WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Price Prediction Using Machine Learning

Nalawade Sakshi[1], Mule Sarita[2], Adsare Monika[3], Dr.Gunjal S.D.[4]

Department of Computer Engineering, Jaihind College of Engineering, Kuran, Pune 410511.

A B S T R A C T

The stock market or stock market is one of thethe most complex and sophisticated methods of doing business.Small goods, trading firms, banking sector, it all dependsin the same body for the distribution of revenue and risk; a more complex model However, this text proposesuse the machine learning algorithm to predict the future price ofshared resources exchange using open source libraries as wellpre-existing algorithms to help create this business format fromunpredictable. We will see how easy this isapplication will bring acceptable results. Result basedcompletely in numbers and occupies a large number of axiomsthat may or may not follow in the real world as the time offorecast.

Keywords:basicconcepts, data analysis, fundamental, implementation, linear regression, stock market, supervised machine learning

1. Introduction

The stock market is one of the oldest ways in which athe average person trades stocks, invests and profitssome money in companies that sell part of itin this court. This program turns out to be a possibilityan investment plan if done wisely. However, the price as wellThe liquidity of this scene is not very predictable and this is truewhen we bring in technology to help us. Machine learningis one of those tools that helps us get what we want. Ithe next 3 paragraphs will explain the key brieflyparts of this text:The stock market, as we know, is a very important tradea platform that affects everyoneindividually and nationally[2]. The basic principle is simple, companies will writetheir shares in companies as they are called smaller assetsshares. They do this to raise money for the company. Athe company calculates its shares at a price called IPO or initialcommunity contribution. This is the price of the companysells stocks and raises money. After that these are sharesyou do not own and you can sell them at any price atbuyer in stocks like BSE or BombayStock Exchange. Traders and buyers continue to sell theseshares for their price, but the company can only keepmoney received during the IPO.This is an ongoing hopehare without getting any extra profit, it translateson a certain increase in the value of the stock after eachprofitable transactions.However, if the company produces morestocks with low initial public offering, market price ofexchanges fall and traders lose. This is exactly whatoccurrence is a cause for fear of people who have itwhen investing in the stock market and the reason for the collapseand a brief rise in stock prices.

2. Literature Survey

Author of [1] discusses Web application for resume screening. Use of NLP pipeline. Text Extraction is done using sections based segmentation. Semi supervised learning to train machine learning model. The system has some drawbacks. The web application shows results at the recruiter side in the form of ranking. One of the feature can be described as Resume is only matched to those job openings which they are interested in and have applied for.

The author of [2] makes use of OCR (Optical Character Recognition) Feature Extraction of Principal Component Analysis. Decision tree classification algorithm is used. But the system works only on Urdu Text. Paper [3] mentions that Conventional SVM can be optimized for text classification. The

conventional SVM is optimized by selecting features using the entropy. The author of [4] compares three classification algorithms. Results show that support vector classifiers with the TF/IDF feature shows more accuracy than naive bayes and KNN.

The author of [5] presents the web document classification based on fuzzy k-NN network, in the process of classification, TF/IDF is adopted for selecting features of document. The results show that classification performance is better than k-NN and SVM, but the speed of classification is bit slow than KNN.

3. Proposed System

2.PREDICTION MODEL

A. Data Analysis Stage

In in this section, we will analyze the available raw data as wellread to see the appropriate qualities forpredicting our selected marker.Now the data that we will use for our system is taken from itwww.quandl.com, a world-class data delivery platform.In this section, we will analyze the available raw data as wellread to see the appropriate qualities forpredicting our selected marker.Now the data that we will use for our system is taken from itwww.quandl.com, a world-class data delivery platform.GOOGL Database Database from WIKI and can beextracted from quandl using the "Wiki / GOOGL" token. Weextracted and used for approximately 14 years of data.Character set of data includes:Open (opening price of shares)High (the highest possible price in a given period) Low (the lowestPossible value for a particular period) Close (closing value ofto share)

Volume (trading price during the day)

Adj. Open
Adj. Up
Adj. Close } Adj. Adjusted prices are lower for more attributes
Adj. Volume

We select the "Close" attribute as our label (The variable ofthey will predict) and then use "Adj. Open, Adj. High, Adj. Close,Adj. Low and Adj. Volume "to remove those featureswill help us better predict the outcome.It should be noted that we use values included in green, fromthese figures have been processed and do not contain standard datacollection errors.

Defines a set of features that we will use:

• Closure: It is an important source of information, as it weredetermines the price to open the market the next day againthe expected dose of the day.

• HL_PCT: this is a feature found in:

$$HL\_PCT = \frac{\text{Adj. High} - \text{Adj. Low}}{Adj.\,Close} \times 100$$

We use percentage conversion, as this helps us reducenumber of jobs, but to keep net informationaffected. High-Low is a good feature because it helps uscreate OHLCV chart shape.

• PCT_change: this is also a feature detected,defined by:

$$PCT\_Change = \frac{Adj.Close - Adj.Open}{Adj.Open} \times 100$$

We do the same with Open and Close as High andDown, as both are very consistent with our predictive modeland help us reduce the number of obsolete jobs like thiswell.

• Capacity: This is the most important decision-making processparameter as the sold volume is very accurateimpact on future stock value than any other factor. Thenwe will use it as our own.We successfully analyzed the data and extracted I useful information we will need in the classifier. Thisa very important step and should be treated with the utmost care.Lack of information or minor error in the acquisition ofuseful information will lead to an errorprediction model and ainefficient separator.In addition, the released features are very specific totheme used and will vary from topic to topic.Normal performance is possible if, and only if, the data of anotherthe title is collected in the same order asprevious article.

B. Training and testing stage

Training and Assessment Phase In this section, we will apply what we doextracted from our data and applied to oursmachine learning model.We will recommend SciPy, Scikit-learn and Matploliblibraries in python to edit our model, train them with itfeatures and labels we have released and reviewedwith the same data.We will first process the data to create data that includes:Submissions are labeled label

percentage of thatwants to predict.

• Dataframe format is converted to Numpy systemformat.

• All NaN data values were deleted before being sent todivides into categories.

• Data is measured in such a way that at any X value,

$X \in [-1,1]$

• Data is divided into test data and training databy type, i.e. label and feature

Now the data is ready to be included in the separator. We willuse a simple separator, that is, a vertical loop,described in the Sklearnlibrary of the Scikit-lear package.Wechose this separator because of its simplicity and because itfulfills our purpose in the right way. The drop line is aa widely used method of data analysis and prediction. Itit actually uses key elements to predict relationshipsamong other variables based on their dependence on othersfeatures. [9] This type of prediction is known as surveillancemachine learning. Supervised learning is the way to do itwe include tagged data, that is, features matchedon their tags. Here we train you to divide in such a wayreads patterns what are the combinations of symbolsresults are in which label.Here, in our case, the divider sees the features and easilylook at their label and remember it. Remember I a combination of features and their labelour case is the stock price a few days later. Then move onand find out which model is followed by features in orderproduce the appropriate label. This is how it is monitored machine learning works [10]. Supervised testmachine learning, we put a combination of features ina trained section and check the output of the splitter withreal marker. This helps us to determine our accuracydivides into categories. Most importantly for our model. Separatorwith less than 95% accuracy is useless.Accuracy is the most important factor in a machine learning model.It is necessary to understand what it isaccuracy also means how to increase its accuracy for the nexttitle below.

4.Results

When the model is ready, we use a template to find itthe results we want in any way we want. In our case, we will trygraph of our results (fig. 1) according to our needswhich we discussed earlier in this article.

Fig. 1. Graph showing stock price of GOOGL from year 2005 till July 2018. Red is the line representing given data and blue is representing the forecasted or the predicted value of stock

It should be according to our needs, and as mentioned above, aless than 95% accurate model is realisticuseless. There are some common methods of calculationaccuracy in machine learning, some are as follows:

• R2 model value.

• Fixed an amount of R2

• RMSE value

• Confusion matrix for separation problems.

Fig. 2. Graph showing the exact amounts of predicted values.

Accuracy is a part that every machine learnsthe engineer is always committed to providing. Aftermodel built, there is endless effort to adjust the model to get the most accurate resultsThere are very common and easy ways to increasethe effectiveness of the model and has been discusses above.

However, let's look at some of the common ways to improveMachine learning algorithm:

• Unlimited preparation

o Decent gradient

o Newton Road

o Group reading

o A decent stochastic gradient

• Limited performance

o Lagrange duality

o SVM in the first two forms

o Limited methods

Many machine learning problems, in the end,development problems, when we reduce workbelow certain limits.

Conclusions

Machine learning as we have seen so far, is a very powerful tool and therefore avoidable, it has a great application. So far we have seen that machine learning relies heavily on data. Therefore, it is important to understand that data is quite valuable, and as simple as it may seem, analyzing the data is not an easy task.Machine learning has found an amazing application and has evolved further into deep learning and neural networks, but the basic idea is pretty much the same for all of them.This document provides a fluid view of how to implement machine learning. There are various ways, methods, and techniques available to manage and solve various problems, in different imaginable situations. This document is limited to supervised machine learning only and seeks to explain only the fundamentals of this complex process.

Acknowledgements

This article is written on the basis of the type of authorproject during his 2018 graduation studies after a majordirection of Neha Agarwal (Asst. Prof.) AUUP, Noida.The author of this article, does not claim any rights to any of thesealgorithms, codes, data, formulas used, definitions, how to solve problems, such as its assets.

REFERENCES

[1] ANDREW MCCALLUM, KAMAL NIGAM, JASON RENNIE, KRISTIE SEYMORE "A MACHINE LEARNING APPROACH TO BUILDING DOMAIN-SPECIFIC SEARCH ENGINE", IJCAI, 1999 - CITESEER

[2] YADAV, SAMEER. (2017). STOCK MARKET VOLATILITY - A STUDY OF INDIAN STOCK MARKET. GLOBAL JOURNAL FOR RESEARCH ANALYSIS. 6. 629-632.

[3] MONTGOMERY, D.C., PECK, E.A. AND VINING, G.G., 2012. INTRODUCTION TO LINEAR REGRESSION ANALYSIS (VOL. 821). JOHN WILEY &SONS.

[4] DRAPER, N.R.; SMITH, H. (1998). APPLIED REGRESSION ANALYSIS (3RD ED.). JOHN WILEY. ISBN0-471-17082-8.

[5] ROBERT S. PINDYCK AND DANIEL L. RUBINFELD (1998, 4H ED.). ECONOMETRIC MODELS AND ECONOMIC FORECASTS

[6] "LINEAR REGRESSION", 1997-1998, YALE.UNIVERSITY HTTP://WWW.STAT.YALE.EDU/COURSES/1997- 98/101/LINREG.HTM

[7] AGARWAL (JULY 14, 2017). "INTRODUCTION TO THE STOCK MARKET". INTELLIGENT ECONOMIST. RETRIEVED DECEMBER 18, 2017.

[8] JASON BROWNLEE, MARCH 2016, "LINEAR REGRESSION FOR MACHINE LEARNING", MACHINE LEARNING MASTERY, VIEWED ON DECEMBER 2018, HTTPS://MACHINELEARNINGMASTERY.COM/LINEARREGRESSION-FOR-MACHINE- LEARNING/

[9] GOOGLE DEVELOPERS, OCT 2018, "DECENDING INTO ML: LINEAR REGRESSION", GOOGLE LLC, HTTPS://DEVELOPERS.GOOGLE.COM/MACHINELEARNING/CRASH-COURSE/DESCENDING-INTO-ML/LINEARREGRESSION

[10] FIESS, N.M. AND MACDONALD, R., 2002. TOWARDS THE FUNDAMENTALS OF TECHNICAL ANALYSIS: ANALYSING THE INFORMATION CONTENT OF HIGH, LOW AND CLOSE PRICES. ECONOMIC MODELLING, 19(3), PP.353-374.

[11] HURWITZ, E. AND MARWALA, T., 2012. COMMON MISTAKES WHEN APPLYING COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING TO STOCK MARKET MODELLING. ARXIV PREPRINT ARXIV:1208.4429