

STOCK MARKET PREDICTION USING MACHINE LEARNING IN PYTHON

*Saloni Kumari and Pawan Kumar

Department of Computer Application Jain University Bangalore, India.

Article Received on 08/04/2022

Article Revised on 28/04/2022

Article Accepted on 18/05/2022

***Corresponding Author**

Saloni Kumari

Department of Computer
Application Jain University
Bangalore, India.

ABSTRACT

Stock market is one among them which needs the prediction future market to invest in the new enterprise or to sell their existing shares to get profit. This need the efficient prediction technique which studies the previous exchanges of stock market and gives the future prediction

based on that. This article proposed the prediction system of stock market price based on the exchange takes place in previous scenario. The system studies the diverging effect of market price of product in a particular time gap and analyse its future trend whether it's loss or gain. During the system of thinking about diverse strategies and variables that should be taken into account, we observed out that strategies like random forest, Support vector machine and regression algorithm. Support vector regression is a beneficial and effective gadget gaining knowledge of approach to apprehend sample of time collection dataset. The data collected for the four years duration which was accumulated to get the expecting prices of the share of the firm. It can produce true prediction end result if the fee of essential parameters may be decided properly. It has been located that the guide vector regression version with RBF kernel indicates higher overall performance while in comparison with different models.

KEYWORDS: Stock market, prediction, support vector machine, Random forest, Regression.

1. INTRODUCTION

Basically, quantitative buyers with a number of cash from inventory markets purchase shares derivatives and equities at a reasonably-priced fee and afterward promoting them at excessive

fee. The fashion in a inventory marketplace prediction isn't a brand new aspect and but this difficulty is saved being mentioned with the aid of using numerous organizations. There are kinds to examine shares which buyers carry out earlier than making an investment in a inventory, first is the essential evaluation, on this evaluation buyers study the intrinsic cost of shares, and overall performance of the industry, economy, political weather etc. to determine that whether or not to make investments or not. On the alternative hand, the technical evaluation it's far an evolution of shares with the aid of using the method of analyzing the data generated with the aid of using marketplace activity, together with beyond fees and volumes. In the current years, growing prominence of system getting to know in numerous industries have enlightened many buyers to use system getting to know strategies to the field, and a number of them have produced pretty promising results. This paper will expand a monetary information predictor software wherein there could be a dataset storing all historic inventory fees and information could be dealt with as schooling units for the software. The predominant cause of the prediction is to lessen uncertainty related to funding selection making.

2. RELATED WORK

Recently, quite a few exciting paintings has been achieved withinside the region of referring to Machine Learning Algorithms for comparing fee styles and predicting inventory fee.

^[1]Employed the assist of vector system layout to assume the monetary communique expense. By using this calculation, they have got predicted the precision result. SVM is hired right here to fathom the direct certain quadratic programming issue. Employing the SVM the carry one-of-a-type solutions for the issue.

Mehak Usmani^[2] the paintings achieved on this paper to assume Karachi sign up exchange. This paper gives a neural tool and Support vector device layout is Objective is to forecast the market execution of Karachi sign up exchange. This paper principle motive for modern is virtually to foresee the Karachi sign up exchange. The effects of the the usage of Gradient-descendant to reap to textbook worth.

Tejas Mankar^[3] on this paper expectation relying at the social feelings using tool gaining knowledge of. The assessment given with the useful resource of the usage of human beings in fashionable nearly the company. Based on the positive and unfavorable commentary of open nearly the company they're foreseeing the financial exchange. Utilizing assumption studies

on the tweets accumulated using the Twitter API and furthermore the surrender estimations of diverse stocks, we try to carry collectively a framework that gauges the stock fee development of diverse organizations.

El Mehraz^[4] artwork accomplished on this paper forecasting the securities any other expectation using Hybrid approach which consolidates Support Vector Regression. 332 Reshma R et al. / Stock Market Prediction Using Machine Learning Techniques Our intention in this survey artwork is to endorse a 1/2 of and 1/2 of approach that joins bolster vector relapse (SVR) and Hedrick-Prescott channel (HP), for development the forecast of listing price with the useful resource of eating analyzing the recorded records of using our encouraged prototypical.

Haiying Huang^[5] on this paper indicates the health of a device that consolidates bolster vector decline and Fourier exchange, for forecasting the stock fee via getting to know the wonderful gen. Fourier exchange is implemented for clam or separating, and the help vector relapse is for model preparing. Their proposed form is a unusual prescient gadget for stock forecasts with in the monetary market.

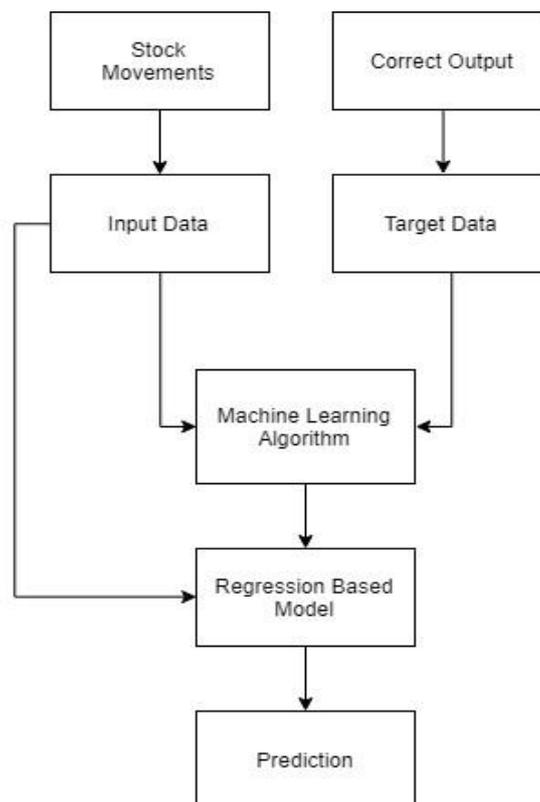
Nonita Sharma^[6] on this paper to anticipate the future economic exchange file esteems relying on verifiable facts. The test assessment is predicated upon on supportable facts of ten years of lists, specifically, CNX Nifty and SP Bombay Stock Exchange Sensex from Indian economic exchanges. In this paper they've implemented LS help that is implemented as a instruction misfortune cappotential to beautify the blunder gauges and consequently making it increasingly charming to take care of the forecast issue.

This paper is offerings throughout the assignment of looking ahead to future estimations of the financial exchange records. The studies paper "Predicting inventory and inventory fee index motion the usage of Trend Deterministic Data Preparation and system gaining knowledge of techniques" written with the aid of using J. Patel, S. Shah, P. Thakkar, and K. Kotecha for the "Expert Systems with Applications" worldwide magazine validated a manner to apply fashion deterministic facts to are expecting inventory fee motion^[3] They performed experiments in the usage of 10 technical signs' alerts as inputs, then they use prediction fashions to are expecting whether or not the inventory will move up or down withinside the coming 10 days, Technical evaluation signs encompass SMA, EMA, Momentum, Stochastic SK, Stochastic SK, MACD, RSI, etc. The prediction fashions they have got used encompass

ANN, SVM, Random Forest, and Naive Bayesian fashions. The version outputs “up” or “down” motion alerts. Experiments have proven random wooded area scored the very best overall performance with 83.56curacy with their inputs.

B. Wanjawa and L. Muchemi validated the ability in predicting inventory fees the usage of ANN, as proven withinside the studies paper “ANN Model to Predict Stock Prices at Stock Exchange Markets”.^[4] They used 70% of the education facts to are expecting the inventory fees for the following 60 days. Through optimizations, they have been capable of are expecting the real final fees inside 0.71% suggest absolute percent error (MAPE), with the very best variance -3.2% amongst all the sixty two days. This validated a excessive ability for the usage of system gaining knowledge of to appropriately are expecting inventory fees. This is one of the key additives in our software in which algorithms should be designed to have excessive accuracy, such that the platform might be beneficial for retail investors.

3. System Architecture



4. Proposed System

LSTM: - Long short-time period memory (LSTM) is an synthetic recurrent neural network (RNN) architecture^[1] used withinside the subject of deep learning. Unlike fashionable feedforward neural networks, LSTM has comments connections. It can procedure now no

longer most effective unmarried facts points (which include images), however additionally whole sequences of facts (which include speech or video). LSTM networks are well-ideal to classifying, processing and making predictions primarily based totally on time collection facts, considering the fact that there may be lags of unknown period among essential activities in a time collection. LSTMs have been evolved to address the vanishing gradient hassle that may be encountered whilst education conventional RNNs.

4.1. Data Collection: Data series is the number one step and preliminary module for the project. Here it offers with amassing the proper dataset for the project. They dataset has been accrued from quant. The proper dataset is accrued to be expecting the inventory rate. Our dataset accrued from the ancient facts.

4.2. Feature Extraction: Feature Extraction is the device to choose the important function to expect the stock rate. Because the raw information is gathered they'll have too many function, but best the important function we need to be awaiting the stock rate. So using the random forest set of regulations deciding on the important function. The random forest set of regulations classify the function similar to the tree form and installation the important function on one aspect and unwanted function on unique aspect. The important function list are displayed with the frequency rate. Based on the frequency rate the important function are arranged. In that dataset they important function are Close, Open, High, Low, Volume and Adj quantity the ones are the important function to be awaiting the stock market.

4.3. Data Split: The dataset that use to expect the stock price is discover to interrupt up into educate records and the check records. The records is typically cut up into education records and finding out records. The education set consists of a identified output and the model learns on this records at the manner to be generalized to unique records later on. They cut up the educate records proposition is extra the check records. The educate may be in 70 percentage, in which the check records may be in 30%. The education dataset is used to educate the model at the same time as the check records is used to expect the accuracy of the model. The records cut up achieved via pass validation.

4.4. Trained Data Result: The device of education the information the usage of SVR and linear regression. The information is professional with the resource of the usage of help vector regression and deliver the predicted stop end result. Using SVR education the stop end

result information the prediction will deliver appropriate accuracy. The stop end result will predicted the open stock price for the day.

5. Implementation Methodology

Random woodland set of rules have a hard and fast of guidelines is used for function extraction. Random forests or random choice forests are an ensemble studying method for class, regression and distinct responsibilities that carry out with the resource of the usage of constructing a large number of choice timber at schooling time and outputting the elegance that is the mode of the commands for class or mean prediction for regression of the character timber. Algorithm 1 Feature Extraction Procedure Input: Data set as CSV file. Output: Selected the vital characteristic Listed.

- 1: Read the dataset.
- 2: Import RandomForestClassifier from sklearn.ensemble.
- 3: Assign the RandomForestClassifier to neighborhood variable model.
- 4: Train Rfc =(nestimators=100,randomstate=0,njobs=-1).
- 5: Create clf =sfm(clf, threshold=0.15)
- 6: Assign sfm to clf
- 7: Get the vital characteristic Using the random woodland set of rules the statistics has been break up. They break up the statistics for schooling and trying out the usage of the move validation. They are match in Random woodland set of rules and break up the usage of move validation. The break up the teach statistics in 70 percent and check in 30 percent. Algorithm

2 Data Train, Test Split Procedure

- 1: Read the dataset.
- 2: Import RandomForestClassifier from sklearn.ensemble.
- 3: Create Xtest, Xtrain, Ytest, Ytrain
- 4: Create featable.
- 5: Assign Date,open,near interior featable.
- 6: Assign traintestsplit(dfx, dfy, testsize=0.2, rs=0).
- 7: X,Y are match the usage of Randomforestclassifier.
- 8: Create clf variable and match randomforest in that variable
- 9: For function in featable do
- 10: Print function
- 11: Get the teach statistics and check statistics This set of rules is used create the unbiased statistics set X and keep the statistics withinside the variable dates. Create the based statistics

set y and keep the statistics withinside the variable costs. Both may be finished through appending the statistics to every of the lists. The unbiased statistics set we need handiest the day from the date, so use the break up characteristic to get simply the day and forged it to an integer even as appending the statistics to the date's listing. Support Vector Regression is used are expecting the result, the usage of SVR teach the dataset to get the accuracy of the prediction and linear regression is likewise used to technique to modelling the connection among a scalar reaction or based variable and one or greater explanatory variables or unbiased variables. Create a characteristic that makes use of three distinct Algorithm three Ticker Data Processing Procedure

1: Read the dataset

2: Create the listing dates and costs

3: For date in dates do

4: Dates append to date.break up[o].

5: For costs in open do

6: Prices append to open

7: Print the dates

8: Print the costs Support Vector Regression SVR fashions with 3 distinct kernels to look which one plays the best. The characteristic will aver 3 parameters, the dates, costs, and the day that we need to do the prediction directly to get the rate. First I will create the 3 SVR fashions with 3 distinct kernels are linear, polynomial, radial foundation characteristic. Also upload withinside the linear regression model. Algorithm four Algorithm Evaluation Procedure Input: The educated dataset Output: The expected open rate for the day because the result

1: Read the dataset

2: Import SVR from sklearn.SVM

3: Import matplotlib.pyplot

4: Create the linear kernel

5: Create the polynomial kernel

6: Create the rbf kernel

7: Train the linear in dates, costs

8: Train the polynomial in dates, costs

9: Train the rbf in dates, costs

10: Create the linear regression

11: Train the linear regression

12: Plot the times in Xlabel

13: Plot the rate in Ylabel

14: Plot dates and costs in poly and linear and rbf

15: Return rbf expected result

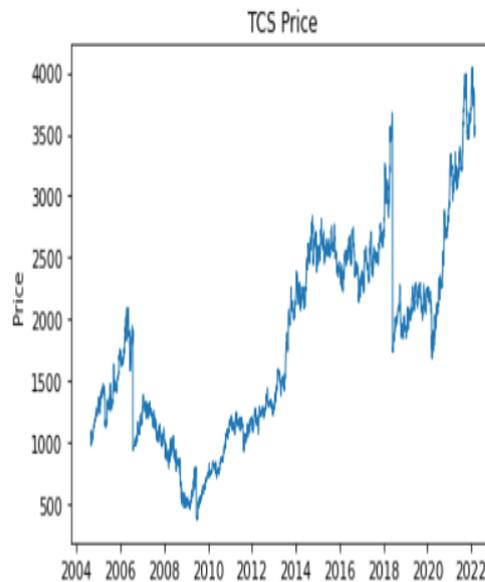
OUTPUT

In [6]: `tcs_data.head()`

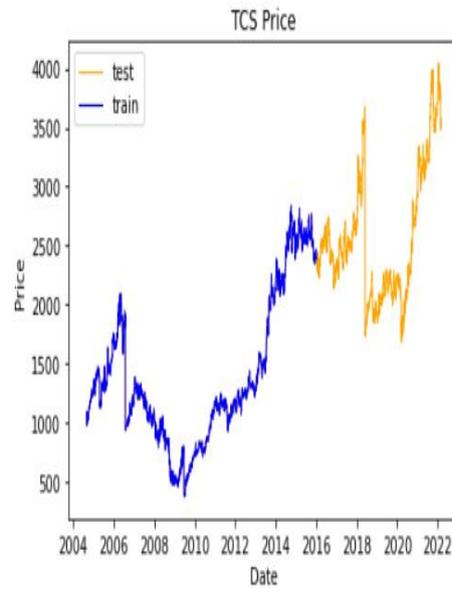
Out[6]:

	Date	Symbol	Series	Prev Close	Open	High	Low	Last	Close	VWAP	Volume	Turnover	Trades	Deliverable Volume	%Deliverble
0	2004-08-27	TCS	EQ	979.00	982.4	982.4	958.55	961.2	962.65	969.94	3830750	3.715586e+14	NaN	976527	0.2549
1	2004-08-30	TCS	EQ	962.65	969.9	990.0	965.00	986.4	986.75	982.65	3058151	3.005106e+14	NaN	701664	0.2294
2	2004-08-31	TCS	EQ	986.75	986.5	990.0	976.00	987.8	988.10	982.18	2649332	2.602133e+14	NaN	695234	0.2624
3	2004-09-01	TCS	EQ	988.10	990.0	995.0	983.60	986.0	987.90	989.68	2491943	2.466236e+14	NaN	790586	0.3173
4	2004-09-02	TCS	EQ	987.90	989.9	1004.6	986.00	994.0	993.65	996.96	2669544	2.661426e+14	NaN	501792	0.1880

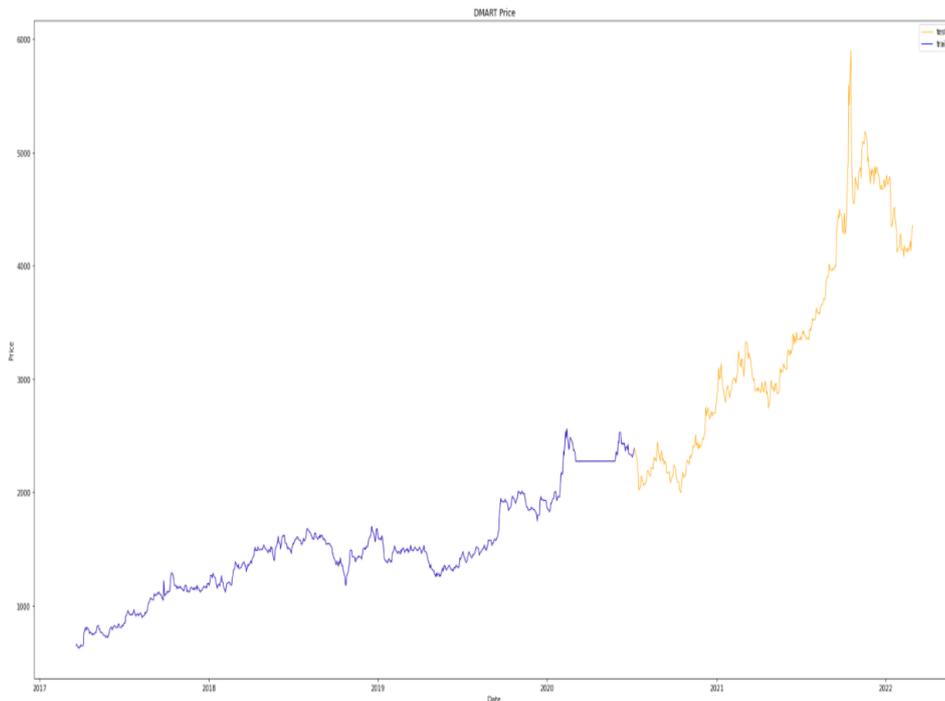
Out[8]: `text(0.5, 1.0, TCS PRICE)`



```
Out[11]: <matplotlib.legend.Legend at 0x238dd9fec0>
```



```
Out[45]: <matplotlib.legend.Legend at 0x275fd495a30>
```



6. DISCUSSION

The proposed approach makes use of Support Vector Machines (SVM) and Decision Trees. The benefit of using Decision trees over Neural Network are:

1. They are easy to program.
2. The top nodes in the tree will give the information about what data affects the prediction.
3. Trees are interpretable and provide visual representation of data.

4. Performs faster than Neural Networks after training.

The benefits of using SVM over neural networks are:

1. SVM has strong founding theory.
2. Global optimum guaranteed.
3. Requires less memory to store the predictive model.
4. Yield more readable results and a geometrical interpretation.

7. CONCLUSION

In this project, we are predicting closing stock price of any given organization, we developed a web application for predicting close stock price using LMS and LSTM algorithms for prediction. We have applied datasets belonging to Google, Nifty50, TCS, Infosys and Reliance Stocks and achieved above 95% accuracy for these datasets.

REFERENCES

1. Hu Z, Zhu J, Tse K. Stocks market prediction using support vector machine. In 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013; 2: 115-118. IEEE.
2. Usmani M, Adil SH, Raza K, Ali SS. Stock market prediction using machine learning techniques. In 2016 3rd international conference on computer and information sciences (ICCOINS), 2016 Aug 15; 322-327. IEEE.
3. Mankar T, Hotchandani T, Madhwani M, Chidrawar A, Lifna CS. Stock market prediction based on social sentiments using machine learning. In 2018 International Conference on Smart City and Emerging Technology (ICSCET), 2018 Jan 5; 1-3. IEEE.
4. Ouahilal M, El Mohajir M, Chahhou M, El Mohajir BE. Optimizing stock market price prediction using a hybrid approach based on HP filter and support vector regression. In 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), 2016 Oct 24; 290-294. IEEE.
5. Huang H, Zhang W, Deng G, Chen J. Predicting stock trend using fourier transform and support vector regression. In 2014 IEEE 17th International Conference on Computational Science and Engineering, 2014; Dec 19; 213-216. IEEE.
6. Sharma N, Juneja A. Combining of random forest estimates using LSboost for stock market index prediction. In 2017 2nd International conference for convergence in technology (I2CT), 2017 Apr 7; 1199-1202. IEEE.

7. Y. Dai and Y. Zhang, "Machine Learning in Stock Price Trend Forecasting," Stanford University; [http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningIn StockPrice TrendForecasting.p df](http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningIn StockPrice TrendForecasting.pdf).
8. J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Systems with Applications: An International Journal*, 2015; 42: 259-268.
9. B. Wanjawa and L. Muchemi, "ANN Model to Predict Stock Prices at Stock Exchange Markets," arXiv: 1502.06434 [q-fin.ST], 2014.
10. D. Mandic and J. Chambers, *Recurrent Neural Networks for Prediction*, Wiley, 2001.