International Journal of Entrepreneurship and Business Management

Vol.1, No.2, (2022)
pp.139-145
e-ISSN: 2808-716X

Investor Sentiment Stock Price on Indonesia Stock Exchange

Mohamad Rafki Nazar
School of Economic and Business Department of Accounting Telkom University, Indonesia
nazarrafki@gmail.com

DOI: https://doi.org/10.54099/ijebm.v1i2.376

ARTICLE INFO

Research Paper

Article history:
Received: 23 September 2022
Revised: 11 November 2022
Accepted: 30 November 2022:

Keywords: Social Network analysis; Invesment, Market Effiesient; Market Analysist; Sentiment Analysis)

ABSTRACT

This study aims to obtain the results of how much value is formed from the relationship between issues and stock prices, and how the dynamics that occur between issues on stock prices and any increase or decrease in stock prices are related to repetitive issues.

The technique used in this research is using Social Network analysis, Investment, Market Effiesient, Market Analysis, sentiment analysis, the data used is based on User Generated Content (UGC), where the data is taken from social media which contains content created in looking for issues related to stock prices, and the movement of rising and falling stock prices taken from the IDX.

The result of this research are stock issues are influenced by positive sentiment from the market with a positive response of 81% and a negative 19%. In addition, 63% are influenced by micro (small) scale external issues. The classification results generated using the Suppori Vector Machine (SVM) model are more suitable than the Naïve Bayes Classifier (NBC)

INTRODUCTION

LQ45, a stock index with the largest and most liquid stock market capitalization constituents, also collapsed. Dropped 39.85 points (-4.14%), LQ45 stuck at 923.23. The decline in stock prices yesterday occurred evenly. Of all the stocks that are LQ45's constituents, 44 of them have dropped in price from the previous closing price (10/8). There was only one stock that did not experience a change in closing price compared to the previous day. (Hasbi Maulana, 2018).

Tuesday (14/8/2018) the Indonesia Stock Exchange (IDX) continued to decline. When the stock market closed, the Jakarta Composite Index (JCI) fell by 91.37 points (-1.56%), before perched at 5,769.87 (Hasbi Maulana, 2018)

LQ45, the stock index with the largest and most liquid stock market capitalization constituents, is also still under pressure. Dropped -20.14 points (-2.18%), LQ45 landed to 903.09. The decline in stock prices yesterday occurred evenly. Of all the stocks that are LQ45's constituents, 38 of them are still lower in price from the previous closing price (13/8). However, there were already four stocks that were able to increase in price and three stocks did not experience a change in closing price compared to the previous day (Hasbi Maulana, 2018)

Investor in the capital markets, they make decisions based on the information received by the stock price decline like this, so that by the time they do buy or sell. As a result, both sides of trading are always changing according to available information, which is ultimately reflected in the volatility of stock prices on the IDX.

The occurrence of a decline in stock price movements cannot be separated from the news and the development of information technology that is growing rapidly and this fast information in the form of existing news is also one of the factors for the speed of message delivery. So with the speed of delivery of messages, causing existing issues to spread quickly. So from the speed of information that is spread, causing the current stock price to be affected quickly.

Rationally, stock prices are formed quickly because of the process of adjusting information related to these stocks. The market is said to be inefficient if all available information is used to obtain abnormal returns in the market. It is said to be a perfect efficient market if there is no information that can be used to obtain abnormal returns in the market. According to Chowdury (2014), there are several things that affect stock prices, including the issuer's fundamental condition, the law of supply and demand, interest rates, foreign exchange, foreign funds on the stock exchange, stock price index, news and rumors.

News and rumors are all news circulating in the community concerning several things, be it economic, social, and political, security issues, to news about the cabinet reshuffle. With this news, investors can predict how conducive the country's security will be so that investment activities can be carried out. This has an impact on the movement of stock prices on volatility behavior in Indonesia stock exchange. However, there is a clear dividing line between news/news and rumors/issues/news is something that can be confirmed, objective. While issues are subjective, they cannot be confirmed, etc. Based on Kominfo data (kominfo.go.id), Indonesia is the sixth active user of social media in the world with a total of 112 million people beating Japan in sixth place, where the growth in the number of internet users is slower. In fact, in 2018, it was estimated that there were 3.6 billion people. So with the speed of growth of technology and information, causing existing information to be distributed quickly. For example, twitter as a social media platform is used as a medium for delivering messages quickly. In fact, not only twitter, mailing lists, forums and other social media became the eleventh.

Although there are many studies have shown that microblogging, such as Twitter, can provide numerous data for sentiment analysis (Agarwal et al., 2011), Harvey (1998) argues that with the increasing number of corporations aiming to have direct investments in emerging markets, there is a crucial need for an asset pricing model that produces appropriate hurdle rates for each of these potential investments. The novelty of this research is to see whether every increase in stock prices or decreases in stock prices is followed by every issue that is repetitive. This study observes the phenomenon of every existing issue, so this is the basis for the author to be interested in conducting research with the title Investor Sentiment Stock Price on Indonesia Stock Exchange

LITERATURE REVIEW

Share

Stock is one form of investment that is in great demand by investors compared to investments in the form of gold, deposits, because stocks have a much higher return, as well as high risk, so that this becomes an attraction for investors. On the other hand, stocks are an investment instrument that are chosen by many investors because they are able to provide two advantages, namely in the form of dividends and capital gains. Bodie et al. (2014:45) explains that capital gain is a gain or profit obtained from an investment where the value exceeds the purchase price, namely an increase in the price of the asset being invested. Shares are traded on the capital market. If the company makes a profit, each shareholder is entitled to a share of the profits to be distributed or dividends in accordance with the proportion of ownership.

Shares are divided into two types, namely common stock and preferred stock. Bodie et al. (2014:42) continues that ordinary shares are part of ownership in a company. Each share entitles the holder to one vote in all matters relating to corporate governance used in the annual general meeting of shareholders and to obtain a share of the company's financial benefits. Meanwhile, preferred stock is also explained by Bodie et al. (2014:46) is ownership in a company but has features similar to equity and debt, which promises to pay a fixed amount every year to the owner. Further explained by Husnan, (2015) preferred stock is said to be "preferred" because the company is obliged to pay dividends regularly to the owners of preferred shares and

has priority over the owners of ordinary shares. Both ordinary and preferred shares, in the transaction are determined by the share price and available volume.

Market Effecient Theory

According to Tandellin (2010: 219), Market efficiency is that the market price shows all available information, so we call it an efficient market, if all the information is already contained in the price, so investors are confused with stock prices whose values are too low or too high.

The concept of an efficient market implies a process of adjusting the price of securities to a new equilibrium price, in response to new information entering the market. Although the price adjustment process does not have to run perfectly, what is important is that the price formed is not biased. Thus, at certain times the market can be overadjusted or underadjusted when reacting to new information, so the new price formed may not be a price that reflects the intrinsic value of the security. So, the important thing about the efficient market mechanism is that the price formed is not biased with the estimated equilibrium price. The equilibrium price is formed after investors have fully assessed the impact of the information.

Big Data

Big Data is a large data set, we cannot be processed using traditional computing techniques. It is not a single technique or tool, but rather involves many areas of business and technology, Smailovic (2013) explains that Big Data can be defined as data that is already very difficult to collect, store and manage or analyze using the usual database system because its volume continues to multiply. Another opinion by Vincent (2016) that Big Data is data that exceeds the processing capacity of conventional databases whether the data is too large, moves too fast, or the data does not match the database structure, so to obtain value from the data, alternative ways are needed to process the data.

RESEARCH METHOD

Data and Source Collection

The data that users post on the web is publicly accessible or at least accessible to their circle of friends. The post may contain user-generated creative content for general publication, this is called user-generated content (UGC). UGC can be created through all social media and other forms of media, Sharma (2017).

In practice, UGC can be used as a source for the application of big data because it has content that can be described, such as the application of sentiment analysis or text classification as an effort to extract information to be used as additional knowledge presented (Moens, Li, Chua, 2014:12). Social media such as: Facebook and Instagram are also categorized as UGC because of the contribution each user makes to the public. The Stock Forum and Twitter are the data sources that are designated as data sources for this research.

Data Analysis Techniques

This research was assisted by using the 2016 x86 version of Semantria for Excel 5.3.141 software and MATLAB version R2017a and Microsoft Excel 2017 with add-ins namely data analysis tools and linear programming solver . The calculations used in this study are stock price predictions (t + 1) based on previously determined sentiments.

Dataset Cleaning

All micro-blogging data is filtered by removing data that is not relevant to the context. For example, there are several companies that are not included in the object of research, so they are deleted, for example $BUMI and $EXCL. if there is such content, it will be deleted because it does not match the content. The purpose of data cleaning is to remove spacing or punctuation. Irrelevant sentences can be problematic when entered in sentiment analysis software. Therefore, the data must be cleaned by deleting which names of shares are not included in the LQ-45 index.

Sentiment Analysis

Sentiment analysis is important in the big data era where the purpose of opinion mining is to get the author's feelings represented as positive and negative comments. At the sentence level, sentiment analysis will classify the opinions contained in a sentence with the orientation of positive and negative opinions.
One sentence describes the orientation of the same opinion of a person or organization. Sentiment analysis at this level can be used to analyze a statement/headline of an issue or analyze investor opinion in a forum. The category of each positive/negative word can be seen in the appendix. After being analyzed, the results show that the most widely used language is Indonesian. However, there are also some words using English or Bahasa in investment. The challenge in this research is that there are many words that have the same meaning, such as “decrease” or “descending” has the same meaning as “down”. Then, there are some slang words in Indonesian such as "cuan" which has the same meaning as "profit". The existing words must be entered as a list of keywords for later analysis using a semantic program.

Extraction of Sentiment
After categorizing the words that have been done above, the next step is to adjust the weight of each word in the semantria software. Semantria is a website that contains tools for providing text analytics and sentiment analysis for everyone.
The computer consistently assesses the existence of positive statements or negative statements that are directly related to existing documents. Examples of word weight settings in table 3.1.

Table 3.1. Example of Word Weight Setting

Say	Weight
"Down"	- 1
"Ride"	+1

The word weight must be between 1 and -1. And the author uses 2 types of sentiment, namely

Sentiment	Weight
"Positive"	-1
"Negative"	+1

Every word that exists, such as negation and auxiliary words, is of course very calculated. So that in the weight setting every word that is negated like "not good" will have a negative meaning.
Each of the existing basic words, can only be entered in one type of category. For example, the word "down", which can only be grouped into one type of group, namely negative and "up" is only grouped in positive words

Weighting

After doing the sentiment setting process, the next step is to enter the word into the semantria software. As explained above, Semantria is a website that contains tools for providing text analytics and sentiment analysis for the public. In finding the relevance value of each data entry and category in this analysis, a score from 0 to 1, indicates a low score provides low certainty and a high score provides high certainty. The weight affects the category relevance score of a particular category and has no effect on the categorization algorithm. The relevance value is calculated based on its own matrix and determines the machine's confidence in certain categories in the text. By default, semantria has a draft threshold that directs the engine to drop a relevance value lower than the threshold because it doesn't fit the category.

The first step in determining tone is to break down the document into its basic parts of speech. POS (Part of Speech) tagging is a technology that identifies all the structural elements of a document or sentence, including verbs, nouns, adjectives, adverbs and much more. To determine the sentiment of a document, it is necessary to identify the parts of speech in the document that show the user's emotions. What must be done first is to create a very large dictionary containing the sentiment keywords that have been carried out in the previous stage and the relative value of each word. This score is determined by how often a given word occurs

near a positive set of words (e.g. good, beautiful, and spectacular) and a negative set of words (e.g. bad, terrible, and terrible).

Using a very large collection of texts (web, via internet search engines) to evaluate the proximity of known positive and negative words to the word under consideration. For example, the word “down thin” is a word that has been combined between positive and negative words. That's the process that has to be done by checking to see if we should associate the word with a positive or negative sentiment, and how closely the words fit together.

Each word is returned with the existing word count. The words are combined using a mathematical operation called the log odds ratio to determine the score for a given phrase. Log odds ratio is a way to measure how strong the relationship 70 between the independent variables associated with in certain populations. For example, with the calculations that have been carried out, the results of the word “down thin” will come out, which is 0.7. It uses an existing algorithm to combine word scores in documents based on an operation. This is what is called semantics in Natural Language Processing.

RESULT AND DISCUSSION

Sentiment Analysis Evaluation and Validation Results

The true positive, true negative, false positive, and false negative based on the grouping on the machine, as a basis for seeing the results of evaluation and performance measurements. Sentiment analysis model as in Figure 4.1

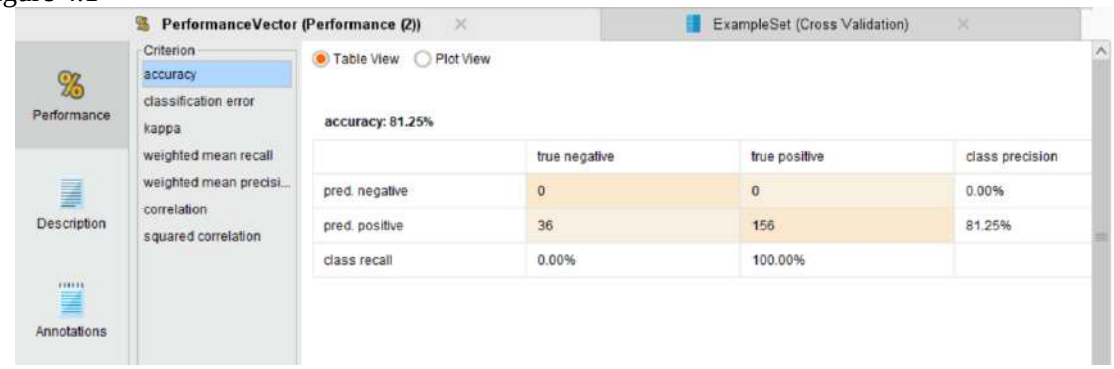

Figure 4.1 Evaluation Results Sentiment Analysis

Spearman Rank Result Test

The test results are calculated using the Spearman rank coefficient formula. Rank Spearman examines the ranking between the existing issues on changes in stock prices. The number of data (n) obtained in this study was 226 data. Then, the result of the difference is squared the difference between pairs (d2), which is 1783425. After that, the author calculates the Spearman rank correlation, the result is 0.07. Based on these results, Sarwono (2006) classifies the intervals of the strength of the correlation relationship in Suliyanto (2011) states that it is stronger if it is close to 1 and weaker if the correlation coefficient is close to 0. The results show 0.07 and explain that the correlation results belong to the 0-0.25 interval (Sarwono, 2006) which shows the strength of the relationship. Very weak between issues and market returns. So this explains that the issue cannot be used to make investment decisions because the results shown are very weak. This shows that investors should not use issues as guidelines/benchmarks for making stock investment decisions.

Stock Price Dynamics

Making stock price dynamics using a scatter plot. The scatter plot (XY) has points that show the relationship between two datasets. The first step is to determine the type of data to be used, whether it includes numeric or non-numeric data. Second, the data used is continuous or not and looking for the data used in this study is classified as numeric or non-numeric data. This study uses numerical attributes that are included in the data mining attribute group and to describe the relationship between the two variables generally uses a scatterplot/chart which is usually used to display and compare numerical values, such as scientific, statistical,

and engineering data. The direction of the relationship and the type of positive relationship move in the same direction if the explanatory variable can affect the response variable. A negative relationship will be shown by the opposite effect (negative slope). There is probably no significant association, in which case the scatter will not show any trend.

Rank Spearman Result Test

Rank pearman result test shows that the issue does not significantly affect changes in the value of stock prices (market return). This study explains sentiment or issues as an explanatory variable and stock prices (market return) as a response variable. Figure 4.14 indicates that there is no significant change in pattern when there are issues in the capital market. Based on the theory, in this case the spread of the line will not show any trend. Positive correlation explains when values increase together and negative correlation when one value decreases when the other increases. Correlation is not said to be causal, it states that correlation does not mean that one thing causes another.

This can be proven by using or examining the relationship between current issues and changes in market returns. The horizontal line (x) shows that the sentiment on the chart ranges from -15 to +20 and on the vertical (y) chart the relationship between changes in market returns which ranges from -10% to 5% indicates that there is a relationship between the two. Although at some points there is not necessarily a correlation, indicating that there is no relationship between the two variables. It has been explained in the theory that there are four types, including positive, negative, neutral, or perfect relationships when b = 0. The graph shows that the relationship between the two is b≠0 which indicates that there is no strong and significant correlation between the two. This proves that the information described in an efficient market does not clearly affect price changes and stock dynamics quickly.

CONCLUSIONS, SUGGESTIONS AND LIMITATIONS

During our observation in the 2018 stock issue. The conclusions of this research are stock issues are influenced by positive sentiment from the market with a positive response of 81% and a negative 19%. In addition, 63% are influenced by micro (small) scale external issues. The classification results generated using the Suppori Vector Machine (SVM) model are more suitable than the Naïve Bayes Classifier (NBC) showing an accuracy In making investment decisions, investors should not use the issue because the issue has a relatively weak correlation with changes in stock prices. The results of statistical calculations show the value generated using non-parametric statistics is 0.07, this is grouped in a very weak correlation. There is no significant pattern change if there is an issue in the capital market by looking at the scatter plot. A scatterplot is often used to identify a potential relationship between two variables, where one can be considered as the explanatory variable and the other can be considered as the response variable.

REFERENCES

Agarwal, A., Ng, W.J. and Liu, Y. (2011) Principle and Applications of Microbubble and Nanobubble Technology for Water Treatment. Chemosphere, 84, 1175-1180

Bodie, Kane, Marcus, (2014), Manajemen Portofolio dan Investasi, edisi 9 buku 1, Salemba Empat Jakarta

Chowdury, Spandan Ghose, Soham Routh, Satyajit Chakrabarti. (2014). News Analytics and Sentiment Analysis to Predict Stock Price Trends- International Journal of Computer Science and Information Technologies,5(3), 3595-3604. Retrieved from International Journal of Computer Science and Information Technologies

Hasbi Maulana (14/8/2018), Hanya satu saham mampu bertahan, ini daftar harga saham LQ45, http://investasi.kontan.co.id/news/38-saham-masih-turun-ini-daftar-harga-saham-lq45-14-agustus-2018

Hasbi Maulana (14/8/2018), 38 saham masih turun, ini daftar harga saham LQ45, http://investasi.kontan.co.id/news/38-saham-masih-turun-ini-daftar-harga-saham-lq45-14-agustus-2018

Harvey, C. (2001). "Asset Pricing in Emerging Markets." Working Paper, National Bureau of Economic Research

Husnan, (2015), Teori Portofolio Analisis Sekuritas, edisi kelima cetakan pertama, UPP STIM YKPN Yogyakarta

Liu, Bing. (2012). Sentiment Analysis and Opinion Mining. San Rafael, Amerika Serikat: Morgan & Claypool Publisher.

Moens, Marie-Francine, Juanz Li, dan Tat-Seng Chua, (2014), Mining User Generated Content. Boca Raton. Amerika Serikat: CRC Press.

Pangemanan, Helga, Deddy P.Koesindarto. (2015). Indonesian Capital Market Behavoour Using Sentiment Measurement In Stockbit Conversation- International Conference on Economics and Business Management, Thailand. Retrieved from International Conference on Economics and Business Management, Phuket, Thailand 82

Sharma, Er.Rooplal, Gurpreet Singh, Er.Baljinder Kaur. (2017). Implementation of Sentimenttal Analysis of Social Media for Stock Prediction in Big Data- International Research Journal of Engineering and Technology (IRJET), 04(07), 56-72. Retrieved from International Research Journal of Engineering and Technology (IRJET)

Smailovic, Jasmina, Miha Grcar, Nafa Lavrac, martin Znidarsic. (2013). Predictive Sentimentt Analysis of Tweets A Stock Market Application. Retrieved from Jozef Stefan International Postgraduate School lamova Cesta 39, 1000 Ljubljana, Slovenia

Suyanto, (2017), Data mining: Untuk Klasifikasi Dan Klasterisasi Data (edisi 1), Informatika Bandung

Vincent, Rohit dan Rohini V. (2016). Analysis of Social Networking Platform to Predict Stock market Changes-International Journal of Computer Trends and Technologies (IJCTT), 33(1), 40-42. Retrieved from International Journal of Computer Trends and Technology (IJCTT)

