Volume: 2 Issue: 12 Pages: 14-18

A Study of Opinion Mining of Twitter Data for Forecasting Stock Market Movements

[*1] Bhushan Jalamkar and [2]Dr. MM Bhonde

[*1] Research Scholar, Department of Computer Science, Shri. Shivaji Science College Amravati, Maharashtra, India.

[2]Associate Professor & Head, Department of Computer, Science, Shri Shivaji Science College, Amravati, Maharashtra, India.

Article Info.
E-ISSN: 2583-6528
Impact Factor (SJIF): 5.231
Peer Reviewed Journal
Available online:
www.alladvancejournal.com

Received: 11/Nov/2023
Accepted: 21/Dec/2023

***Corresponding Author**
Bhushan Jalamkar
Research Scholar, Department of Computer Science, Shri. Shivaji Science College Amravati, Maharashtra, India.

Abstract
Income estimation is an important topic that has attracted the attention of finance researchers for many years. It involves the assumption that material information disclosed in the past has some predictive relationship with future stock performance. This study aims to help stock market investors determine the best time to buy or sell stocks based on information obtained from historical prices. The decision will be made according to the decision tree classifier, one of the data mining techniques. To develop the model, we use the CRISP-DM method on real historical data of three large companies listed on the Amman Stock Exchange (ASE).

Keywords: Data Mining, Data Mining, Data Classification, Decision Tree.

Introduction

Stock market forecasting has always been a useful tool for many researchers around the world because it is very useful. Bitcoins absorb and react to price information instantly, making it a great investment option. Scientists and researchers have expressed interest in studying stock price forecasting because it can help understand trading behaviour and changes. The stock market has experienced many significant events that affect stock price predictions. Trading data is expressed as changes in stock prices. External factors affecting product prices (political decisions, etc.) are regulated by national regulatory authorities [26]. The stock market is a linear, non-parametric system that is difficult to model with good accuracy [1]. Investors always try to predict stock prices, find good stocks and find the right time to buy or sell. According to [2], [3-4], some studies have used macroeconomic, business and company-based strategic analysis to achieve this goal. The authors [5] and [6] consider that the critical analysis depends on the fair value of the stock and the expected return on investment. Analysis of potential performance and company performance. Therefore, stock prices can be predicted. Many people believe that it is the only good thing in the long run. However, fundamental analysis is usually not suitable for short- and medium-term forecasting. Some other studies use technical analysis techniques where trading rules are developed based on historical data of stock prices and trading volumes [2]. Technical analysis refers to various techniques that aim to predict future price movements using past stock prices and volume information, as discussed in [5] and [7]. This assumes that history repeats itself and that future market trends can be determined by studying historical price data. Therefore, it is believed that there are price trends and patterns that can be identified and used for profit. Most methods used in technical analysis are highly subjective in nature and not statistically valid. Recently, data mining and artificial intelligence technologies such as decision trees, layered layers, and neural networks have been used in this field [8].

Data mining [9] refers to the extraction or search of information from large repositories or data. Some features are the definition of concepts or classes, relationships and relations, classification, prediction, grouping, analysis, differences and differences and similarities. Classification of data can be done in several ways; one of these methods is to

use decision trees for classification. It graphically illustrates all the possibilities and how to achieve them. Decision trees and neural networks can be trained using appropriate learning techniques.

Literature Review

The work in [2] uses decision tree techniques to build on Lin's [12] work, which attempts to modify the filter rule, i.e., buy when the stock price is k% higher than the previous local lows and sell when it reaches the local lows. The share price fell by k% from previous local highs. The modification of the filter rules proposed in [12] is intended to provide three different decisions regarding the analysis. Real tests using products from Taiwanese electronics companies show that Lin's method outperforms traditional filters.

According to [2] in Lin's work, the process of mutual exchange of points is only related to previous information; No forward-looking information is included. Research [2] aims to improve filtering rules and Lin's research by segmenting content with respect to past and future data. The researchers used data from the Taiwan Stock Exchange and Nasdaq to perform the test. The test results show that the proposed method is better than the Lin method and the two-warehouse filtering method. Over the past 20 years, the financial landscape has changed in many important ways.

The development of powerful communication and business tools has expanded the possibilities of businessmen. Income estimation is an important topic that has attracted the attention of finance researchers for many years. It includes the assumption that important information published in the past has some relevance to the future product [10].

To extract these relationships from existing data, data mining techniques are new technologies that can be used to extract information from this data. Therefore, some researchers focus on the analysis and use of mathematics and other research. The field of artificial intelligence and data mining technology is receiving intense attention [11]. Some models were designed and based on the above methods, the authors of [5] conducted a study on the use of backpropagation neural network (BPNN) to create products and sell product alerts. The code designation of the neural networks was NN5. The system was trained and tested on historical price data of banks in Hong Kong and Shanghai from January 2004 to December 2005. The results showed that the system was able to predict the short-term value with an accuracy of approximately 74% with the valid standard.

The model in [11] uses the concept of serial topology and creates a new decision, a two-layer decision tree, for stock price prediction. The model developed by the authors differs from other studies in two aspects; First, the decision model is transformed into an unbiased decision model to reduce variance. Second, a two-level decision tree is used to improve the purchasing process. The results showed that the decision-making model was developed as the best purchasing method and was more efficient than purchasing. The authors [10] proposed a product return prediction method using data mining and neural networks. This study attempts to use various statistical analysis tools in machine learning for data mining to explore the predictive power of finance and business.

The authors investigated the performance of neural network models for level prediction and classification. The results show that for the same exposure to risk, trading strategies based on neural network classification models yield more results than trading strategies based on other ideas. Research [13] has begun to compare the work of Fama and the French model [14-15] with neural electrical tools to try to predict stock prices in the Chinese market. The aim of this study is to demonstrate the accuracy of a neural network in predicting stock prices of companies traded on the Shanghai Stock Exchange. To demonstrate the accuracy of the ANN, the authors compared the predictive ability of the Fama and French model with different models and different types of neural networks. The results of this study show that the neural network allows investors to increase the betting power of stock selection, and more importantly, that simple univariate models appear to be more effective in predicting returns than multivariate models.

Al-Haddad *et al*., [16] presented a study that aimed to provide evidence on whether the governance and performance indicators of Jordanian industrial companies listed on the Amman Stock Exchange (ASE) are affected by the variables that were proposed and to provide important indicators of the relationship corporate governance and firm performance that Jordanian industrial firms can use to address the agency problem. The random sample of the study consists of (44) Jordanian industrial firms. The study found a positive direct relationship between corporate governance and company performance.

Hajizadeh *et al*. [17] provided an overview of applications of data mining techniques such as decision tree, neural network, association rules, and factor analysis, and in stock markets. Predicting stock prices or financial markets has been one of the biggest challenges for the AI community. Various technical, fundamental and statistical indicators have been proposed and used with varying results.

Soni [18] reviewed the recent literature on machine learning and artificial intelligence techniques used to predict stock market movements. Artificial Neural Networks (ANNs) are identified as the dominant machine learning technique in stock market prediction.

El-Baky *et al*., [19] proposed a new approach for rapid stock market price prediction. The proposed approach uses a novel high-speed time-delay neural network (HSTDNN). The authors used MATLAB to simulate the results to confirm the theoretical calculations of this approach.

Khan and Parkinson [22] proposed a technique for processing security event logs and revealing causality between items found in the log. They could generate event-based rules to discover relationships between dependent and independent variables. Cause and effect relationships are ultimately discovered by generating associative rules prior to CR discovery. Their technique could provide causal relationships between security log entries.

Ding *et al* [23] focused on different aspects of the study namely quantile regression, causality, stock returns and oil returns. They used different stock indices in their study. They found a causal relationship between oil prices and stock returns. They also found asymmetric causality for some stocks. They used the concept of quantile causality using quintile regression models. They wanted to further investigate the transfer mechanisms involved in the discovery of causal relationships in quantiles.

Choudhury *et al* [24] attempted to find a causal relationship between business cycles and stock market volatility. We used both nonlinear and linear bivariate causality for multivariate analysis. They were able to find evidence of causality in empirical studies. In the stock market, they found a link between volatility and different business cycles. They also found a link between economic activity and stock market

volatility. They found that a country's policy regarding economic activity affects stock market volatility. Their research gained importance due to the global financial crisis at the time.

According to Lee *et al.* [25], causal data analysis is about discovering interesting facts related to the causal relationship between dependent and independent variables. Cause and effect relationships can exist in data from different domains. Causal data analysis techniques are useful for uncovering hidden relationships.

The Methodology of the Study

A data mining methodology is designed to ensure that data mining efforts lead to a stable model that successfully solves the problem it aims to solve. Various data mining methods have been proposed that serve as blueprints for organizing the data collection process, analyzing the data, disseminating the results, implementing the results, and tracking improvements [9]. CRISP-DM (Cross Industry Standard Process for Data Mining) [20] is used to build a model that analyzes stock trends using a decision tree technique. This method was proposed in the mid-1990s by a European consortium of companies to serve as an unprotected standard process model for data mining. This model includes the following five steps:

- Choose how to build the model.
- Evaluate your model using one of the known evaluation methods.
- Install a model on the stock market to predict the best action to take (buy or sell a stock).
- Be aware of what and why you are modeling.
- Understand the reason and purpose of stock price analysis.
- Prepare data used in classification models.

The main reason and goal of creating a model is to help investors in the stock market decide the best time to buy or sell stocks based on the knowledge gained from the historical prices of those stocks.

Table 1: Attribute Description

Attribute	Description	Possible Values
Previous	Previous day close price of the stock	
Open	Current day open price of the stock	Positive, Negative
Min	Min Current day minimum price of the stock	Positive, Negative, Equal
Max	Max Current Day maximum price of the stock	Positive, Negative
Last	Last Current day close price of the stock	Positive, Negative, Equal
Action	Action The action taken by the investor on this stock	Buy, Sell

The Oracle Amman Stock Exchange (ASE) database contains historical prices of 230 listed companies since 2000. As the volume of this data is very large and complex, it was decided to select 3 companies trading on this exchange. Companies are selected based on the following five factors representing changing company size and liquidity: market capitalization, trading days, turnover rate, transaction value and number of traded shares as well as trading representative companies are also considered while selecting. These companies. These companies are "Arab Bank" whose trade mark is "ARBK", "United Arab Investors Company" of which four are in the banking sector, "United Arab Investors Company" whose trade mark is "UAIC" which is in services, and "Engineering, Electronics in the Middle East". and Electronics". Heavy industry with the code "MECE" belongs to the "complex" business sector. The period selected is from April 2005 to May 2007, which reflects the current situation and market conditions at that time. Behaviour, like other behaviours, has been identified as marginal and has no direct impact on work. Table 1 shows six selected behaviours along with their explanations and potential consequences. An investor's action while buying or selling shares is characteristic and is called "action". The data for this vehicle was also obtained from the ASE database, which is the daily workplace of one of the largest owners of the products in question. A net position can be buying or selling shares on a given day.

Table 2: Sample of historical data before selecting relevant attributes and before generalization

Previous	Open	Min	Max	Last	Action
25.8	24.7	23.8	26.82	24.86	Sell
25.4	25.41	26.75	25.4	26.69	Sell
26.11	24.11	24.10	26.11	23.25	Buy
26.1	25.10	23.15	26.1	24.11	Sell
25.11	25.15	22.11	25.11	24.33	Buy
24.61	23.45	25.31	22.11	23.52	Buy

At the beginning, when the data was collected, all the values of the selected attributes were continuous numerical values. Data transformation was applied by generalizing the data to a higher-level concept so that all values become discrete. The criterion that was created to transform the numerical values of each attribute into discrete values depended on the closing price of the stock on the previous day. If the values of the open, min, max, last attributes were higher than the value of the previous attribute for the same trading day, the numerical values of the attributes were replaced by the Positive value. If the values of the above attributes were lower than the value of the previous attribute, the numerical values of the attributes were replaced by negative ones. If the values of these attributes were equal to the value of the previous attribute, the values were replaced with the Equal value. Table 2 shows a sample of continuous numeric data values before manually selecting 6 attributes and before generalizing them to discrete values, while Table 3 shows the same sample after selecting six attributes and transforming them to discrete values.

The profit ratio is used to rank attributes and build a decision tree where each attribute is ranked according to its profit ratio. When the decision tree model was applied to data from three companies using WEKA software version 3.5 [21], the root attribute for both ARBK and UAIC was Open, while the attribute Last was the root for the MECE company decision tree.

Table 3: Sample after selecting the 6 attributes and after transforming them to discrete values

Previous	Open	Min	Max	Last	Action
Positive	Positive	Negative	Positive	Positive	Sell
Positive	Positive	Negative	Positive	Positive	Buy
Negative	Positive	Positive	Positive	Negative	Buy
Positive	Positive	Negative	Positive	Negative	Sell
Positive	Positive	Positive	Positive	Positive	Buy
Negative	Negative	Negative	Positive	Positive	Buy

As the tree building process continues, all remaining attributes have been used to continue the process. After constructing the complete decision tree, a set of classification

rules was generated by following all paths of the tree. The maximum number of attributes that were used in some of the generated classification rules was four attributes, while some classification rules used only 1 attribute. Both ID3 and algorithms were used in the creation of decision trees, and the pruning technique used to reduce the size of the created decision trees. Table 4 summarizes the number of classification rules that resulted from creating decision trees for each company using the algorithm.

Table 4: Summary of the number of the classification rules

Company	Number of Classification Rules Without Pruning	Number of Classifications Rules with Pruning
ARBK	20	11
UAIC	29	8
MECE	20	7

Deploying the model, the classification rules that were generated from the decision tree model can be used and integrated in a system that predict the best action and timing for the investors, either to buy or sell the stocks on that day.

Results and Discussion

CRISP methodology was used to build the model. This includes model estimation using only one or more specific estimation methods. For model evaluation, WEKA software was used to calculate the accuracy of the classification model. Two evaluation methods are used. K-CV (K=10 times) and the percentile method (68% of the data is used for training and the rest for testing). Both evaluation methods are used for decision tree classification methods ID3 and C4.5. Table 4 shows the accuracies of all the classifications created using the classification method and the evaluation method.

As the table shows, the classification accuracy obtained from the decision tree model is not very high for the training data used and varies from one company to another. The reason for such low accuracy is that internal financial factors influence the performance of a stock market; company news, financial news and general market indicators. Also, external factors can affect the company's performance in the market; political events and political decisions. Therefore, it may be difficult to have a model that provides high classification accuracy for all companies, while the performance of these companies varies.

Conclusions

This study presents a proposal to use a decision tree classification of historical stock prices to generate decision rules for buying or selling stock offers. The model developed in this way can be a useful tool to make good stock decisions based on historical stock price analysis to extract predictive information from this historical data. The results of the proposed model are not perfect because many factors affect the stock market, such as political events, general economic conditions and investor expectations.

For future work, there is still much room to test and improve the proposed model by evaluating all listed companies. Evaluating large learning methods such as neural networks, genetic algorithms, and association rules can be a rich area for future research. Finally, another rich area for future research is to investigate factors that affect market behavior, such as trading volume, news, and financial news, which can affect our stock price.

References

1. Wang YF. “Mining stock price using fuzzy rough set system”, Expert Systems with Applications. 2003; 24:13-23.
2. Wu MC, Lin SY, Lin CH. “An effective application of decision tree to stock trading”, Expert Systems with Applications. 2006; 31:270-274.
3. Al-Debie M, Walker M. “Fundamental information analysis: An extension and UK evidence”, Journal of Accounting Research. 1999; 31(3):261-280.
4. Lev B, Thiagarajan R. “Fundamental information analysis”, Journal of Accounting Research. 1993; 31(2):190-215.
5. Tsang PM, Kwok P, Choy SO, Kwan R, Ng SC, Mak J *et al*. “Design and implementation of NN5 for Hong Kong stock price forecasting”, Engineering Applications of Artificial Intelligence. 2007; 20:453-461.
6. Ritchie JC. Fundamental Analysis: A Backto-the-Basics Investment Guide to Selecting Quality Stocks. Irwin Professional Publishing, 1996.
7. Murphy JJ. Technical Analysis of the Financial Markets: a Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance, 1999.
8. Wang YF. “Predicting stock price using fuzzy grey prediction system”, Expert Systems with Applications. 2002; 22:33-39.
9. Han J, Kamber M, Jian P. “Data Mining Concepts and Techniques”. San Francisco, CA: Morgan Kaufmann Publishers, 2011.
10. Enke D, Thawornwong S. “The use of data mining and neural networks for forecasting stock market returns”, Expert Systems with Applications. 2005; 29:927-940.
11. Wang JL, Chan SH. “Stock market trading rule discovery using two-layer bias decision tree”, Expert Systems with Applications. 2006; 30(4):605-611.
12. Lin CH. Profitability of a filter trading rule on the Taiwan stock exchange market. Master thesis, Department of Industrial Engineering and Management, National Chiao Tung University, 2004.
13. Cao Q, Leggio KB, Schniederjans MJ. “A comparison between Fama and French’s model and artificial neural networks in predicting the Chinese stock market”, Computers & Operations Research. 2005; 32:2499-2512.
14. Fama EF, French KR. “Common risk factors in the returns on stocks and bonds”, The Journal of Finance. 1993; 33:3-56.
15. Fama EF, French KR. “The cross-section of expected stock returns”, The Journal of Finance. 1992; 47:427-465.
16. Al-Haddad W, Alzurqan S, Al_Sufy S. The Effect of Corporate Governance on the Performance of Jordanian Industrial Companies: An empirical study on Amman Stock Exchange. International Journal of Humanities and Social Science, 2011, 1.
17. Hajizadeh E, Ardakani H, Shahrabi J. Application of data mining techniques in stock markets: A survey, Journal of Economics and International Finance. 2010; 2(7):109-118.
18. Soni S. Applications of ANNs in Stock Market Prediction: A Survey, International Journal of Computer Science & Engineering Technology (IJCSET). 2011; 2(3)71-83.
19. Hazem M, El-Bakry Wael A. Awad, Fast Forecasting of Stock Market Prices by using New High Speed Time Delay Neural Networks, International Journal of

Computer, and Information Engineering. 2010; 4(2):138-144.
20. Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, *et al*. CRISPDM 1.0: Step-by-step data mining guide, 2000.
21. Witten I, Frank E, Hall M. "Data Mining: Practical Machine Learning Tools and Techniques", 3rd Edition, Morgan Kaufmann Publishers, 2011.
22. Saad Khan and Simon Parkinson. Causal Connections Mining Within Security Event Logs. ACM, 2017, 1-5.
23. Haoyuan Dinga, Hyung-Gun Kimb, Sung Y. Parkc Crude oil and stock markets: Causal relationships in tails. Elsevier. 2016; 59:58-69.
24. Taufiq Choudhrya, Fotios I. Papadimitrioua, Sarosh Shabib. Stock market volatility and business cycle: Evidence from linear and nonlinear causality tests. Elsevier, 2016, 1-41.
25. Jiuyong Li, Thuc Duy Le, Lin Liu, And Jixue Liu, Zhou Jin, Bingyu Sun, Saisai Ma. From Observational Studies to Causal Rule Mining. ACM Transactions on Intelligent Systems and Technology. 2015; 7(2):1-27.
26. Wikipedia: http://en.wikipedia.org/wiki/List of financial regulatory authorities by country