Retraction

Retraction: Intelligent Monitoring and Forecasting Using Machine Learning Techniques (J. Phys.: Conf. Ser. 1916 012175)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Intelligent Monitoring and Forecasting Using Machine Learning Techniques

Sathyabama S, Stemina S C, SumithraDevi T and Yasini N
1Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India
Email: sathyabama.s@skct.edu.in, 17tucs236@skct.edu.in, 17tucs239@skct.edu.in, 17tucs258@skct.edu.in

Abstract. Stock market prediction is a crucial task and a prominent research area in the financial domain as investing in the stock market involves greater risk. Various events may affect public sentiments and emotions differently, which may affect the trend of stock market prices. Because of dependency on numerous factors, the stock prices are dynamic, and not static, highly noisy and nonlinear time series data. Due to its great learning ability, machine learning has been applied to this research area. The enormous stock market volatility emphasizes the need to effectively assess the role of external factors such as news in stock prediction. Machine learning algorithms have made a great impact in predicting stocks accurately. Methods based on learning for stock price prediction are found to be popular and a lot of strategies have been used to improve the performance of the learning-based predictors. However, performing successful predictions in the stock market is still a challenge. In this paper, we use machine learning algorithms and predict the stock market trade to examine if it increases or decreases. Compared with existing learning-based methods, the effectiveness of this new enhanced learning-based method is demonstrated by using a Naïve Bayes classifier which is found to be consistent and gives the maximum efficiency.

1. Introduction

A stock, in general, represents ownership states on business by a particular individual person or a group of people. The stock market comprises prospects and suppliers of stock. Forecasting the trends in stock market prices is a very demanding chore due to the many unpredictable factors involved and many fickle that impact the market value in a particular day such as economic state, shareholder sentiments in a direction of a specific company, political affairs, etc. Because of this, stock markets are easily influenced by fast alterations, causing random variations in the stock price. Stock market price forecasting is contemplated to be an accidental procedure with variations that are more noticeable for short time windows since stock market series are generally vigorous, non-parametric, chaotic, and clamorous. One of the significant speculation exercises is swapping the stocks on share markets. Investors need to know whether some stock will get high or low over a specific time. To acquire the correct output, the approach used is to implement machine learning techniques. A system is crucial to construct which will work for higher accuracy and it should contemplate all key factors that could
impact the outcome. There are various different methods of implementing the prediction system. With the advancement of the digital and technology era, the prediction has moved up into the technological domain. The most prominent, convenient and promising technique involves the use of Artificial Neural Networks(ANN), Recurrent Neural Networks, which is the implementation of machine learning techniques. Machine learning trains the system to learn from the past experiences which can be used to predict the future. Recently, many researchers are using machine learning techniques for prediction. These predictions in the stock market were used to form stock trade prices. It produces a steady and linear curve. Investors choose to buy those stocks whose prices are expected to increase in the future. The uncertainty in the stock market holds back people from investing in stocks. In real-life there is a need to meticulously predict the stock trade. The dataset of this stock market price forecasting includes details such as the top headlines about the stocks.

The previous model includes the prediction of stock market using PSO (Particle Swarm Optimization) algorithm and SVM (Support Vector Machine) Classifiers. The efficiency is less when compared with the other algorithms. The proposed model includes analyzing the dataset with Naive Bayes Classifier which is found to be consistent and gives the maximum efficiency when compared with the other algorithms which was used earlier.

2. Related Work

[1] proposed deep learning and a hybrid algorithm using machine learning classifiers and social media, news to predict stock market. [2] used the decision trees, NNs, and SVMs for a financial expert system that was capable of improving the prediction accuracy. In addition to that, an effective system was designed for the trading on the stock exchange with the limitation in the forecasting accuracy due to the dependence of them in the time-series analysis of the market. [3] depicted the enhanced Fuzzy Random Auto-Regression (FR-AR) model, as the predicting accuracy of the FR-AR model was enhanced by using this method and was able to solve the biasness between the machine and man error. [4] designed the Coupled matrix and tensor factorization method that was capable of forecasting the changes in the trade with the missing values but the drawback was it was not able to consider other data sources. The stock market is considered non-static, volatile, complex, and the forecasting of the stock market prices is one of the risky tasks. The ANN methods were unable to do this task [5]. One of the demanding fields for the person investing in the stock market is its prediction for getting the gain in the trade. It is essential for them to know about the status [6].

3. Dataset Descriptions

A data set comprises two modules. In the dataset rows and columns are the two modules. In addition to that, a key characteristic of a data set is that it is sorted so that each row holds one observation. In this paper Stock Market Google Dataset is used [7].

The attributes used in the dataset are

- **Date:**
  - It gives the date of the news displayed.

- **Headlines of the newspaper:**
  - It consists of the top headlines of the newspaper related to the stock market.

- **Label:**
  - The label column consists of two values
    - 0 -> stock price will not
increase on next day 1 -> stock price will increase on next day

4. Methodology
The framework in figure 1 represents the module description of the analysis.

![Overall Architectures](image)

**Figure 1.** Overall Architectures

4.1 Data Set Collection
The collection of data is a well-organized process of collecting and analyzing particular information to put forward the solutions to relevant questions and evaluate the results. Its main focus is on detecting out a particular subject matter [8-16]. Data is gathered to put through hypothesis testing which seeks to describe a phenomenon. The foremost reason for collecting data is done through quantitative or qualitative methods to make sure that the probity of the research question is maintained. Data collection rescues the researcher's time and funds that would if not be prodigal without a deeper understanding of the topic or subject matter. Gathering data as a proof is needed to prove the need for the modification [17].

4.2 Data Preprocessing
The process of data pre-processing is preparing the raw data and making it suitable for work or a preliminary step that takes all of the available information to organize it, sort it, and merge it. It is the first and Decisive phase while creating a machine learning model. It may affect how outcomes of the final data processing can be interpreted. It is used to remove

- Missing/Incomplete values
- Inconsistent values
- Irrelevant values
- Duplicate values

Retracted
Noisy data/Outliers

In this dataset, there are a lot of symbols and to remove that the data is being pre-processed. And also, apart from alphabets, all the other characters are replaced with a blank.

4.3 Data Splitting
Data splitting is the process of dividing the dataset into two portions. One part of the data is used as a training set and the other part is used as a testing set.

Training Set:
This part of the dataset is used for training purposes. The machine learning model observes and learns various things so that it would encounter the learned experience in the future.

Test Set:
This part of the dataset is used for testing purposes. Once the phases in the training dataset are completed, then the machine learning model is tested to observe whether it encounters the data correctly.

4.4 Machine Learning Algorithms

4.4.1 Naïve Bayes Algorithm

Naive Bayes classifiers can be realized on high-dimensional datasets. Naive Bayes classifier prophesies the possibility of each class based on the feature vector. It is a straightforward technique for building classifiers. It is based on the Bayes’ Theorem. It is not a single algorithm. It is a family of algorithms where they share a common principle, that is, every feature is not dependent of each other. For instance, a fruit may be considered to be an orange if it is orange in color, round and of medium size of about 10cm. A naïve Bayes classifier figures out each of these characteristics to put up independently to the probability that this fruit is an orange, any way of feasible correlations between the color, roundness, and diameter characteristics.

Figure 2 pictorially represents the overview of naïve bayes algorithm.

4.4.2 Random Forest Algorithm

Random forest is a type of supervised classification algorithm. It is nothing but the collection of several decision trees. Each decision tree will offer its predictive value. By considering the outcome of all the decision trees the value with higher votes is identified and the one with a higher vote is chosen. However, the characteristics of data can influence their performance. There is a straight association
betwixt the number of trees in the forest and the output it can get. The lower the number of trees, the lower the precise of the output. It can also catch the missing values. The problem of overfitting is also ignored by using this algorithm. Figure 3 pictorially represents the overview of random forest classifier algorithm.

![Random Forest Classifier](image)

**Figure 3.** Random Forest Classifier

5. **Experimental Results**

Comparative analysis of Naive Bayes and Random Forest in pictorial representation in figure 4 and 5.

![Random Forest Classification](image)

| precision | recall | f1-score | support |
|-----------|--------|----------|----------|
| 1         | 0.86   | 0.70     | 0.82     | 186      |
| 0         | 0.81   | 0.84     |          | 102      |
| micro avg | 0.83   | 0.80     |          | 378      |
| macro avg | 0.83   | 0.80     |          | 378      |
| weighted avg | 0.83 | 0.80     |          | 378      |

**Figure 4.** Random Forest Classification

![Naive Bayes Classification](image)

| precision | recall | f1-score | support |
|-----------|--------|----------|----------|
| 1         | 0.70   | 0.82     | 0.87     | 186      |
| 0         | 0.77   | 1.00     |          | 192      |
| micro avg | 0.85   | 0.85     |          | 378      |
| macro avg | 0.85   | 0.85     |          | 378      |
| weighted avg | 0.85 | 0.85     |          | 378      |

**Figure 5.** Naive Bayes Classification

From the comparative analysis it is found that Naive Bayes provide maximum accuracy when compared with Randomforest classifier in the forecasting of stock prices in figure 6.
6. Conclusion

Two techniques have been utilized in this paper: Random Forest and NB (Naive Bayes) on the dataset. And these algorithms have given output separately based on its efficiency. Naïve Bayes classifier has produced maximum accuracy when compared with Random forest. Use of recently introduced machine learning techniques within the prediction of stocks have yielded promising results and thereby marked the utilization of them in profitable exchange schemes. It has led to the conclusion that it is possible to forecast the price of the stock markets with more accuracy and efficiency using machine learning techniques. In the future, the stock market price forecasting system can be enhanced by using a bigger dataset than the one that is being used currently.

References

[1] Stock market prediction using machine learning classifiers and social media, news W Khan, MA Ghazanfar, MA Azam, A Karimi… - Journal of Ambient …, 2020 - Springer
[2] Stock market one-day ahead movement prediction using disparate data sources B Weng, MA Ahmed, FM Megahed - Expert Systems with Applications, 2017 - Elsevier
[3] Fuzzy random auto-regression time series model in enrolment university forecasting R Efendi, NA Samsudin, N Arbaiy… - 2017 5th International …, 2017 - ieeexplore.ieee.org
[4] Enhancement of stock market predicting using an improved fundamental analysis-based approach YJ Chen, YM Chen, CL Lu - Soft Computing, 2017 - Springer
[5] Improving stock market prediction via heterogeneous information fusion X Zhang, Y Zhang, S Wang, Y Yao, B Fang… - Knowledge-Based …, 2018 - Elsevier
[6] Stock prediction using deep learning R Singh, S Srivastava - Multimedia Tools and Applications, 2017 - Springer
[7] Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach SK Chandar - Cluster Computing, 2019 - Springer
[8] Stock Market Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data M Kesavan, J Karthiraman… - 2020 4th International …, 2020 - ieeexplore.ieee.org
[9] Deep learning for stock market prediction M Nabipour, P Nayyeri, H Jabani, A Mosavi, E Salvana - Entropy, 2020 - mdpi.com
[10] Stock Market Prediction Using a Deep Learning Approach T Damronsakmethee… - 2020 12th International …, 2020 - ieeexplore.ieee.org
[11] 2019 Index IEEE Transactions on Knowledge and Data Engineering Vol. 31 G Acs, AH Akbarnejad, S Amer-Yahia… - IEEE Transactions …, 2020 - ieeexplore.ieee.org
[12] Stock market prediction using optimized deep-convlstm model A Kelotra, P Pandey - Big Data, 2020 - liebertpub.com

[13] D. Devikanniga, A. Ramu, and A. Haldorai, Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm, EAI Endorsed Transactions on Energy Web, p. 164177, Jul. 2018. doi:10.4108/eai.13-7-2018.164177

[14] H. Anandakumar and K. Umamaheswari, Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers, Cluster Computing, vol. 20, no. 2, pp. 1505–1515, Mar. 2017

[15] Indian Stock Market Prediction using Deep Learning A Maiti - 2020 IEEE REGION 10 CONFERENCE (TENCON), 2020 - ieeexplore.ieee.org

[16] Stock Market prediction on High frequency data using Long-Short Term Memory Z Lanbouri, S Achchab - Procedia Computer Science, 2020 - Elsevier

[17] Predicting Stock Price Using Sentimental Analysis through Twitter Data NN Reddy, E Naresh, VK BP - 2020 IEEE International …, 2020 - ieeexplore.ieee.org