

CAN INVESTOR SENTIMENT IN SOCIAL MEDIA BE USED TO MAKE INVESTMENT DECISION IN STOCK MARKET?

Aldila Rizkiana, Institut Teknologi Bandung

ABSTRACT

The development of information technology affects the investor's behavior in stock investment. Thus data in Indonesia Social Media, hold important information to predict the movement of the stock price. Unfortunately, the research about investor sentiment in social media repeatedly using the same methods that focus on the amount of stock return and mainly analyzes the stock market in the US. So, we will use logistic regression to model stock buy/ sell decision using investor sentiment as the predictor. We find that the logistic regression model we developed is not fit, and thus, investor sentiment in Social Media alone cannot be used as the predictor of stock buy/ sell decision. We argue there are five possible reasons why we did not find a significant relationship between investor sentiment and stock return. There are the speed of information diffusion on stock prices, data sources, market structure, market capitalization, and investor sentiment measurement methods.

Keywords: Investor Sentiment, Stock Market, Social Media, Investment.

INTRODUCTION

The stock market in Indonesia becomes more attractive to the investor. This attractiveness is proved by the growing number of Single Investor Identification (SID), a unique investor ID, in Indonesia (KSEI 2016). The number of SIDs in Indonesia doubled from around 400,000 in 2015 to 1,000,000 in 2017. It shows that investment in the stock market is increasingly more popular in Indonesia. One of the reasons for this popularity is the potential profit they will get when they invest in the stock market. These investors must determine which one is the right stock to invest in, and when is the right time to buy or sell the stock they have, to maximize their profit (Lee and Jo 1999). To make this decision, investors often use several approach, such as fundamental analysis, which is based on fundamental data, such as financial report and market conditions (Nassirtoussi et al. 2014) and technical analysis, which is based on a set of rules using historical stock price and trading volume (Nazário et al. 2017). To get all of the information needed to do these two analyses, investor's search and share stock-related information to get opinions from the other investors or analysts.

Due to a recent information technology development, the investors can easily find all of the information on the internet, whether by searching it on Google, stock forum, or social media. They also will be able to share the information they got to the other investors by using the stock forum or social media. It shows that the internet changes how the investors get the information and how the investor acts with the information they have (Barber and Odean 2001). This phenomenon is especially true in Indonesia, as one of the largest active social media users in the world, with a penetration rate of 40% and growth rate up to 34% per year (Hootsuite and We Are Social 2017) because the social media can be used by the investors to search and share the information about stock market (Wieczner 2015). So, it is possible that the information on social media, as measured by investor sentiment, can be a valuable source of information for the investors.

Bukovina (2016) argues that through information demand and the reaction of society mechanism, the investor sentiments can be captured on social media and can be used as a proxy of investor sentiment measure. The information demand mechanism begins with the investor use an investment forum or social media as a publicly available source of information about the stock market because they have limited sources and access to the professional database. While, the reaction of society mechanism occurs because social media enable the investors to create, share, and respond to the existing information about the stock market. These two mechanisms make social media become a valuable source of data about attention, opinions, emotions, or social mood shared by social media users. Many researches finds that there is a relationship between investor sentiment on social media with stock price movement, for example, Mao et al. (2011) found that Twitter Investor Sentiment can be used as the predictor of stock returns. Other researchers also found that there is a correlation between tweet sentiment and stock return (Sprenger et al. 2014) and can be used to improve the predictability of stock return (Jin et al., 2014). Similar results also founded on a different platform, such as Facebook (Siganos et al., 2014), stock forum (Sprenger et al. 2014), and stocktwits (Sun et al., 2016). Different from previous research which is conducted on the US stock market, the research about investor sentiment in Indonesia also shows that investor sentiment in social media can be used to predict stock buy/sell timing (Rizkiana et al. 2017).

Unfortunately, the growing number of research about the relationship between social media and stock price repeatedly using the same method to solve the problem, such as linear regression (Mao et al. 2011), Vector Auto Regression (Li et al., 2016), and correlation (Li et al. 2016) that focus on the amount of expected return for the investors while using continuous dependent variable. We argue while it is true that investor is interested in the amount of expected return, the information that will benefit investor the most is the information about the decision to buy or sell the stock, which can be reflected using binary variable. The question we want to answer in this research is whether the investor sentiment alone is enough to determine the buy-sell stock decision. As far as we know, the investor sentiment research using binary variable as the dependent variable is scant.

To empirically test this relationship, we use logistic regression model, a specialized form of regression that is formulated to predict and explain binary (two-group) categorical variables, in our case is whether to buy or sell the stock, rather than a metric dependent measure (Hair et al., 2014). Using logistic regression, we can test whether the investor sentiment in social media can be used as a predictor of whether investors must buy or sell the stock.

Therefore, our research has four contributions. First, we add to the current investor sentiment literature by using logistic regression model to empirically test the relationship between the investor sentiment and buy-sell stock decision. Second, different from previous research that uses many variables to predict stock returns, in our research, we will use investor sentiment in social media as the predictor to simplify the decision process and isolate the effect of investor sentiment to the buy-sell decision, similar to Rizkiana et al. (2017). Third, we try to generalize the findings on previous research using data on social media and the stock market outside of US. Fourth, we analyze the relationship between investor sentiment and the stock market on the individual company level, not on the aggregate stock market.

Method

We will use Stockbit, a microblogging platform for the financial and investing community in Indonesia, to measure the investor sentiments. Different from the other microblogging platform in Indonesia, Stockbit specializes itself on ideas and news related to

the stock market so we can minimize the noise, i.e., comment unrelated to the stock market, that usually presents on other social media data, such as Twitter and Facebook. We limit our research to the top five banking company in Indonesia because it generated enough chat in Stockbit to be analysed. The methodology we used in this research for each stock is as following:

1. Calculate the daily return of stocks: the daily return of a stock can be calculated by (1)

$$r_t = \frac{\text{closeprice}_t - \text{closeprice}_{t-1}}{\text{closeprice}_{t-1}} \times 100\% \quad (1)$$

Where, closeprice_t is closing the price of company stock in days t. Closing price data is collected from yahoo.finance.com using the company's stock ticker from 10 April 2017 to 1 December 2017 and generated 165 data point.

2. Convert the daily stock return to the binary variable, i.e. buy (coded as 1) or sell (coded as 0) depending on the next day's return. We coded as 1 (buy the stock) when the next day (t+1) stock return is positive and coded as 0 (sell the stock) when the next day (t+1) return is negative. It means that we will buy stock on day t when the predicted next day t+1 stock return is positive and sell the stock when the predicted next day t+1 stock return is negative.
3. Calculate the Twitter Investor Sentiment (TIS). The predictor used as the proxy of investor sentiment is Twitter Investor Sentiment, adapted from Twitter Investor Sentiment (Mao et al. 2011). We collect the data from Stockbit (<http://www.stockbit.com>) through 3 steps. First, we filter the comment on Stockbit using stock ticker to ensure only comment related to one specific company shows up. Second, we scrape the comment from the website using the python program to extract information about username, comment content, and date and time for each comment posted on the website. Third, we clean the invalid data, i.e., unreadable data, invalid comments, etc. Fourth, we manually categorize the comment as bullish if it contains positive comments and bearish if it contains negative tweets about company stock performance. Fifth, we calculate investor sentiment for day t, TIS by using (2):

$$TIS(T) = \frac{N_{bull\ t}}{N_{bull\ t} + N_{bear\ t}} \quad (2)$$

Where $N_{bull\ t}$ is the number of the bullish tweet in time t, $N_{bear\ t}$ is the number of bearish tweet in time t.

4. Using Logistic Regression to estimate the effect of the covariates, TIS as a independent variable, on the outcome of investor decision, i.e., buy or sell as the dependent variable. Our logistic regression model is as follow:

$$Odds_{buy\ or\ sell\ i} = e^{b_0 + b_1 TIS} \quad (3)$$

So the result is we will have five logistic regression model. One model for each stock. Before we can use the models to decide whether we buy or sell the stock, we must ensure these two models have acceptable goodness of fit. We use Nagelkerke R square and Hosmer-Lemeshow to assess model fitness. Nagelkerke R square had a range of 0 to 1 and interpreted as reflecting the amount of variation accounted for by the logistic model, with 1.0 indicating perfect model fit (Hair et al. 2014). As for Hosmer-Lemeshow statistics, the logistic regression model is fit when it is significant (P-value < 0.05).

RESULTS AND DISCUSSION

Data Collection

We collected 165 stock price data for each stock, i.e. BBKA, BBNI, BBRI, BBTN, BMRI, from 10 April 2017 to 01 December 2017. Then, we calculate the one-day return and coded it to 1 or 0. One-day stock return for TLKM has an average of 0.184% and a standard deviation of 0.015. We also collected the tweets from Stockbit for the same periode using each stock ticker as the search term and manually categorize the tweets as bullish or bearish, resulted in a total of 294 tweets for BBKA, 528 tweets for BBNI, 280 tweets for BBRI, 427 tweets for BBTN, and 612 tweets for BMRI.

Data Processing

We use IBM SPSS 26 to test the logistic regression model, using TIS for independent variable and buy or sell (coded as 1 or 0) for dependent variable. The result of goodness of fit test for the model is presented in Table 1.

Stock	Nagelkerke R square	Hosmer-Lemeshow
BBCA	0.000	0.362
BBNI	0.010	0.019
BBRI	0.024	0.199
BBTN	0.041	0.075
BMRI	0.012	0.783

As can be seen in Table 1, all five model are not fit because Hosmer-Lemeshow p-value > 0.05 and low nagelkerke R square. This result shows that the model cannot be used to predict the buy or sell stock decision. It means we cannot only use the investor sentiments from Stockbit on a given day to make stock buy or sell decisions because the predictive power is low. These results are different from previous research which found the significant relationship between investor sentiment on social media and stock price movement (Mao et al. 2011; Siganos et al. 2014; Sprenger et al. 2014; Sun et al. 2016).

We argue, there are at least five possible reasons for these findings. The first one is related to the speed of information diffusion on stock prices. As explained by Sul et al., (2014), it is possible that the investor sentiments on social media spread slowly to investors and takes longer to be incorporated into stock prices. We only use previous day investor sentiment to predict next-day stock return, in other words, one day lag. Sun et al. (2016) confirms this in their research that investor sentiment slowly spread, so it will take longer to be reflected in the stock market, i.e., more than one day lag needed. There is another possibility; instead of spreading too slow, the investor sentiments spread too fast, i.e., less than one day, so there will be stock price reversal the next day (Renault 2017).

The second one is related to the data source. As stated by Mao et al. (2011), different data sources have a different effect on stock price movement. Initially, we choose Stockbit Social Media because we will be able to get high-quality data relevant to Stock Market (Sun et al. 2016) and easier data identification for grouping the social media comments based on stock ticker (López-Cabarcos et al., 2017). However, these results show that there is a possibility that the investor sentiment in Stockbit did not reflect the sentiment in the Stock Market because only 10% of retail investors in Indonesia registered as Stockbit's users (Suryadhi 2015). Therefore, the investors in Stockbit did not have a significant impact on the stock market. Initially, we choose Stockbit Social Media because we will be able to get high-quality data relevant to Stock Market (Sun et al. 2016) and easier data identification for grouping the social media comments based on stock ticker (López-Cabarcos et al. 2017).

The third one is related to market structure. Indonesia's stock market has high proportions of foreign investors, around 60% of the total stock market (Setiawan 2018). It means that foreign investor's trading activities is the main driving force for the movement of the stock price. Because we use the data in Stockbit, which is the social media platform for local investors, the investor sentiment formed that affects the trading activities for local investors is not strong enough to move the stock price.

The fourth one is related to market capitalization. Baker (2007) found that stocks with low capitalization, young, or not profitable are more sensitive to investor sentiments. All five stock we used in this research is stock with high capitalization and has excellent financial

performance, so the effect of investor sentiment will be less and not significant. Further research needs to be conducted on low capitalization, young, or not profitable stock to test this argument.

The fifth one is the selection of methods for calculating investor sentiment from social media. In line with the opinion of Baker (2007) that the method of measuring and quantifying the impact of investor sentiment on stock returns is something that must be considered because it can influence the results of research on investor sentiment. In this study, the method of calculating investor sentiment Mao et al. (2011) which assumes that comment only has a positive +1 or -1 negative to ease the calculation of decision making for the investors. In reality, one comment can have very positive values, for example, if there is news about the 200% increase in profit, while other comments can have a slightly positive value, for example, news about the 2% increase in profit. So, the investor sentiment does not fully reflect the contents of user comments. There is one possible solution that can be used to anticipate this weakness. That is following Renault (2017) methodology, which combines the use of positive and negative dictionary words with each investor sentiment score with machine learning to weight each comment according to the word occurrence of the words in the comments.

Based on these results, unfortunately, we cannot use the investor sentiments in Stockbit Social Media to decide whether to buy or sell stocks. However, we have several suggestions for further researches. First, develop a new investor sentiment index based on several investor sentiment indicators using Principal Component Analysis. Second, test the model on other company stock, especially low capitalization stock. Third, use weighted-lexicon for calculating investor sentiment. Fourth, consider using the other social media platforms to test the model, for example, Twitter, Facebook. Fifth, develop the model to analyze a longer time horizon, i.e. monthly, yearly.

CONCLUSION

In this research, we try to use investor sentiment in social media Stockbit for Indonesia investors to predict whether we must buy or sell the stock on a given day. We use a logistic regression approach to address the problem. The results show that investor sentiments on Stockbit alone cannot be used to determine stock buy/sell decisions on the same trading day. We have different findings from previous research, which shows investor sentiment has a significant effect on stock prices. We argue there are five possible reasons for our different findings. First, the speed of information diffusion on stock prices, data sources, market structure, market capitalization, and investor sentiment measurement methods.

ACKNOWLEDGEMENTS

This research is supported by The Eminent Research of Higher Education Program from the Ministry of Research, Technology, and Higher Education of Indonesia.

REFERENCES

- Baker, Malcolm. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, 21(2),129–52.
- Barber, B.M., & Odean, T. (2001). The Internet and the Investor. *Journal of Economic Perspectives*, 15(1), 41–54.
- Bukovina, J. (2016). Social media big data and capital markets-an overview. *Journal of Behavioral and Experimental Finance*, 11,18–26.
- Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). *On Multivariate Data Analysis Joseph F . Hair Jr . William C . Black Seventh Edition*. London: Pearson Education Limited.

- Sul, H.K., Dennis, A.R., & Yuan, L.I. (2014). Trading on twitter: The financial information content of emotion in social media. *2014 47th Hawaii International Conference on System Sciences*, 806–15.
- Hootsuite and We Are Social. (2017). Digital in 2017: Southeast Asia. Retrieved <http://www.slideshare.net/wearesocialsg/digital-in-2017-southeast-asia?ref=https://aseanup.com/southeast-asia-digital-social-mobile/>.
- Jin, X., Guo, D., & Liu, H. (2014). Enhanced stock prediction using social network and statistical model. *Advanced Research and Technology in Industry Applications (WARTIA), 2014 IEEE Workshop On* 1199–1203.
- KSEI. (2016). *Terobosan 19 Tahun KSEI: KSEI Menjadi Kustodian Sentral Terbaik Di Asia Tenggara (Press Release)*. Retrieved http://www.ksei.co.id/files/uploads/press_releases/press_file/id-id/129_siaran_pers_terobosan_19_tahun_ksei_ksei_menjadi_kustodian_sentral_terbaik_di_asia_tenggara_20161230150415.pdf.
- Lee, K.H., & Jo, G.S. (1999). Expert system for predicting stock market timing using a candlestick chart. *Expert Systems with Applications*, 16(4), 357–64.
- Li, Q., Zhou, B., & Liu, Q. (2016). Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion. *Proceedings of 2016 IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2016* 359–64.
- López-Cabarcos, M.Á., Piñeiro-Chousa, J., & Pérez-Pico, A.M. (2017). the impact technical and non-technical investors have on the stock market: evidence from the sentiment extracted from social networks. *Journal of Behavioral and Experimental Finance*, 15,15–20.
- Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing Survey, News, Twitter and Search Engine Data. *ArXiv Preprint* 10.
- Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y., & Ngo, D.C.L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16), 7653–70.
- Nazário, R.T., eSilva, J.L., Sobreiro, V.A., & Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance*. 66, 115-126
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the u.s. stock market. *Journal of Banking and Finance*, 84, 25–40.
- Rizkiana, A., Sari, H., Hardjomijojo, P., Prihartono, B., & Yudhistira. T. (2017). Analyzing the Impact of Investor Sentiment in Social Media to Stock Return : Survival Analysis Approach. *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* 519–23.
- Setiawan, S.R.D. (2018). JK Soroti Banyaknya Investor Asing Di Pasar Modal Indonesia - Kompas. *Ekonomi Kompas*. Retrieved <https://ekonomi.kompas.com/read/2018/01/02/124917826/jk-soroti-banyaknya-investor-asing-di-pasar-modal-indonesia>.
- Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook’s daily sentiment and international stock markets. *Journal of Economic Behavior and Organization*, 107, 730–43.
- Sprenger, T.O., Tumasjan, A., Sandner, P.G., & Welpe, I.M. (2014). Tweets and trades: the information content of stock microblogs. *European Financial Management*, 20(5), 926–57.
- Sun, A., Lachanski, M., & Fabozzi, F.J. (2016). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. *International Review of Financial Analysis*, 48, 272–81.
- Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking and Finance*, 73,147–64.
- Suryadhi, A. (2015). Stockbit: Platform Intelijen Investasi Racikan Mantan Pemain Bola.” Retrieved February 26, 2018 <https://inet.detik.com/cyberlife/d-3026465/stockbit-platform-intelijen-investasi-racikan-mantan-pemain-bola>.
- Tetlock, P.C. (2007). *Giving content to investor sentiment: The role of media in the stock market*. Published by Wiley for the American Finance Association Stable URL : <http://www.jstor.org/stable/4622297> The Role of Media in the Stock Market Giving Content to Investor Sentiment.” 62(3):1139–68.
- Wieczner, J. (2015). How Investors Are Using Social Media to Make Money. *Forbes.Com*. Retrieved April 3, 2017 (<http://fortune.com/2015/12/07/dataminr-hedge-funds-twitter-data/>).