Indonesia Capital Market Behavior Using Sentiment Measurement in Stockbit Conversation

Helga Pangemanan and Deddy P. Koesrindartoto

School of Business and Management, Institute Technology Bandung

Abstract: *In this study, data is collected from Stockbit posts, a micro-blogging platform, which is a community of Indonesia markets. Indonesia is the world's third-largest number of social media users and this paper exclusively processes the social media post in order to predict stock market indicators in Indonesia. This is the research gap from the previous research because this study has not been researched yet in Indonesia. Specifically, the stock market indicators are analyzed separately in all sectors of the Indonesian market, 9 sectors. We collected the tweets feeds about these nine market sectors for three months. Moreover, the difference with previous studies is from the methodology, which uses sentiment analysis method in Semantria for Excel program. This study proposes to improve the previous sentiment analysis method, which had already determined the term of mood public beforehand. First, this paper will check what kind of sentiment words frequently used by the public Indonesia on Stockbit and then will classify and defined them in several terms. The result of this study is all Indonesia market sectors have been proved that they are influenced by public sentiment in social media. In addition, within all Indonesia financial market sectors, finance index sector has the strongest relationship with public sentiment. Therefore, seems that just posting on social media forum can give an influence to the market returns.*

Keywords: *Micro-blogging Data, Capital Market, Financial Sectors, Sentiment Analysis, Behavioural Finance, Regression*

1. Introduction

The research about maximizing return investment has been analyzed for years. [1] [5] [6] [7] There are many theories that provide the investor to gain the maximum return. These theories have been researched in many ways to help the traders who are hard in making the decision in order to maximize their wealth. Recently, the theory which has been researched to support the decision making is using social media, which is known by micro-blogging data. Mining micro-blogging data is very recent topic to help the traders in forecasting stock market behaviour. There are several arguments from the previous research which support this approach. For example, Sheng Yu and Subhash Kak [6] stated that if the data on social media are extracted & analyzed properly, it can be useful to predict the certain human related events, such as finance, product marketing & politics.

Nowadays, social media which is known as place of sharing information has become a vital information for investor decision making. Barber and Odean [2] in their journal stated that the investor could act to the information that is delivered to them. Social media also often uses as share information about the trade among the investors. Also, in Indonesia there are a lot of community users that share information about stock market, which can be used in maximizing the function of micro-blogging services. In this case, the information has already represented the majority of the investors these days.

Moreover, micro-blogging data is easier and less expensive than traditional source such as surveys. In addition, the other benefit of using micro-blogging is a real time assessment which can be exploited during the trading day.

2. Literature Review

Based on previous research, there are a lot of researches from foreign country that research about social media that influence their stock, Dow Jones Industrial (DJIA), with different methods. For example, Olivia

Sheng and Chong Oh [4] measured micro-blogging data through sentiment analysis and stated the result that micro-blogging has a power to influence future stock price. This result is also supported by the journal from Johan Bollen et.al. [3] which is public mood from social media can increase the accuracy of future market stock. The accuracy result is up to 86.7% in predicting the daily market return of DJIA. Xue Zhang et al. [7] at their journal use posting volume of Twitter as methodology to predict future market. The methodology was classified into 3 base line of Twitter characteristic, the number of tweets per day, number of followers per day and number of retweets per day. The public data mood also has been defined beforehand in terms of hope, fear and worry. The result appears that when public tweet a lot of emotions, which are hope, fear, and worry, the next day index is going down and vice versa. Therefore, by looking the emotions on Twitter the traders can predict the index of following day.

Even though these literatures show the use of micro-blogging data to forecast stock market behaviour, the results need to be interpreted with caution. Lots of methodology can be used in this field and can have a different interpretation on the result. According to the recent research from Nuno Oliverira et al. [5], there is scarce evidence for the utility of the sentiment variables when predicting returns, and of posting volume indicators when forecasting volatility in US index. In conclusion, most of these studies haven't performed the robust evaluation. However, mining micro-blogging data to forecast stock market behaviour still appears to present promising results.

Different with previous research, the main goal of this study focuses on Indonesia market while the majority of previous research predicts US index. There is lack of the research of mining micro-blogging data in Indonesia Market even though Indonesia is including as the countries which use a lot of social media. By adopting the recent previous research, this study will use sentiment analysis in predicting Indonesia stock market behaviour. Similarity with other studies, micro-blogging data specifically will be taken from one popular trading forum resource in Indonesia, Stockbit (https://stockbit.com/). This resource is more interesting than Twitter, since Stockbit was designed with features that are focused with behavioural investors and traders that would be more useful as a special social media investment. This resource also can help to reduce noise because containing sign ($), which can be found the specific stock that needed (e.g. $IHSG, $SMGR). Also of note, this study proposes to improve the previous sentiment analysis which had already determined the term of mood public (hope, fear, and worry). First, the sentiment words will be determined by looking the frequent words that use by public on Stockbit. Contrast with other studies, which use Naïve Bayes method in sentiment analysis, this paper, will use Semantria for Excel program which also represent in sentiment analysis function. In determining the Indonesia stock market behaviour, this paper will analyze the stock market in Indonesia in separate ways, based on the stock sectors in Indonesia. From these sectors, linear regression will be used to determine whether social media can influence Indonesia market behaviour and which sector that has high relationship with the social media

3. Methodology

3.1. Stockbit and Stock Market Data

Contributions to the congress are welcome from throughout the world. Manuscripts may be submitted to Data was collected for 9 sectors in Indonesia, which are agriculture, mining, basic industry & chemical, miscellaneous industrial, consumer goods, property & real estate, finance, trade, services & investments and the last is infrastructure, utility & transportation sector. These sectors were chosen because they can represent the stock market Indonesia and can determine which sectors are most affected by social media. Other reason is each sector has several stocks which can provide more data which is very significant important in micro-blogging data. The more data obtained the better result that will be got. For each sector, we retrieved Stockbit and sector market data price from January 1, 2015 to March 31, 2015, in total of 62 trading days.

Stockbit (stockbit.com) is a community of Indonesian stocks, where traders and investors Indonesia to gather and share ideas. Stockbit platform is also integrated with other social media platforms, such as Twitter and Facebook, making it easy to share ideas on Stockbit to Facebook and Twitter account user. The main feature of

Stockbit is that "stream" consisting of ideas, chart, links and other financial data. Stockbit uses the sign "$" before the stock code as marking an idea to simplify the user in finding information about a certain individual stocks. We selected Stockbit content because it is exclusively about investing, resulting in a less noisy data set than collecting from a more generalist micro-blogging service. The data was filtered by the stock sign (i.e. $SMGR) according to stock that needed. Therefore, we separate the data stock sign name in accordance with the respective sector, ranging from total 1,000 to 2,000 tweets per each sector.

The sector market variables are considered in daily return. Price data were collected from Yahoo Finance (http://finance.yahoo.com/). Market returns determine changes in the asset value. To calculate market returns, this paper use the adjusted close price. Following is the formula to calculate market returns (Rt).

$$Rt = \frac{Pt - P(t-1)}{P(t-1)} \tag{1}$$

where P_t is the adjusted close price of day t and $P_{(t-1)}$ is the adjusted close price of the previous day. Market returns is very useful for the traders in making trading strategy in order to minimize the risk.

3.2. The technique of Analyzing Data

After collecting the data, below are the stages for analyzing the data:

1. Filtering the Data Set

All micro-blogging data will be filtered by excel formula =CLEAN (data). The purpose of this clearance is to delete the space or punctuation mark. The other filters are cleaning other stocks which are not included in each sector. For example, in one tweet there will be other name stock included (i.e. “Sell $SRIL now and buy $SMGR). This sentence can be a problem when enter the sentiment software analysis. Therefore, these data should be cleaned by deleting the sentence, which include a stock name that does not belong to the sector.

2. Sentiment Analysis

Sentiment analysis method will be used to determine the quantitative measurements of tweets’ sentiments: Positive (+1), Neutral (0), and Negative (-1). The relevant data will be input in Semantria Program for Microsoft Excel. The program needs some keywords for each group sentiment (positive, neural, and negative) to identify the tweets’ sentiment. Below are the keywords tables for each group sentiment:

TABLE I: Positive Keywords

Positive (+1)				
melesat	terbang	go	loncat	uptrend
melejit	hijau	rebound	melambung	rebound
take/collect	ambil	mantap	menembus	lonjak
nice/good	bagus	accumulated	akumulasi	tingkatkan
ganteng	naik	bullish	tambah	luar biasa
profit	panen	positive	positif	koleksi
buy	beli	tarik	borong	hajar

TABLE II: Negative Keywords

Negative (-1)				
turun	ambrol	terjungkal	dipangkas	hancur
anjlok	jeblok	lepas	terjun	cut
longsor	nyangkut	buang	jatuh	pangkas
sell/jual	merah	negative/negatif	downtrend	gagal
buang	payah	bearish	minus	loss
banting	CL	cutloss	merosot	sayonara
terperosok	terjebak	nyungsep	ambles	stuck

TABLE III: Neutral Keywords

Neutral (0)		
cermati	simak	ayo
perhatikan	tahan	waiting
amati	batal	prediksi
jebol	perkiraan	favorite

The majority keywords above are in Bahasa Indonesia. However, there are also some words using English (i.e. “profit”, etc) and the trading term (i.e. bullish, etc). The challenge in this research is there are a lot of similar words with same meaning from keywords above, such as "naiknya" or "kenaikan" has same meaning with "naik". The similar words also have to put into the keywords list. Then, there are also slang on Indonesia words such as "hijo" or "ijo" or "hejo" has same meaning with "hijau". These words should also be put on keywords list in order to be analyzed by Semantria program. Then, each tweet will be weighted and categorized based on the Semantria system.

3. Linear Regression

The result of the Semantria for Excel will be used as an independent variable. It will be pooled per day according to trading days. Because there is no trading in weekends and public holiday, the sentiment on that day will be add to the following day. The dependent variable is market return which has been discussed above (R_t). Then,test each sector using linear regression model:

$$Yi = \alpha + \beta Xi + \varepsilon i. \tag{2}$$

The regression model will show the p-value that represent the significant level of each variable. In this case, the variable is each sectors. By looking through p-value, we can know whether the predictor variable (sentiment) are related to changes in the response variable (market returns). The standard number in this model is the common alpha level, 0.05. Another interpretation of this model is by using the R^2, whereas give the measurement of how good the independent variable (sentiment) can influence the dependent variable (market returns). Previously, there are some standard requirements that must be met before performing linear regression test.

4. Experiment Result

Each sector has been tested using four different models of sentiment scoring, which is daily, cumulative daily, average daily, and cumulative average daily sentiment model. The purpose of these model test is to find the appropriate model in each sector using p-value and R-square. Before, each model has also been tested by the standard requirements of linear regresion. All sectors in cumulative daily and cumulative average daily sentiment model didn't pass the standard requirements. The model which passes all the requirement in all sectors is daily sentiment model. Some sectors in average daily sentiment model didn't pass the requirements too. Therefore, the expereimental result of this study will show in daily sentiment model.

In this model, all sentiment score from each tweet in one day will be summed. For example, below shows the movement of the daily sentiment for each sectors (Figure 1 – 9). As the comparation, in the figures also show the movement of daily return for the sectors. From these charts, the public sentiment and sector index return of all sectors generally have the same pattern in daily movement.

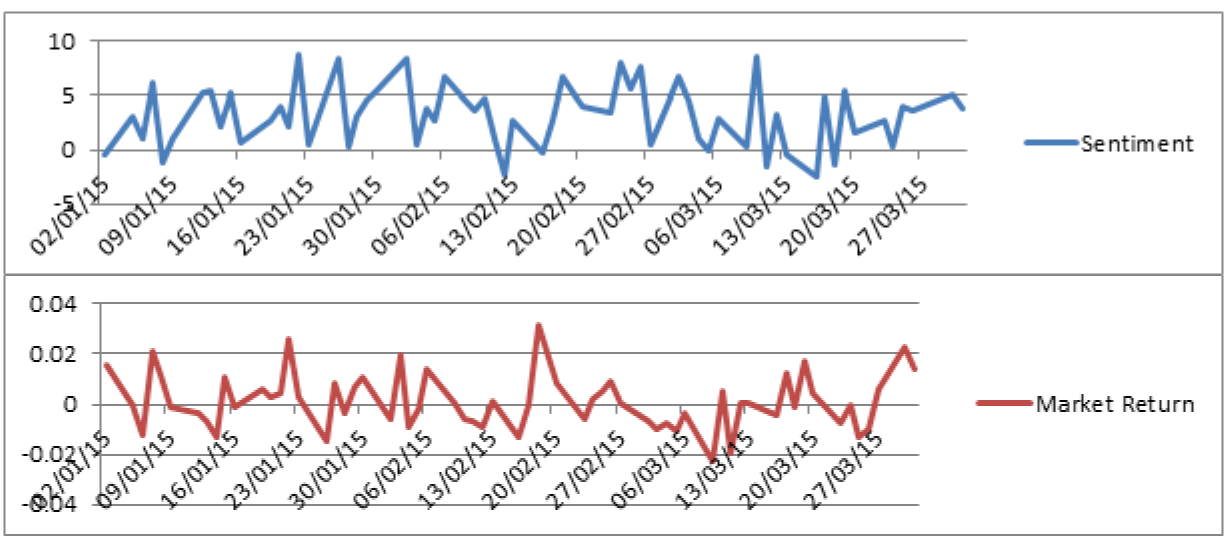

Fig. 1: Construction, Property & Real Estate Sector

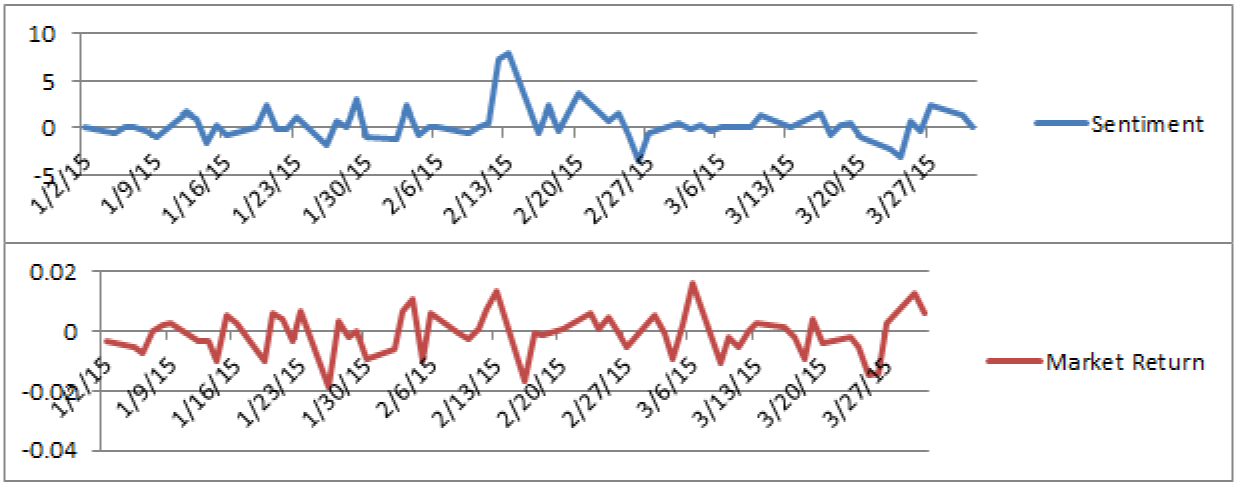

Fig. 2: Infrastructure, Utility & Transportation Sector

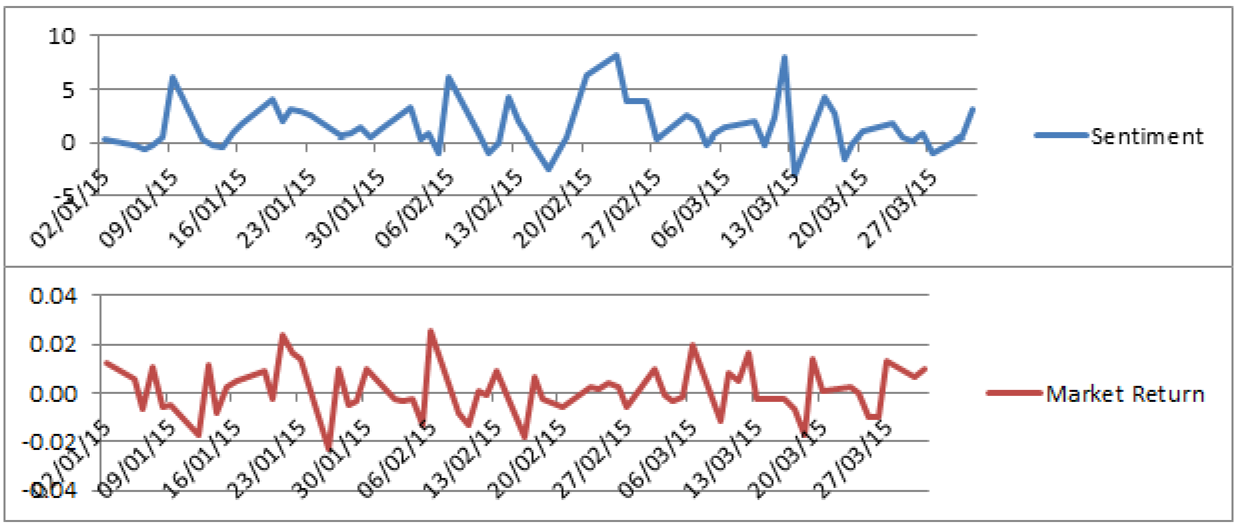

Fig. 3: Consumer Goods Sector

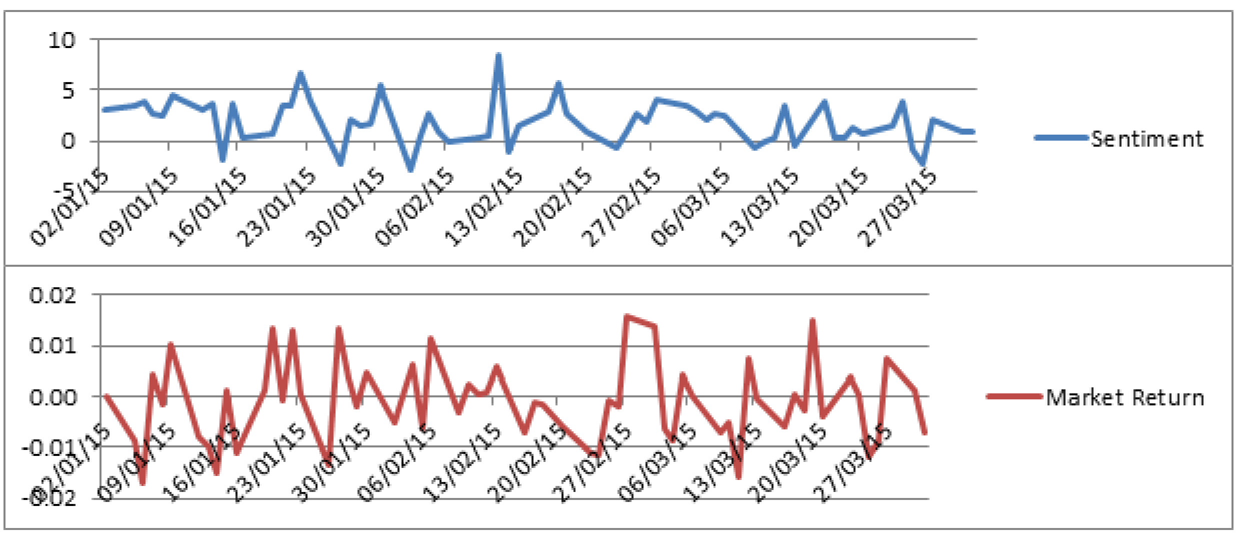

Fig. 4: Mining Sector

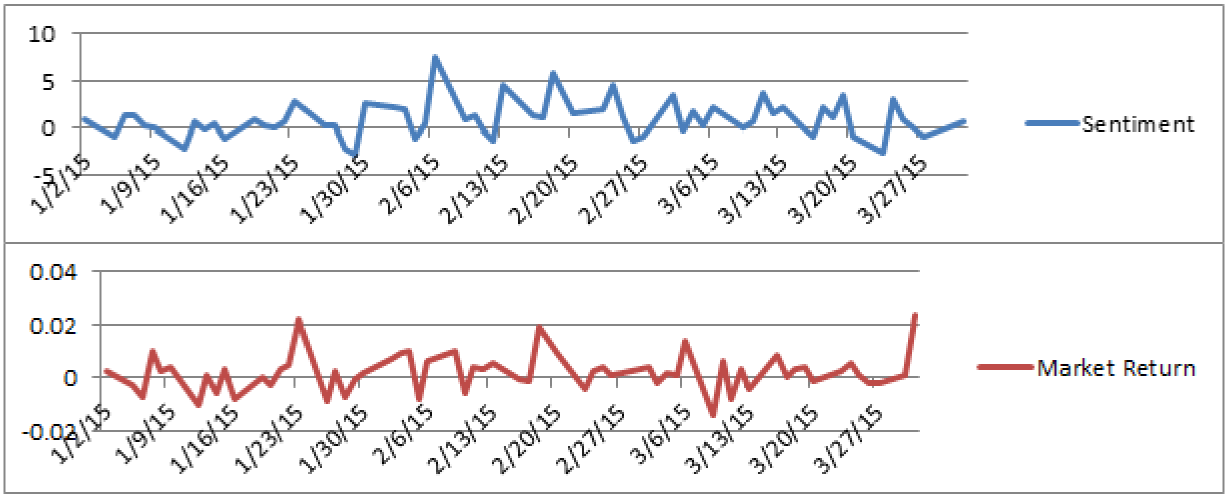

Fig. 5: Finance Sector

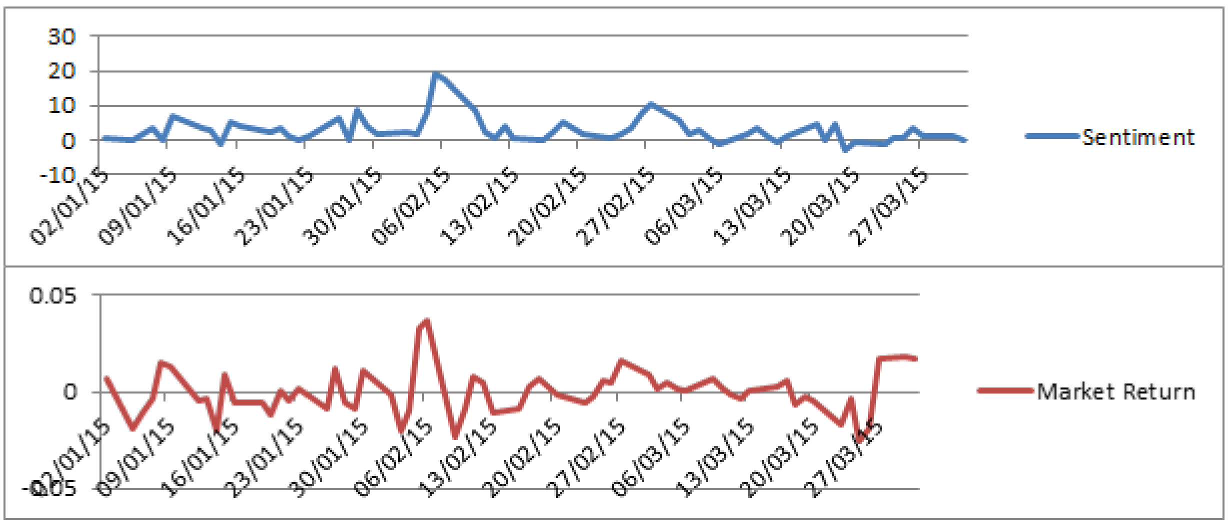

Fig. 6: Agriculture Sector

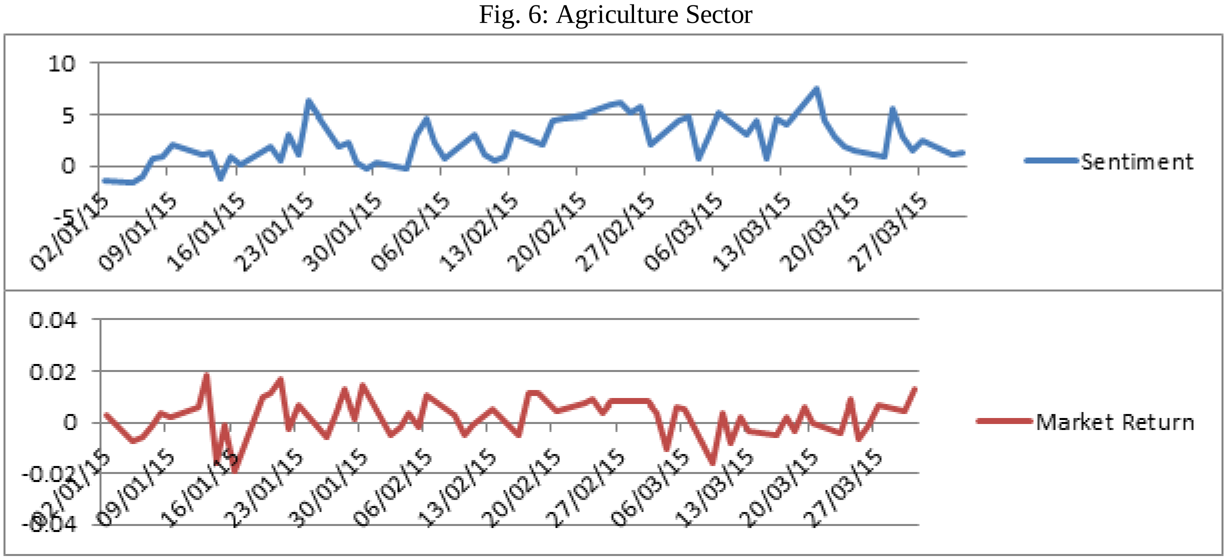

Fig.7: Trade & Investment Services Sector

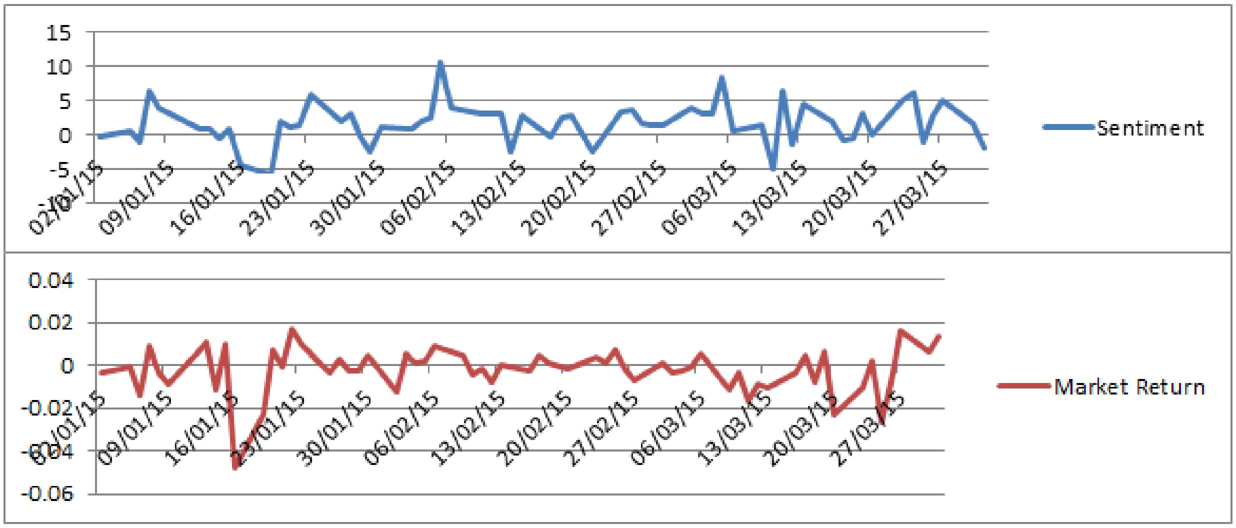

Fig. 8: Basic Industry & Chemicals Industry Sector

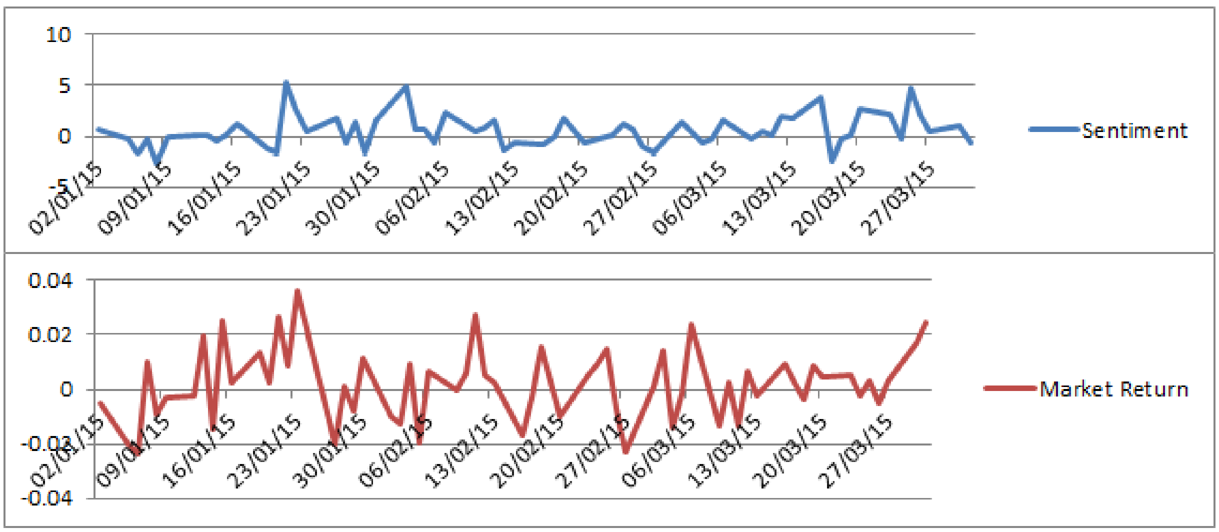

Fig. 9: Miscellaneous Index Sector

- Model Summary

After all sectors have been tested using linear regression method, we can conclude that the public sentiments do inlfluence the market returns by looking from p-value and R-square result. The higher number of R-square, the stronger that the influnce of public sentiment to market returns and more appropriate that model used.

TABLE IV: Market Sector Model Summary

No	*Symbol*	*Market Sectors*	*P-value*	*R-square*
1	JKPROP	Construction, Property & Real Estate	0,007825	0,112067869
2	JKCONS	Consumer Goods Index	0,007732	0,112389844
3	JKTRAD	Trade, Service & Investment	0,019572	0,087513390
4	JKMING	Mining Index	0,002090	0,147100908
5	JKBIND	Basic Industry and Chemicals Industry	0,003492	0,133556318
6	JKAGRI	Agriculture Index	0,001047	0,165173978
7	JKMISC	Miscellaneous Index	0,005753	0,120280356
8	JKINFA	Infrastructure, Utility & Transportation	0,003509	0,133427378
9	JKFINA	Finance Index	0,000749	0,173816249

The table above shows the number of significant level of P-value and R-square. Based on the data above, the daily sentiment model of all sectors is accepted and applicable to use on population using 95% level of confidence. Based on the R-square, the highest score occurs in Finance Index Sector. It means that finance market return as a dependent variable is 17.38% affected by social media, and 82.61% is affected by the other factor.

5. Conclusion

The main purpose of this study is to show that social media can be used to predict the Indonesia stock market investment. This study is really supported by the knowledge about Indonesia included as the three big countries which often use social media in their life. Using 95% of confident level, the result shows that all Indonesia market sectors have been proved that they are influenced by public sentiment. Athought the impact of public sentiment less than 20%, we can say that sentiment publict is really have an impact to the market returns. On the other hand, the remaining influence, 80% can be from other variables, such as politics, macroeconomics issues, etc. According to regression result, within all Indonesia financial market sector, finance index sector has the strongest influence from public sentiment, 17.38%, than the other sectors.

In future work, we suggest exploring more keywords which are suitable for Indonesia sentiment. The keywords are the main part of the result accuracy. Also, the other suggestion is other influencers; such as politics issue can also be analyzed in predicting the market returns.

6. Acknowledgements

The authors want to thank Stockbit as the main source in providing the data. This study also has been done with a big helped from my senior, Eneng Nur Hasanah. She is a Master Student at the School of Business and Management, Institut Teknologi Bandung (SBM ITB) and the one who introduces the Semantria for Excel Progam.

7. References

[1] Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. Social Computing Lab, HP Labs. http://dx.doi.org/10.1109/wi-iat.2010.63

[2] Barber, B. M., & Odean, T. (2001). The Internet and The Investor. Journal of Economic Perspectives, 41-54. http://dx.doi.org/10.1257/jep.15.1.41

[3] Bollen, J., Mao, H., & Zeng, X. J. (2011). Twitter Mood Predicts The Stock Market. Computational Science. http://dx.doi.org/10.1016/j.jocs.2010.12.007

[4] Oh, C., & Sheng, O. R. (2011). Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. ICIS 2011 Proceedings .

[5] Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. Progress in Artificial Intelligence EPIA. http://dx.doi.org/10.1007/978-3-642-40669-0_31

[6] Yu, S., & Kak, S. (2012). A Survey of Prediction Using Social Media. CoRR abs/1203.1647.

[7] Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators through Twitter. Procedia - Social and Behavioral Sciences , 55-62. http://dx.doi.org/10.1016/j.sbspro.2011.10.562