Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN : 2348-4470
p-ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

Stock Market Prediction using Machine Learning

Prof. S .P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyam Anand, Tushar Mukherjee

Department of Computer Engineering, AISSMS IOIT,
SavitribaiPhule Pune University, Pune, India.

ABSTRACT: *The coremotive of this project is to analysethe market and predict its performance using various machinelearning techniques. The predictive architecture will use various attributes as the input and will predictwhether the market value will be Positive or Negative. The various attributes that are used in the model include Oil rates, Foreign Exchange Rate, Interest rate, Gold and Silver rates, NEWS, twitter news feedand Pattern Matching. The different ML techniques including Support Vector Machine with regression (SVR) and Recurrent Neural Network are used. The algorithm SVRproduced the most efficient results among all.*

TECHNICAL KEYWORDS: *Machine Learning, Regression, Meta Learner, Data Mining, Sentiment Analysis, Neural Networks.*

I. INTRODUCTION

Stock market prediction is the methoddetermining the future value of the stock or other financial instrument traded on an exchange. The successful prediction of the future price could result in improved profits. The ability to predict the value of stock price efficiently is very important for the market dealers or investors to maximize their profits. Web is rich textual information resource such as financial news, twitter sentiments even that is unmanageable to one. But one can use this abundance textual information to get datasets of various companies.

The main goal of this project is to predict the stock market performance using various machine learning techniques and algorithms. The model will be usingvarious attributes as the input and will predict stock price as positive or negative. Various attributes being used in thisproject include oil rates, Foreign exchange rate, interest rate, gold & silver rate, News and twitter news feed and Pattern Matching.

We will be using machine learning techniques involving algorithms like Support Vector Machine with Regression Technology (SVR), Recurrent Neural networks(RNN). Once the base model becomes ready, we are planning to have to apply the STACKING technique of machine learning to get more accurate result.

II. SURVEY DETAILS

The following algorithms/techniques will be used in the implementation of the machine learning architecture-

1.1. Support Vector Machine (SVM):

SVM is a classification and regression algorithm we are going to use for time series prediction of historical data. It works by first classifying the available data and computing a graph indicating the value of the stock over the period of time. Then, various techniques for regression are applied including for analysing and computing the trends. The regression part of the algorithm is the main component resulting in the prediction function. SVM applied with regression gives reliable efficiency on which the customers can trust easily.The implementation is depicted using the following algorithm:

Stock_future.py

1. We initialize all the libraries required.
2. Download the historical database of the stock using Yahoo Finance API.
3. Stock_url = yahoo_stock_code + yahoo_excess_code returns the json using the request.get(stock_url).
4. Convert the json to readable format that is UTF-8 format.
5. Code variable stores the stock code for stock which is stored in ResultSet in result in the first row under the column symbol.
6. Variable named output_path_new is created that stores the address of a file that stores the first 40 records going backwards from today.

7. Store all the records downloaded from yahoo's website on the file whose address is stored in output_file using the function download_quotes in the getData.py file
8. Then we copy the last 40 entries of the file to the newly created file.
9. We convert the years, month, days into number of days passed since 1 Jan 2012 and store in an array named date.
10. Create the classifiers i.e.; SVR_LINEAR and SVR_RBF
11. Fit the data into the classifiers with Price and date.
12. Plot the values using matploit onto a graph.

GetData.py

1. The main function calls the download_quotes function
2. The start is set to epoch for 1 Jan 2012 and end is stored as epoch of today.
3. We get the cookie and crumb values form the get_cookie_crumb function that takes stock symbol.
4. A file is created where the downloaded data will be stored. This downloading of the csv file is done by the get_data function.
5. It returns the address of the csv file to the stock_future.py file
6. The get_now_epoch returns the epoch of the current instant of time.
7. The get_data function save the data on a file named symbol.csv where symbol is the stock symbol.
8. It then uses the Yahoo Finance API to download the data with the help of the stock symbol, characters that depict the high, low etc. of the stock along with the cookie and the crumb value.
9. It then stores the downloaded data into the file.
10. The get_cookie_crub gets the cookie and no of lines value from the get_page_data function
11. Crumb value is returned by the split_crumb_store.
12. Get_page_data returns the no of lines and the cookie.
13. It uses the Yahoo Finance API to get the cookie value using the function get_cookie_value.
14. Lines is a variable that stores the no of lines in the file.
15. Get_cookie_value returns the cookie value
16. Find_crub_store returns the line number containing the crumb value
17. Function split_crumb_store extracts the crumb value from the line containing the crumb value.

III. PROPOSED SYSTEM

The basic architecture of the software will consist of the following components –

a. Input Datasets
b. Basic algorithms
c. Meta Learner
d. Final Prediction

The input datasets will first be processed through the basic algorithms and the outputs from theses algorithms will be feeded to the meta-learner. The meta-learner will perform the task of optimizing the outputs of the algorithms. It itself is another more powerful algorithm which will improve the overall efficiency of the system. The meta-learner upon analysing all the outputs will produce the final prediction for the consumer.

SYSTEM ARCHITECTURE DESIGN

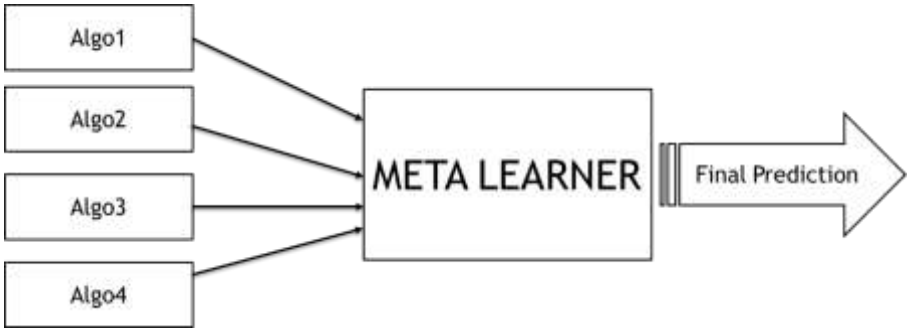

Figure 1: System Architecture Design

OUTCOME OF THE PROJECT

The final output will prompt the consumer about the value of the stock for the next opening and will also prompt the user whether to buy or sell the stock. A candle chart/graph will be generated showing the variations in the value of the stock in the past and various patterns formed. It will show the customer various details they are looking for so that they can decide whether to invest in the particular stock or not.

IV. CONCLUSION AND POSSIBLE FUTURE WORK

In this paper, we have examined and implemented various classification and regression techniques for developing an efficient stock prediction software. We have used the algorithms – support vector machine (SVM) with regression and recurrent neural network. The above mentioned algorithms are used for classification as well as regression purposes. We can efficiently predict the market value of the stock for the upcoming day by analysing various factors like – historical database, sentiment analysis and pattern recognition. A graph is generated showing the historical trends for the particular stock and the prediction is also given and advice is also given to the users whether they should invest in that stock or should sell it to get maximum benefits. Finally, the user will be able to deal in the market smartly without any hassle.

REFERENCES

[1] S. Qamar, and S. H. Adil, Comparative analysis of data mining techniques for financial data using parallel processing, In Proceedings of the7th International Conference on Frontiers of Information Technology,2009.

[2] S. H. Adil, and S. Qamar. Implementation of association rule mining using CUDA, International Conference on Emerging Technologies (ICET),2009.

[3] S. S. A. Ali, M. Moinuddin, K. Raza, and S. H. Adil, An adaptive learning rate for RBFNN using time-domain feedback analysis, The Scientific World Journal, 2014.

[4] A. W. Lo, and A. C. MacKinlay, Stock market prices do not followrandom walks: Evidence from a simple specification test, Review of financial studies, vol. 1, no. 1, pp. 41-66, 1988.

[5] C. D. R. Aurangzeb, Factors Affecting Performance of Stock Market:Evidence from South Asian Countries, International journal of academicresearch in business and social sciences, vol. 2, no. 9, 2012.

[6] S. Fatima, and G. Hussain, Statistical models of KSE100 index using hybrid financial systems, Neurocomputing, vol. 71, no. 13, pp. 27422746, 2008.

