ARPN Journal of Systems and Software Volume 1 No. 4, JULY 2011, Brisbane, Australia.
- M. Ballings, D. VandenPoel, N. Hespeels, R. Gryp Evaluating multiple classifiers for stock price direction prediction Expert Syst Appl, 42 (20) (2015), pp. 7046-7056.
- R. Dash, P.K. Dash A hybrid stock trading framework integrating technical analysis with machine learning techniques J Finance Data Sci, 2 (1) (2016).
- E.A. Gerlein, M. McGinnity, A. Belatreche, S. Coleman Evaluating machine learning classification for financial trading: an empirical approach Expert Syst Appl, 54 (1) (2016), pp. 193-207
- J. Patel, S. Shah, P. Thakkar, K. Kotecha Predicting stock market index using fusion of machine learning techniques Expert Syst Appl, 42 (4) (2015), pp. 2162-2172
- K. Kim Financial time series forecasting using support vector machines Neurocomputing, 55 (1-2) (2003), pp. 307-319
- R.K. Nayak, D. Mishra, A.K. Rath A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices Appl Soft Comput, 35 (1) (2015), pp. 670-680
- S. Barak, M. Modarres Developing an approach to evaluate stocks by forecasting effective features with datamining methods Expert Syst Appl, 42 (3) (2015), pp. 1325-1339
- C.-Y. Yeh, C.-W. Huang, S.-J. Lee A multiple-kernel support vector regression approach for stock market price forecasting Expert Syst Appl, 38 (3) (2011), pp. 2177-2186
- Y. Xiao, J. Xiao, F. Lu, S. Wang Ensemble ANNs-PSO-GA approach for day-ahead stock e-exchange prices forecasting

AUTHORS PROFILE



Mr. Surendra Babu K N, completed BE (CSE) from HMSIT, Tumakuru, MTech (Computer Engineering) from SJCE Mysore and pursuing PhD in School of Computing and Information Technology, REVA University. He has 12 years of Teaching experience and 2 years of Industry experience as a Software Engineer. Currently he is serving at School of Computing and Information Technology, REVA University, Bengaluru as Assistant Professor. He has guided many BE Projects. His research area is Machine learning and Deep Learning, particularly analyzing and developing predictive models to predict retail sales in retail shops.



Stock Price Prediction using KNN and Linear Regression



Miss Priyanka C N
B.Tech. Final year(Computer Science)
REVA University
Bangalore, India
Area of interest-machine learning and data science



Mr Suraj Kr Jaiswal
B.Tech. Final year (Computer Science)
REVA University
Bangalore, India
Area of interest-web development



Miss Poornima S P
B.Tech. Final year (Computer Science)
REVA University
Bangalore, India
Area of interest-web development and machine learning



Miss Reshma P
B.Tech. Final year (Computer Science)
REVA University
Bangalore, India
Area of interest-machine learning and data science