Pune Institute of Computer Technology
Dhankawadi, Pune

A SEMINAR REPORT
ON

Analysis of Machine Learning algorithms on stock prices

SUBMITTED BY

Pratik Ghodke
Roll No. 31218
Class TE-2

Under the guidance of
Prof. R. A. Kulkarni

DEPARTMENT OF COMPUTER ENGINEERING
Academic Year 2019-20

DEPARTMENT OF COMPUTER ENGINEERING

Pune Institute of Computer Technology
Dhankawadi, Pune-43

CERTIFICATE

This is to certify that the Seminar report entitled

"Analysis of Machine Learning algorithms on stock prices"

Submitted by
Pratik Ghodke Roll No. 31218

has satisfactorily completed a seminar report under the guidance of Prof. R. A. Kulkarni towards the partial fulfillment of third year Computer Engineering Semester II, Academic Year 2019-20 of Savitribai Phule Pune University.

Prof. R. A. Kulkarni
Internal Guide

Prof. M.S.Takalikar
Head
Department of Computer Engineering

Place:
Date:

ACKNOWLEDGEMENT

I sincerely thank our Seminar Coordinator Prof. B.D.Zope and Head of Department Prof. M.S.Takalikar for their support.

I also sincerely convey my gratitude to my guide Prof. R. A. Kulkarni, Department of Computer Engineering for her constant support, providing all the help, motivation and encouragement from beginning till end to make this seminar a grand success.

Contents

List of Tables

List of Figures

Abstract

Brokerage corporations, banking sector depend on stock market to make revenue. It is very risky to predict stock prices. In existing methods, the stock price forecasting is implemented using ANN, data mining also using machine learning but fuzzy forecasting methods try to forecast prices by determining patterns only. The main objective is to predict the future worth of stock prices using existing supervised machine learning algorithms and compare the efficiency of algorithms. I'm going to use existing stock market data to train machine learning algorithms. I'm going to use Linear regression, K-nearest neighbor, random forest regression, SVM. Finally, a simple trading model is established to study the performance of the proposed prediction algorithm against other benchmarks.

Keywords

Data analysis, Stock market, Algorithms, Efficiency comparison, Linear Regression.

1 INTRODUCTION

Prediction of stock trend has long been an interesting topic and is extensively studied by researchers from different fields. A correct prediction can lead to huge profits for the seller and the broker. The stock prices are unpredictable however, if analysis is done wisely, future stock price can be predicted. When we say prediction, the first method that comes to our mind is plot the graph and extrapolate graph to predict values. we know stock prices are not linear, though with neat observations, we can obtain the relation with time and stock prices. To obtain relation, we are going to use Machine Learning algorithms.

In this report, stock prices are predicted using continuous machine learning algorithms. i.e. Linear Regression, SVM, k-nearest neighbor, Random forest. And the efficiency of algorithms is being compared.

2 MOTIVATION

With a successful model for stock prediction, we can gain insight about market behavior over time, spotting trends that would otherwise not have been noticed. Machine Leaning would be efficient method to forecast the stock prices.

2.1 Financial

Many people are interested in financial market but they need guidance to invest their money. With successful implementation, we can provide guidance to invest money with minimal human intervention.

2.2 Accuracy

Investors always look for better and more accurate stock price forecasting model. Another motivation is to provide more accurate machine learning model to predict stock prices.

2.3 News Channels

Many news channel broadcast stock market news along with their predictions if stock prices will rise or fall. so this can be helpful for news channels.

3 LITERATURE SURVEY

In last few decades, forecasting of stock returns has become an important field of research. In many cases, researchers have tried to establish linear relationship between the input variables and stock returns. It involves an assumption that the stock prices data obtained from past few years has some predictive relationship to the future stock returns. In order to obtain such relationships, researchers have focused on Machine Learning techniques to forecast stock prices. Significant work has been done throughout the world in this field.

3.1 Deep Neural network

H.Gunduz, Z. Cataltepe and Y. Yaslan predicted stock prices using deep neural network techniques. M. Billah, S. Waheed and A. Hanifa suggested improvements to stock predication using neural networks through the use of a training algorithm which they designed on their own.

3.2 Data mining approach

Nghiem Van Tinh, Nguyen Tien Duy proposed the methods by using k-means clustering algorithm. Most of the fuzzy forecasting methods based on fuzzy time series used the static length of intervals, i.e. the same length of intervals. The drawback of the static length of intervals is that the historical data are roughly put into intervals, even if the variance of the historical data is not high.

3.3 Machine Learning Approach

Anass Nahil proposed a new method on stock market prediction which will help many investors to invest their money in right time by which they will get more benefit in near future. Their proposed method was about support vector machine (SVM). It is a popular tool in time series forecasting for the capital investment industry. This machine learning technique which is based on a discriminative classifier algorithm, forecasts more accurately the financial data. In this paper, we'll compare efficiency of all the machine. This paper compares efficiency of continuous machine learning algorithms on available dataset.

4 CHALLENGES

4.1 Needs large dataset

To train the model, test on the data, to check the accuracy and forecast the stock prices we need large dataset so that the result would be accurate.

4.2 Enhanced model

As we know, there's been a huge research done on this topic, so we have to come up with better and enhanced solution than the previous ones .

5 DATA ANALYSIS PHASE

In this stage, we shall look for the available raw data and keep the suitable attributes and get rid of unnecessary attribute. The dataset is obtained from www.quandl.com.

The attributes of dataset include:

- Open (opening price of stock)
- High (Highest possible price at an instance of time)
- Low (Lowest possible price at an instance of time)
- Close (Closing price of stock)
- Volume (Total times traded during a day)
- Ex-dividend
- Split ratio
- Adj. open
- Adj. high
- Adj. low
- Adj. close
- Adj. volume

It repetitively divides the working area(plot) into sub part by identifying lines.Operations are carried with optimization.Effciency reduces with increase in dataset.

Correlation Graph:

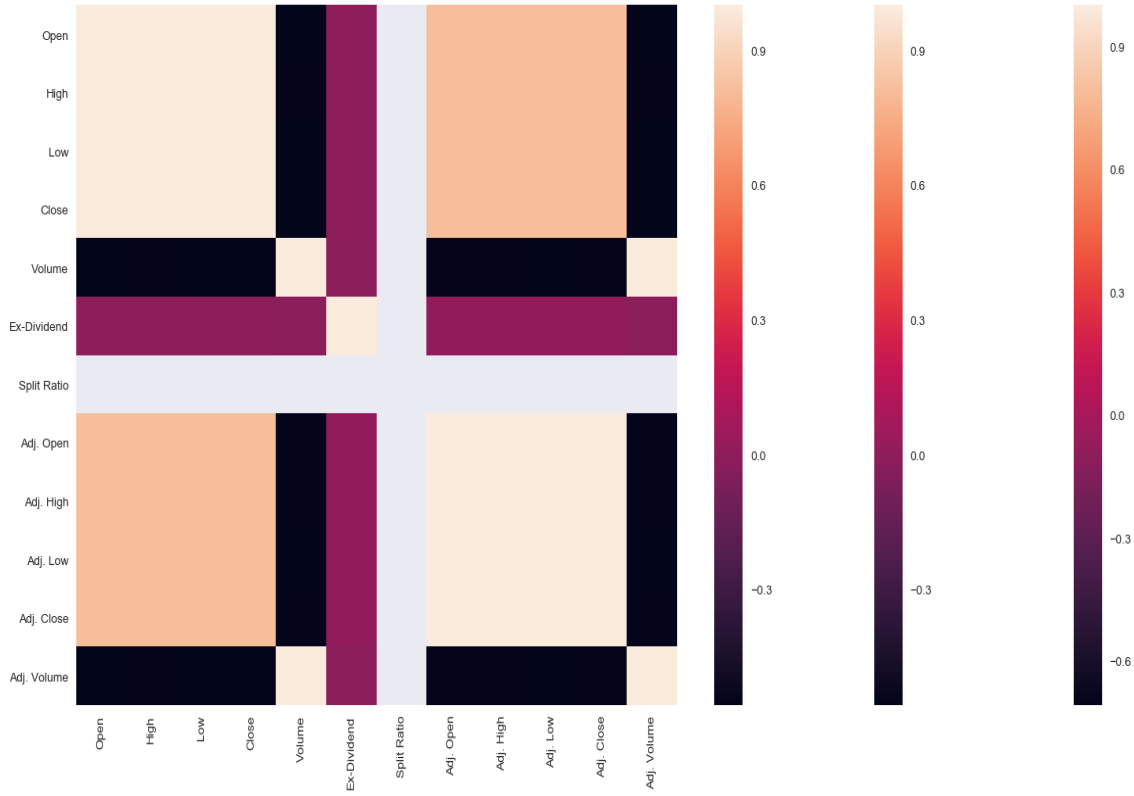

Figure 1: correlation before processing dataset

The above graph represents correlation between the attributes. Similar color represents the attributes are related. By observation we can conclude:

- Split ratio has constant value irrespective of change in values in other attributes.
- Open, High, Low, Close attributes are correlated Hence we don't need all of them.
- Also Adj. open, Adj. High, Adj. low, Adj. Close are correlated hence we would not use them directly.

To eliminate unnecessary variables and retain the information, we are defining new variables as:

- $\mathrm{HL}_P CT = \frac{Adj.high - Adj.low}{Adj.Close} \times 100$
- $\mathrm{PCT}_C hange = \frac{Adj.Close - Adj.open}{Adj.open} \times 100$

Finally, we are left with following attributes:

- Adj. Close
- HL_PCT
- PCT_CHANGE
- Adj. Volume

Correlation Graph:

Figure 2: correlation after processing dataset

From graph, we can see All attributes are independent of each other hence we are done with removing redundant attributes. Now we have to split data into train data and test data. Then we have to apply machine learning algorithms to train data and we have to test the efficiency using test data.

6 DISCUSSION ON IMPLEMENTATION AND RESULTS

6.1 Algorithms

6.1.1 Linear regression:

Linear regression is a supervised algorithm. It predicts dependent variable value from different attributes. We have implemented linear regression on the dataset.

6.1.2 K-nearest neighbor:

KNN works by finding distances between a query and all the examples in the data, selecting the specified number examples closest to the query, then averages the labels in case of regression.

6.1.3 Support vector machine:

It can be used in both classification and regression problems. Since we are dealing with regression problem. We are using existing implemented SVR algorithm in python.

6.1.4 Random forest regression

Random forest regression makes use of multiple decision trees. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

6.2 Accuracy comparison:

Table 1: Prediction comparison

S.No	Algorithm	Accuracy
1	Linear regression	0.9769
2	K- nearest neighbor	0.7514
3	Support vector machine	0.7915
4	Random forest regression	0.9817

6.3 stock price forecasting

Figure 3: stock price forecast result

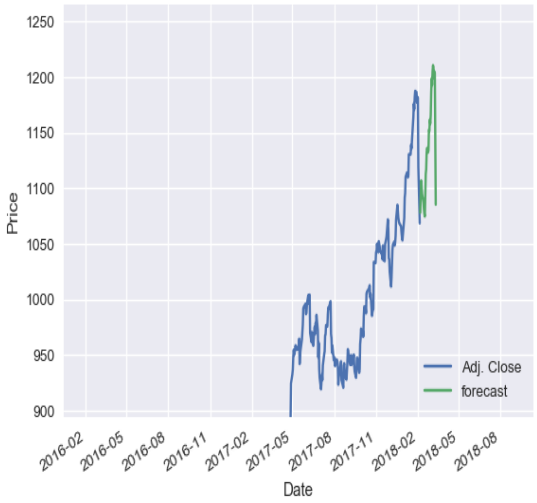

Figure 4: stock price forecast result

7 Conclusion and Future Enhancement

7.1 Conclusion

The Report summarizes the application of different machine learning algorithms for forecasting stock prices. In a study of machine learning algorithms such as Linear regression, SVM, KNN, Random forest regression to predict stock prices, we come to know that linear regression and random forest regression are working most efficiently for given dataset with an accuracy of 97.69% . and 98.17% respectively.

7.2 Future Enhancement

- Efficiency of an algorithms can be increased with further addition of dataset.
- Dependent variables can be increased in dataset to increase efficiency.
- In case of Linear regression, we can use methods such as R squared to obtain best fit graph.

References

[1] J. Chou and T. Nguyen, "Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine-Learning Regression," in IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3132-3142, July 2018.

[2] Shunrong Shen, Haomiao Jiang, and T. Zhang, "Stock Market Forecasting Using Machine Learning Algorithms," IEEE Transactions on Neural Networks, vol. 84, no. 4, pp. 21, 2015

[3] . K. Pahwa and N. Agarwal, "Stock Market Analysis using Supervised Machine Learning," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 197-200.

[4] J. Moody, et al., "Learning to trade via direct reinforcement," IEEE Transactions on Neural Networks, vol. 12, no. 4, Jul. 2001.

attach your review and visit log here.....

Document Information

Analyzed document	STC_REPORT_31218 (1).pdf (D68011215)
Submitted	4/13/2020 6:51:00 AM
Submitted by	Rekha Kulkarni
Submitter email	rakulkarni@pict.edu
Similarity	18%
Analysis address	rakulkarni.pict@analysis.urkund.com

Sources included in the report

SA	URL: Shubham_Yenchanwad_31380.pdf Fetched: 4/7/2020 2:18:00 PM	3
SA	URL: 31204_STCReport.pdf Fetched: 4/4/2020 12:54:00 PM	1
W	URL: https://sinarict.com/bqg/random-forest-regressor.html Fetched: 4/13/2020 6:53:00 AM	1
SA	URL: STC_Report_31217.pdf Fetched: 4/10/2020 12:06:00 AM	2
SA	URL: 31216_STC_Report.pdf Fetched: 4/5/2020 3:13:00 AM	1