International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Sentiment Classification from Social Media for Stock Prediction with Data Mining

Natassya Afdalena Triany, Sani M. Isa

Abstract: ***Social media currently plays an important role as a means of exchanging information. Through social media, information is obtained that can be used to see people's sentiments about a product or an event. Social media is a viable option to attract public sentiment through a method called sentiment analysis. The thing done is attracting sentiment from internet users through the posts made. In this way, sentiment data can be collected quickly and easily. Current economic behavior has proven that financial decisions are driven significantly by sentiment. The level of collective optimism or pessimism in society can influence investor decisions. Sentiment can also be interpreted as something that is felt by someone, both positive and negative. Sentiments and perceptions are psychological constructs and therefore difficult to measure in the analysis. This study focuses on sentiment analysis of information obtained from Twitter about stocks. For sentiment classification process ensemble methods of Naïve Bayes and SVM is used. Sentiment results are classified as positive or negative. We are expecting to see if there is connection between sentiment analysis from social media in predicting movement of IHSG stock price. As a result, we obtained strong correlation with coefficient of correlation r= 0.56609.***

Keywords: ***Sentiment Analysis, Twitter, Stock Index, Classification, Data Mining.***

I. INTRODUCTION

Social media has grown rapidly. The microblogging site has become a very popular communication tool among internet users. This can be seen with the emergence of millions of messages every day on popular websites that provide microblogging services such as Twitter, Tumblr, and Facebook. Users write about their lives, share opinions on various topics and discuss current issues.

The advent of social media combined with an easy-to-use feature microblogging service has dramatically changed the lives of people with more and more people sharing their thoughts, expressing their opinions. Social media is a viable option to attract public sentiment through a method called sentiment analysis [1]. The thing done is attracting sentiment from internet users through the posts made. In this way, sentiment data can be collected quickly and easily.

Social media currently plays an important role as a means of exchanging information. One example is Twitter, which has attracted millions of users to post and get information. Current economic behavior has proven that financial decisions are driven significantly by sentiment. The level of collective optimism or pessimism in society can influence investor decisions. Sentiment can also be interpreted as something that is felt by someone, both positive and negative [2]. Sentiments and perceptions are psychological constructs and therefore difficult to measure in the analysis. News articles have been used as the main source for text content analysis. For example, news articles are used to analyze public mood, where stock price movements can be predicted [3]. However, sentiment analysis on social media is a difficult process. Texts are usually short, contain many spelling mistakes, unusual grammatical constructions, etc. [4] In addition, the literature shows conflicting results in sentiment analysis for stock market predictions.

In general, there are two approaches in conducting sentiment analysis: using Machine Learning, Knowledge-Based. In Machine Learning based approach, Machine Learning requires a dataset to be used as training data. Therefore, efforts are needed to collect and conduct class tags on the sample dataset, besides the training process also requires time [5]. The accuracy of the Machine Learning classification approach is very good, but the domain classification performance depends on the dataset used during training [6]. The methods that this category such as: Naïve Bayes, SVM, Neural Network, Maximum Entropy

Knowledge-Based is the Sentiment Analysis approach at the word level, where the entity being processed is the word. The method included in this approach is Lexicon Based. Knowledge-based relies on a dictionary or lexicon dictionary that is used to assess the features obtained. a Supervised Machine Learning-based approach provides a high accuracy in detection of text polarity [7], but the results depend on the data set used. However, lexicon-based has good classification performance in cross domain and can be enhanced easily with source of additional knowledge for sentiment classification [8].

Figure 1 below shows the development of social media users in Indonesia. The total users reached 150 million users; this means wanting to use the internet to socialize through social media. The number of social media users reaches 56% of the total population of Indonesia, with mobile users reaching 130 million.

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Natassya Afdalena Triany*, Department of information technology from BINUS University Indonesia
Sani Muhamad Isa, Lecturer and researcher in the Department of Computer Science, BINUS University Indonesia

Through the same survey, Twitter took the 4th position as the most used social media, under YouTube, Facebook and Instagram. With a number of users reaching 6.43 million users (based on a survey in January 2019)

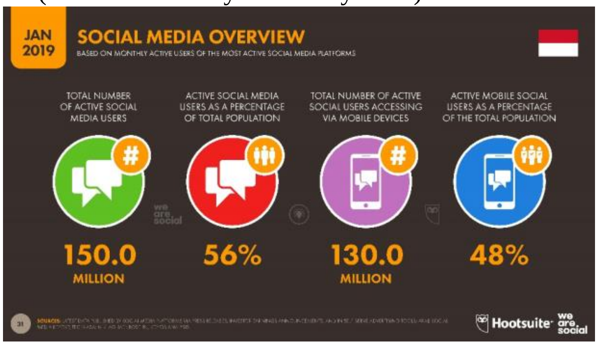

Fig. 1 Infographics of the use of social media in Indonesia

Twitter is a social networking service where users post and support via messages. This media is popular because it spreads very large instant messages (e.g. Tweets). Twitter provides a good tool for creating and presenting opinions. tweets can be used to analyze opinions about current issues, product reviews, movie reviews, elections

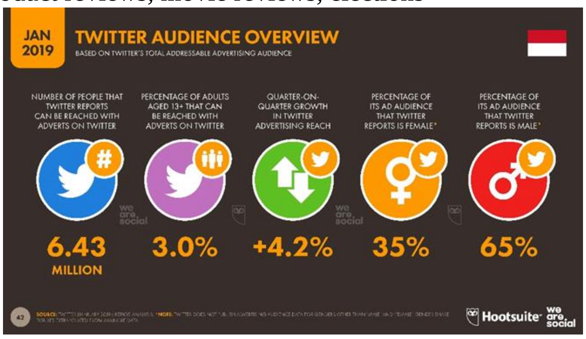

Fig. 2 Use of social Twitter social media in Indonesia

Current economic behavior has proven that financial decisions are driven significantly by sentiment. The level of collective optimism or pessimism in society can influence investor decisions. The sentiment is defined as the views or opinions expressed. The sentiment is a person's feelings which he expresses both in text and verbal form. This can be defined as personal positive or negative feelings [9]

Sentiments and perceptions are psychological constructs and therefore difficult to measure in the analysis. The process of identifying and grouping opinions expressed in a piece of text is known as Sentiment Analysis. However, sentiment analysis on social media is a difficult process. Texts are usually short, contain many spelling mistakes, unusual grammatical constructions, etc. [10]

Stock is one investment option that provides large profits in a relatively short time and allows investors to get profits in large quantities in a short time. However, along with fluctuations in stock prices, shares can also make investors experience large losses in a short time. For this reason, investors need a prediction tool that can help them in making stock investment decisions. For this reason, investors need a prediction tool that can help them in making stock investment decisions. Prediction activities are usually divided into three categories, namely short term, medium-term and long term. The prediction period for the short-term is prediction in minutes, hours and days until the weekly period. Medium-term prediction periods range from one week to one-month prediction period, whereas for long term predictions include prediction periods of one to several years into the future [11].

Things to consider if you are going to invest in shares, the individual must be able to understand the stock by reading its movements as a whole, which is also called an index. To analyze stock movements, it can be done using fundamental analysis and technical analysis. For investors and prospective investors in making decisions always try to get fundamental and technical information with various analyzes to reduce the risk of investing in shares.

Many investors desire is to use forecasting methods that can guarantee easy profits and minimize investment risk from the stock market. This is a challenge for researchers and market investors in deciding to buy or sell existing shares and is a motivating factor for researchers to evolve and develop new prediction models that can minimize risks in investing in capital market stocks.

II. STATEMENT OF PROBLEM

Stock price prediction is considered as one of the difficult processes [12] to succeed in financial forecasting due to the complex nature of the stock market. This is also influenced by very random, irregular, non-linear and noise stock market data. Stock prices depend on many factors, including macroeconomics and various news. [13].

Current stock market classification models are still suffering from low classification accuracy [14], [15]. Hence, this weakness in the models has a direct effect on the reliability of stock market indicators such as “series of statistical figures” and “financial reports” that explains the stock behavior [16],[17].

Much research has been done on the topic of sentiment analysis for stock predictions. However, not many people raised the topic of stock predictions on the main index of the Indonesia Stock Exchange (IDX), namely *Index Harga Saham Gabungan* (IHSG). To achieve the best prediction model for stocks, it is necessary to collect information about the stock including news and periodic reports of the company. However, the focus of this research is to find correlation between sentiment and stock price prediction. Author will collect tweets related to the IHSG stock prices for the same period of time, then decide the polarity of tweets and check correlation for the tweets and stock prices. The research will be carried out using the Ensemble Naïve Bayes approach and Support Vector Machine (SVM).

III. REVIEW OF LITERATURE

Ajla Kirlić et al. [18] in their study on stock price predictions using sentiment analysis from Twitter. the author collects tweets related to Microsoft Company and share prices for the same time period, then decides on the polarity of the tweet and restores the bond to the tweet and the share price. The study found a strong positive between tweet sentiment related to Microsoft Company and Microsoft's stock price.

Retrieval Number: E2226039520/2020©BEIESP
DOI: 10.35940/ijitee.E2226.039520
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

A. Pappu Rajan1 et al [19] in their study on Web Sentiment Analysis: Comparison of Sentiments with Stock Prices Using Correlation and Regression. The study is to investigate and predict possible relation between sentiment analysis scores and stock prices. The result is there is significant positive correlation between sentiment and its respective stock prices and the study leads to higher accuracy in prediction when they predict the closing prices of a firm by using a combination of opening prices and overall sentiment (although the information criterion is low)

Kevin P. Christ and Dale S. Bremmer [20] in their study on the relationship between consumer sentiment and stock prices. The study obtains three empirical results. First, there is no long-run relationship between stock indices and consumer confidence. Second, using granger-causality test on short-run relationship, there is indicate that stock prices affect consumer confidence, but no vice versa. And the last, changes of consumer confidence have no statistically significant effect on stock prices. Chuan Ju Wang et al. [21], this paper attempted to identify the importance of sentiment words in financial reports on financial risk. They use regression and ranking techniques to analyze the relationship between sentiment words and financial risk using a special financial sentiment lexicon. The experimental results are based on a bag-of-words model, a model trained with sentiment words only produces performance comparable to that in the original text, which increases the meaning of financial sentiment words on risk prediction.

IV. RESEARCH METHODOLOGY

Generally, steps in this research are collecting data, pre-processing, clustering, feature extraction, and text classification. Figure 3 is the proposed model which consists of data collection process, sentiment analysis process.

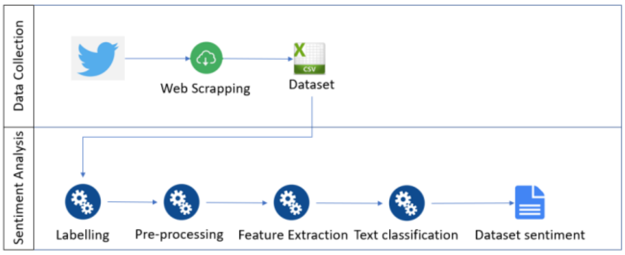

Fig. 3 System Design

A. Data Collection

Data collected from 1 January 2016 to 31 December 2018. Historical data and external factor data were taken from Yahoo Finance. Sentiment data comes from Twitter data. The data collected will be divided into 2 parts, namely training data and testing data. The tweet data collected is only limited to Indonesian tweets. The data will be processed and grouped into two types, namely: positive sentiment, negative sentiment. Twitter data collection uses the Twitter API which is then stored in CSV format.

Information text that used to derived from tweets contain words such as: "*saham* or *IHSG* or *IDXDailyTrading* or *YukNabungSaham* or *MauMulaiKapan* or *BI* or *BBRI* or *ICBP* or *TLKM* or *UNVR* or *GGRM* or *BBNI*". The total data of tweets successfully retrieved was 119391 tweets from the period January 1, 2016 to December 31, 2018. Data that not related to the topic will be deleted using Ms. Excel. From the above process, total of 52073 tweets was obtained. For data training, 4500 tweets were randomly selected, and the data labeling process was carried out manually. From the total data, 37753 tweets were used as data training.

B. Sentiment Analysis

1) Data Labelling

The sentiment analysis will be carried out on this stage. Data from twitter will be labelling manually. Data labeling affects the accuracy of the classification results. The acquired dataset will be labeled for each row of data obtained. Data labeling is done manually to ensure that each row of data set gets the appropriate label. Each row of data will be given a Positive or Negative label.

2) Pre-processing

Pre-processing is done to eliminate elements that are not important or disturbing and make it easier to process data to the next stage. Some stages in pre-processing are as follows:

- Case Folding.

This stage serves to change or uniform the character letters in a tweet into lowercase characters.

- Tokenization.

Tokenization functions to break sentences into words. Tokenization is done by looking at the space in each comment and each word will be broken down based on the space.

- Stop Word.

This stage is done to eliminate words that are not important in the classification process, such as words: which are, but, or, to, in, with, and so on.

3) Feature Extraction

Feature extraction is carried out to identify aspects of the topic being commented on. Feature extraction reduces the number of variables by converting a large number of attributes into a reduced set of features. Feature extraction tries to obtain meaningful low dimensional representation of high dimensional data. In other words, Feature extraction can guarantee less information loss and a higher discriminatory power than feature selection [22]. TF-IDF (Term Frequency-Inverse Document Frequency) is a well-recognized method for evaluating the importance of words in documents [23]. The term frequency will calculate the frequency of words appearing from each text information. IDF (Inverse Document Frequency) is used to calculate the importance of a term. There are several terms such as "is", "an", "and" etc. which often occur but are not important

4) Text Classification

After getting the data set from labelling, preprocessing and feature extraction process, the process continues with the text classification process.

The classification method used is Naïve Bayes and SVM. Multiple learning algorithms is used on ensemble methods to get better predictive performance better than any of the constituent learning algorithms alone. A machine learning ensemble only consists of a concrete finite set of alternative models and allows for much more flexible structure to exist among those alternatives.

Unlike a statistical ensemble in statistical mechanics that consists of infinite alternative models. Ensembles combine multiple hypotheses to form a better hypothesis.

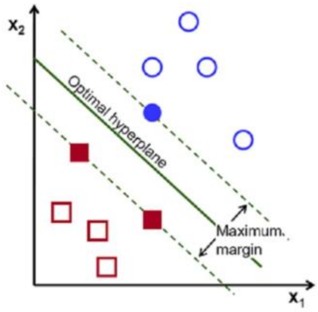

Fig. 4 Illustration of SVM (sources: www.aitrends.com)

SVM and Naïve Bayes is two of machine learning algorithm that used as a basic method for text classification. Their performance varies greatly depending on variant models, datasets and features used [24]. Naïve Bayes gives better results in text with shorter sentences, whereas Support Vector Machines gives better results for longer documents.

Ensembles of Naïve Bayes and SVM are carried out to achieve an effective and efficient classification process [25]. Ensemble methods is a multi-model system in which different classifiers or techniques are strategically combined to classify or predict statistics from the complex problem with better accuracy. The objective of this method is to minimize the likelihood of a poor selection from the model. Two conditions are needed to be fulfilled to achieve a good ensemble: accuracy and diversity of predictions [26].

Voting and averaging are two of the easiest ways to combine more than one machine learning algorithms. Voting and averaging are easily understood and implemented. For classification process using voting mechanism, for regression is using averaging. In majority voting (hard voting), each algorithm makes a prediction for each test instance and the prediction result is the one that receives more than half of the votes. In the cases of a tie, the ensemble method will select the class based on the ascending sort order.

Evaluation of the results of the study was carried out using the Confusion Matrix to measure performance and accuracy. To calculate the level of calculation using the following formula:

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$$

Where:

tp is the number of samples that are categorized correctly for positive samples,

tn is the number of samples that are categorized correctly for negative samples,

fp is the number of samples categorized incorrectly for positive samples,

fn is the number of samples that are categorized incorrectly for negative samples

5) Correlation

After sentiment data is obtained, a check will be made whether there is a correlation between sentiment analysis on changes in stock prices. Pearson r correlation will be used in this process.

Correlation reflects the relationship between variables, finding correlation does not mean one of the variables causes changes in other variables [27]. The main object of correlation is measuring the strength or degree of association between two variables. Association measurements are useful for measuring the strength and direction of the relationship between two or more variables.

To facilitate the interpretation of the strength of the relationship between the two variables given the following criteria:

0	: There is no correlation between the two variables
0 - 0.25	: Correlation is very weak
0.25 - 0.5	: Correlation is sufficient
0.5 - 0.75	: Strong correlation
0.75 - 0.99	: Correlation is very strong
1	: Perfect correlation

The correlation matrix is used to test the relationship between intervals and/or ratio variables. The Pearson correlation matrix will show the direction, strength, and significance of the bivariate relationship between all variables measured at intervals or ratio levels. Correlation is obtained by assessing variations in one variable because other variables also vary.

$$r_{xy} = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{n}(x_i - \bar{x})^2(y_i - y)^2}}$$

V. RESULT AND DISCUSSION

Tweets were collected for over the period from January 1st, 2016 to December 31st, 2018. The information used is derived from tweets that contain words such as: "*saham* or *ihsg* or IDXDailyTrading *atau YukNabungSaham* or *MauMulaiKapan* or *BI* or *BBRI* or *ICBP* or *TLKM* or *UNVR* or *GGRM* or *BBNI*". In total were collected 52073 tweets. And from Yahoo! Finance, Stock prices were collected for the same period of time.

The example of tweet and sentiment is shown below:

Retrieval Number: E2226039520/2020©BEIESP
DOI: 10.35940/ijitee.E2226.039520
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 1 Samples of collected tweets and their polarity

Tweet	Datetime	Polarity
Analisa Saham #IHSG Enam Keuntungan Jika DKE Diterapkan http:// goo.gl/KcPjsk Â	1/4/2016	Positif
[News] Jokowi: BUMN RI Jangan Jago Kandang, Serang Negara Lain http:// bit.ly/21v2Pv7 Â #valbury #saham	2/29/2016	Positif
10:37 WIB - Pergerakan IHSG Naik 0.22% ke Level 4.593 #IHSG #SAHAM #INDEKS	1/28/2016	Positif
Saham Bumi Teknokultura Terus Diborong, Ada Apa? http:// bit.ly/1msaLei Â #saham	1/3/2016	Negatif
IHSG perpanjang reli menanti keputusan BI rate http:// investasi.kontan.co.id/news/ihsg-perp anjang-reli-menanti-keputusan-bi-rate Â â€¦ #IHSG	2/22/2016	Positif
[News] Pasar perbankan syariah ditargetkan di atas 5% http:// bit.ly/21FrUnm Â #valbury #saham	3/3/2016	Positif
[News] Dolar Turun, Aktivitas Money Changer DKI Turun Separuh http:// bit.ly/1TkLCzX Â #valbury #saham	3/7/2016	Positif
Global dan Asia memerah, IHSG pun tak berdaya http:// investasi.kontan.co.id/news/global-da n-asia-memerah-ihsg-pun-tak-berdaya Â â€¦ #IHSG	1/14/2016	Negatif
# BI #Kembali Turunkan #BI #Rate #Jadi 6,75% http:// bit.ly/1pwhLbA Â	3/17/2016	Positif
Analisa Saham #IHSG Jokowi Ingin Tunjukkan Ada 'Tikus' di Persoalan Blok Masela http:// goo.gl/KcPjsk Â	3/5/2016	Negatif
Bursa masih meneruskan tarian poco2nya. Maju mundur. #saham	1/14/2016	Negatif
[News] MAIN akan kurangi ketergantungan bahan baku ekspor http:// bit.ly/1QeTzkj Â #valbury #saham	2/25/2016	Positif

Table 2 Sample of stock exchange market data

Date	Open	High	Low	Close	Adj Close	Volume
7/1/2016	5,027.62	5,039.69	4,971.58	4,971.58	4,971.43	3,924,520,000
7/11/2016	5,021.24	5,080.30	5,018.40	5,069.02	5,068.87	4,160,283,300
7/12/2016	5,092.17	5,120.13	5,075.56	5,099.53	5,099.38	5,657,618,500
7/13/2016	5,112.99	5,133.93	5,090.59	5,133.93	5,133.78	5,701,978,600
7/14/2016	5,125.69	5,131.03	5,081.03	5,083.54	5,083.39	5,097,381,300
7/15/2016	5,090.25	5,130.34	5,090.25	5,110.18	5,110.03	4,844,274,800
7/18/2016	5,107.92	5,131.71	5,093.15	5,127.50	5,127.35	4,773,677,100
7/19/2016	5,132.90	5,193.90	5,132.90	5,172.83	5,172.68	5,062,287,000
7/20/2016	5,185.59	5,242.82	5,185.59	5,242.82	5,242.67	4,953,995,800
7/21/2016	5,246.30	5,268.87	5,213.99	5,216.97	5,216.82	6,127,696,000
7/22/2016	5,215.17	5,215.53	5,179.62	5,197.25	5,197.10	3,938,603,400
7/25/2016	5,201.45	5,227.11	5,197.81	5,220.80	5,220.65	3,632,723,800
7/26/2016	5,218.27	5,231.50	5,200.62	5,224.40	5,224.24	4,911,860,600

Here, the current and the proposed scheme were examined by the experimental conclusions. Naïve Bayes based sentiment classification and SVM based sentiment classification approaches are considered as existing system. Combined method-based sentiment classification approach is suggested to enhance the system performance. The methods were distinguished by the metrics like precision, recall, f-measure and classification accuracy.

Precision is determined as a computation of correctness or quality; in fact, the recall is a computation of completeness or quantity. And, high precision represent that the approaches returned desirably more appropriate results than irrelevant. It concludes that the combined method-based sentiment classification method has shown the high accuracy results.

Recall is explained as the numeral of an appropriate documents recovered through a search partitioned by the complete numeral of accessible relevant documents. Recall is also the number of true positives classified through the total number of elements which are efficiently owe to the positive class

F-measure calculates the combined value of accuracy and recall as the harmonic mean of precision and recall. The f-measure value is acquired as follows.

A summary of the result:

Table 3 Classifier results summary

Classifier	Naïve Bayes	SVM	Ensemble Method
Recall	1.00	0.96	0.96
Precision	0.78	0.85	0.85
Accuracy	0.78	0.84	0.84
F1 Measure	0.87	0.90	0.90

Based on the results of the classification, then a calculation is performed to find the average sentiment per day which will be compared with the prediction results per day.

Sentiment data is then combined with IHSG historical data that has been downloaded via Yahoo! Finance to be compared and processed to look for correlations between the sentiments obtained by the prediction of the movement of the IHSG. The calculated correlation is the sentiment obtained on the D-day with the predicted movement of the IHSG on the same day. The following is a sample data that are combined:

Table 4 Samples of combined data

Date	Open	High	Low	Close	Predict	Avg Sentiment
7/1/2016	5,027.62	5,039.69	4,971.58	4,971.58	0.988853573	0.870748299
7/11/2016	5,021.24	5,080.30	5,018.40	5,069.02	1.019599403	0.956521739
7/12/2016	5,092.17	5,120.13	5,075.56	5,099.53	1.006018915	0.876712329
7/13/2016	5,112.99	5,133.93	5,090.59	5,133.93	1.00674572	0.895061728
7/14/2016	5,125.69	5,131.03	5,081.03	5,083.54	0.990184907	0.766917293
7/15/2016	5,090.25	5,130.34	5,090.25	5,110.18	1.005240443	0.876033058
7/18/2016	5,107.92	5,131.71	5,093.15	5,127.50	1.003389313	0.853846154
7/19/2016	5,132.90	5,193.90	5,132.90	5,172.83	1.008840566	0.904761905
7/20/2016	5,185.59	5,242.82	5,185.59	5,242.82	1.013530311	0.871621622
7/21/2016	5,246.30	5,268.87	5,213.99	5,216.97	0.995069447	0.868965517
7/22/2016	5,215.17	5,215.53	5,179.62	5,197.25	0.996220028	0.728000000
7/25/2016	5,201.45	5,227.11	5,197.81	5,220.80	1.004531242	0.829457364
7/26/2016	5,218.27	5,231.50	5,200.62	5,224.40	1.000689549	0.721428571
7/27/2016	5,245.40	5,301.93	5,245.40	5,274.36	1.009562821	0.917241379
7/28/2016	5,278.50	5,299.21	5,255.47	5,299.21	1.004711472	0.827160494
7/29/2016	5,300.90	5,334.12	5,215.99	5,215.99	0.984295772	0.773333333

The results obtained using the Pearson correlation matrix is r = 0.56609, which can be interpreted that the public sentiment obtained from Twitter has a strong correlation IHSG stock exchange and tweet's polarity for the same period of time. Strong correlation means that with an increase on one variable, another variable is increased too and vice versa.

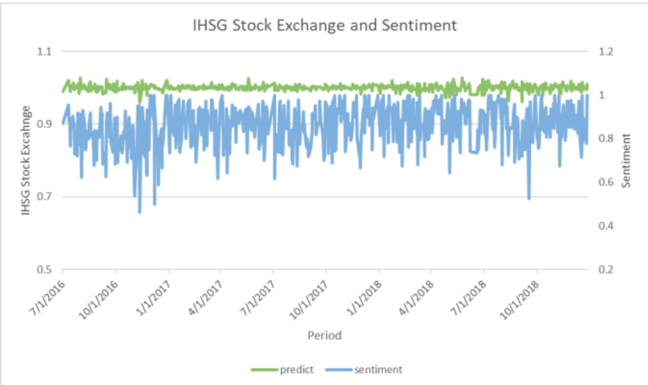

Fig. 5 Comparison of IHSG stock exchange and sentiment

VI. CONCLUSION

The study found strong correlation among sentiment of tweets related to IHSG / stock and IHSG stock prices. Although some limitations in our research like manual labelling the dataset, the results were showing affection of Tweeter public opinion to the IHSG stock exchange market and movements of the stock prices. We have expectation that our research will make contribution on the field related to sentiment analysis and stock prediction, although there are there are many research papers on the same topic.

In this work, we only consider twitter data to analyzed public opinion or sentiment that might be biased because not everyone shares their opinions on twitter. As a future work, we are expecting to examine sentiment from other social media is having effect on stock price movements and to obtain what has more impact to IHSG stock exchange market.

ACKNOWLEDGMENT

The author would like to thank Nurul Hidayati for her help in labelling process of training data for classification process and Adi Saputra for his help during the process.

REFERENCES

1. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis, 2(1).
2. Tyagi, A., & Sharma, N. (2018). Sentiment Analysis using Logistic Regression and Effective Word Score Heuristic, (April).
3. Zhang, X., Shi, J., Wang, D., & Fang, B. (2018). Exploiting investors social network for stock prediction in China ' s market. Journal of Computational Science, 28, 294–303. https://doi.org/10.1016/j.jocs.2017.10.013
4. Hai, T., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems With Applications, 42(24), 9603–9611. https://doi.org/10.1016/j.eswa.2015.07.052
5. J, A. M., & Chooralil, V. S. (2015). A Literature Survey on Web-Based Traffic Sentiment Analysis: Methods and Applications, 6(12), 926–930.

6. Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-Based Sentiment Analysis in the Social Web Lexicon-Based Sentiment Analysis in the Social Web, (October 2015)
7. Zhou, L. (2014). Movie Review Mining : a Comparison between Supervised and Unsupervised Classification Approaches Movie Review Mining : a Comparison between Supervised and Unsupervised, (July).
8. Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-Based Sentiment Analysis in the Social Web Lexicon-Based Sentiment Analysis in the Social Web, (October 2015)
9. Tyagi, A., & Sharma, N. (2018). Sentiment Analysis using Logistic Regression and Effective Word Score Heuristic, (April).
10. Hai, T., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems With Applications, 42(24), 9603–9611.
11. Devi, M. K. N., Bhaskaran, D. V. M., & G.Prem Kumar. (2015). Cuckoo Optimized SVM for Stock Market Prediction.
12. Cakra, Y. E., & Distiawan Trisedya, B. (2016). Stock price prediction using linear regression based on sentiment analysis. ICACSIS 2015 - 2015 International Conference on Advanced Computer Science and Information Systems, Proceedings, 147–154. https://doi.org/10.1109/ICACSIS.2015.7415179
13. Derakhshan, A., & Beigy, H. (2019). Engineering Applications of Artificial Intelligence Sentiment analysis on stock social media for stock price movement. Engineering Applications of Artificial Intelligence, 85(August 2018), 569–578. https://doi.org/10.1016/j.engappai.2019.07.002
14. Romanowski, A. (2015). Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction, 5, 1349–1354. https://doi.org/10.15439/2015F230
15. Mudinas, A., Zhang, D., & Levene, M. (2018). Market Trend Prediction using Sentiment Analysis : Lessons Learned and Paths Forward.
16. Bollen, J., & Pepe, A. (2011). Modeling Public Mood and Emotion : Twitter Sentiment and Socio-Economic Phenomena, 450–453.
17. Ludwig, S., Ruyter, K. De, Friedman, M., & Brüggen, E. C. (2013). More Than Words : The Influence of Affective Content and Linguistic Style Matches in Online Reviews on, 77(January), 87–103.
18. Kirlić, A., Orhan, Z., Hasovic, A., & Kevser-gokgol, M. (2018). Stock market prediction using Twitter sentiment analysis, 2(1), 1–4
19. Rajan, A. P., & Victor, S. P. (2016). Web Sentiment Analysis : Comparison of Sentiments with Stock Prices Using Correlation and Regression, (June)
20. Christ, K. P., & Bremmer, D. S. (2003). The Relationship Between Consumer Sentiment and Stock Prices. 78th Annual Conference of the Western Economics Association International.
21. Wang, C., Tsai, M., Liu, T., & Chang, C. (2015). Financial Sentiment Analysis for Risk Prediction Financial Sentiment Analysis for Risk Prediction, (April).
22. Kim, K. (2018). An improved semi-supervised dimensionality reduction using feature weighting: Application to sentiment analysis. Expert Systems With Applications. Elsevier Ltd. https://doi.org/10.1016/j.eswa.2018.05.023
23. Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). ScienceDirect ScienceDirect The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science, 152, 341–348. https://doi.org/10.1016/j.procs.2019.05.008
24. Wang, S., & Manning, C. D. C. (2012). Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2(1), 90–94. Retrieved from http://dl.acm.org/citation.cfm?id=2390665.2390688%5Cnhttp://dl.acm.org/ft_gateway.cfm?id=2390688&type=pdf%5Cnhttp://aclanthology.info/papers/baselines-and-bigrams-simple-good-sentiment-and-topic-classification
25. Korovkinas, K., & Danėnas, P. (2017). SVM and Naïve Bayes Classification Ensemble Method for Sentiment Analysis. Baltic Journal of Modern Computing, 5(4).
26. Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis : Bayesian Ensemble Learning. Decision Support Systems, 68, 26–38.
27. Sekaran, U., & Bougie, R. (2016). Research Methods for Business : A Skill-Building Approach (7th Editio). West Sussex, United Kingdom: Jhon WIley and Sons Ltd

AUTHORS PROFILE

Natassya Afdalena Triany, currently pursuing master degree in information technology from BINUS Graduate Program – Master of Computer Science. She received her bachelor's degree from Maranatha Christian University

Sani Muhamad Isa is a lecturer and researcher in the Department of Computer Science, BINUS Graduate Program - Master of Computer Science. He received his master's degree as well as doctoral degree in computer science from Universitas Indonesia (UI), and his bachelor's degree from Universitas Padjadjaran, Bandung, Indonesia. He has a lot of experiences in teaching and research in remote sensing and biomedical engineering field.