Get Some Returns through Online News Sentiment on Big Cap Market

Kezia Vincy Renaldi[1] and Deddy P. Koesrindartoto[2]

School of Business and Management, Institute Technology Bandung

***Abstract*:** *This study analyzes the impact of sentiment on online news to abnormal stock returns on some big capital market in Indonesia, such as ASII, BBCA, BBRI, and TLKM. These companies are the top two of Indonesia's big capitalization that always compete to be the first rank. The big capital company can represent the mover of market index. The data is obtained from some popular online news, specifically about stock, like Kontan, Media Indonesia, and so on for one year period; in order to achieve the amount of data needed in doing research. We transform the data from textual data into numeral data by using Semantria in Microsoft Excel program to analyze the sentiment. The use of this program for the methodology is the extent of previous study which uses the average negative word list from the articles as the sentiment. In order to give score on Semantria, the bag of words of positive, negative, and neutral, which used in Bahasa, has to be classified first by checking the keywords that appeared frequently on the online news. The result of this study is the sentiment on online news has impact to excess return on big capitalization company. Therefore, the online news gives the information to the investors that can influence them in their financial decision making.*

Keywords: *Online news, Excess Returns, Sentiment Analysis, Behavioural Finance, Regression, Big Capital Market*

1. Introduction

Everyday people hear and read both good and bad news around the world easily from the online media. The news itself becomes important information and knowledge that allow them to broaden their perspectives and at the same time, it might affect the readers on how they see things around them. The fact that world news is updated rapidly in the Internet shows that downloading the latest information every single minute becomes a habit to people nowadays because they believe information is power, and this is one of the key things they need in order to survive in the 21st century. This is clearly the evidence of how humans are very much dependent on the social media as these media have become a part of their lives today.

This also applies for the investors. In the past, most investors used to get the information from various traditional sources, like financial analysts and newspapers. These traditional sources do not give investors the freedom to receive the information wherever they are. It is very troublesome that investors have to make appointments or calls with their own financial analyst, or they have to wait for the next newspaper to get the information on the next day. Even though these two sources are able to provide predictions of future stock prices, they could be wrong. Moreover, the information from the analysts and newspapers is not the most updated because by the time the investors receive the information, the prices of stocks must have changed. In this situation, they, hence, feel uncertain in trading their own assets due to the lack of accessibility. They might lose their returns in the end since the stock movement is very fast. The stock prices can move higher or lower within seconds.

Nowadays, we allow ourselves to be surrounded by various technologies, and one of them that always revolves around us is the social media. Social media are places for people to communicate with others easily and quickly, regardless of the distance. Living with the social media becomes a lifestyle that people do not mind adopting nowadays. What we usually understand about social media is merely social networks, such as Facebook and LinkedIn. Ellison and Boyd (2007), who are researchers of the social media, define social network sites as online services that allow people to build and share a public profile, to create a connection with other users, and to explore the connections which are partly made by others as well. People, often spontaneously, share stories about their daily lives and share how they feel in social networks, and even for some, they try to gain popularity in the networks. However, those social network sites are only one form of the social media.

According to 'Out:think', other forms of the social media are bookmarking sites (e.g. Delicious and StumbleUp), social news (e.g. Digg and Reddit), media sharing (e.g. YouTube and Flickr), micro-blogging (e.g. Twitter) and blog comments and forums (e.g. Kaskus). Among these various types, there is one that is very crucial for businessmen, especially investors, in their decision making to buy stocks, and that is the social news. Hence, while the social media can satisfy humans as social beings, it is also useful for business purposes.

Social news sites are services that allow people to share posts regarding different news items, and the users of the sites are able to vote and/or comment on the news posted. These sites are the places where business people, especially investors, can get every bit of useful information about stocks. Kontan.co.id and Pasarmodal.Inilah.com are two examples of Indonesian social news site which provides news including stocks news, and both are the most commonly used website by investors and businessmen in the country. Unlike in the past, most investors today have experienced the advancement of technology as there are enough and credible sources from social news sites. The articles posted can be quickly and freely accessed anywhere and anytime, which allows investors to trade easily. Since now they are the ones responsible for their own assets, they can also avoided paying incentives to their financial analysts that can decrease returns by using the online information from the articles. This information, which is more cost-efficient than spreading questionnaires or doing surveys, can also be used in doing researches, not only in marketing, but also in financing,

In finance, there is also the study that correlates the social media with stock returns, like in 'Customers as Advisors: The Role of Social Media in Financial Markets' written by Chen, De, Hu, and Hwang (2011). This research analyses the correlation of social media, specifically online news, indicating excess stock returns by using the data and articles in Indonesia.

2. Literature Review

According to the previous study by Chen, De, Hu, and Hwang (2011), it shows that there is correlation between negative words on the online news and the abnormal return. In that study, they used two most popular online articles and list the negative word based on Loughran and McDonald list. The negative words are used for one of the independent variables that give high impact on abnormal return using linear regression. The result is that the power of online news gives high impact on stock prices and for contemporaneous and subsequent stock returns. Besides, they also find that the articles that get more attention have stronger effect from online news.

Another research is conducted by Chong Oh and Olivia R Liu Sheng (2011). They measured sentiment both manually and automatically by using some approaches. Bag of words (Schumaker & Chen, 2009) approach has the highest score among the other approaches. For the result, they stated that the posting from micro blogging have a strong value in predicting future market movements. In another study by Xue Zhang, Hauke Fuehres, and Peter A. Gloor (2010), they used mood words in a tweet and measured some emotions everyday by counting emotional words in tweet. The researchers divided the emotional words into 2 groups, which are positive words, such as hope and happy, and negative words, such as fear, worry, nervous, anxious, and upset. As the result, when the tweet shows more emotional words, such as hope, fear, and worry, the market index goes down on the next day, and vice versa. These emotional words are proven to be used when the economic is uncertainty.

The research related to social media has developed from time to time. It is proven by the methods in processing the sentiment of each research are different from one another. The common social media type that is used is micro blogging, like Twitter since its content consist of 140 characters. Being different from micro blogging, the characters of the content in online news are unlimited. In micro blogging, the sentiment can be categorized as positive, negative, and neutral, whereas online news actually categorized as neutral since the content of it do not have opinion from the people. However, online news can be classified based on the positive, negative, and neutral news, or it can be named as review, seen from the overall content.

By adopting the previous research, this research is seeking for the correlation between the sentiment on online news and excess returns. The difference between the previous is on the variable. There are some independent variables that used in indicating excess return in the previous. However, in Indonesia there is no sufficient data about the variables that used, like earnings surprise. Therefore, this research uses excess return as the dependent variable, whereas the sentiment as the independent variable. The sentiment on online news is the only independent variable in this study in order to obtain the result on how strong the correlation between them. This research is expected to prove that the sentiment, as the only independent variable, can influence the excess

return. Also, some previous studies use sentiment from micro blogging that have less characters on the content. This is the challenge using online news which has the longer content and mostly the content consist of multiple-ticker in the news. Thus, it should be clean first one by one manually in getting the exact sentiment for each stock since its measurement uses Semantria program in Microsoft Excel. Semantria program can transform textual data into the numerical data by listing the bag of words of positive, negative, and neutral. These words are from the most frequently used words in online news. This methodology is the extent from the previous that uses the average negative words on each article. In defining Indonesia's market behavior, the data price for excess return are from Jakarta Stock Exchange (JKSE), as well as the online news and keywords being used are in Bahasa. The stock that is chosen is categorized as big capitalization company, such as Astra International Tbk. (ASII), Bank Central Asia Tbk. (BBCA), Bank Rakyat Indonesia Tbk. (BBRI), and Telekomunikasi Indonesia Tbk. (TLKM), which can be represented as the mover of market index in Indonesia. These two big companies have competed to be the first rank in Indonesian market.

3. Methodology

3.1 Data of Online News and Excess Returns

The data in this research contains two sets of secondary data. The first set is for excess return. In obtaining excess return, it needs the price data of each stock. While the second set is the sentiment from online news for one year period in 2014. The stocks in this study are ASII, BBCA, BBRI, and TLKM as mentioned above. In addition, the data of online news from one company is not enough to represent the correlation. It needs another sample with the same characteristics to conclude the result. Furthermore, the more data is obtained, the better the result would be. For each company, the data of online news was collected from January 1, 2014 to December 31, 2014. The data from one online news website is not sufficient; therefore this research uses lot of website. For the data price, the period of data is the same as the sentiment, which has 245 trading days in 2014.

Data from online news have to be classified first by using the keywords before being mined. The keywords, comprise the stock ticker of each stock, the company names itself, and also the observed year, are used to collect the related articles since the data of online news are abundantly, but not all the articles are relevant with the criteria above. Besides the content, the date of each news has to be included, as well as the keywords in order to discover the most frequently words appeared. For each company, the data needed is range from 500 to 600 news.

The other set data is excess return. Excess return is defined as the returns portfolio from an investment that exceeds the market index return over a period of time. It is often used for quantifying the value added of the portfolio investors. The excess return is calculated from different return of each stock with the market returns. For the return itself, it can be measured from the price data of each stock, which obtain from Yahoo Finance (http://finance.yahoo.com/). The adjusted close price is used for calculating returns since it is affected by corporate actions, like stock dividend, cash dividend, stock splits, and right offerings in representing firm's equity value accurately. The formula in measuring returns (R_t):

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

where P_t is the adjusted close price of day t and P_{t-1} is the adjusted close price of the previous day. After measuring the return of each stock and market, excess return can be calculated by subtracting returns of stock with return of market. The result can be either positive or negative since excess return is not happened every time, but is usually triggered by some events, like mergers, acquisitions, dividend announcement, lawsuits, interest rate increases, and announcement of company earning.

3.2 Data Processing

After the data is collected, processing data requires some steps, which are:

1. Filtering the data

In spite of the online news data have been classified by keywords, the content is not 'clean' enough, which means that the content of online news can consist of multiple-ticker news. For example, this research needs the data about stock ticker ASII, but the content in the news can also be about the other stock ticker that has opposite news with the researched stock ticker. This can make false sentiment scoring. Therefore, before sentiment scoring, it must be sure to have the 'clean' data by deleting the part of the news that is irrelevant.

2. Sentiment Analysis

Sentiment analysis that is used for determining the content of the text refers to a positive, negative, or neutral. In the process of analyzing the sentiment, it needs a bag of words that represents the overall content, and those words could be positive, negative, or neutral depending on the most frequently appeared words in the content that are the same as the words that have been collected before. The score of each sentiment is +1, -1, or 0, respectively. All online news data will be input into Semantria Program in the Microsoft Excel. Also, the keywords and score have to be input in order to classify and give sentiment score on online news. Below are the tables of keywords for each sentiment category:

TABEL I: Positive Keywords

Positive (+1)				
support	'kerja sama'	beli / buy	'sell on strength' / SOS	'buy on weakness' /BOW
rebond	optimistis	menguat	uptrend	akumulasi
naik	melambung	tumbuh	gainer	rekomendasi
kenaikan	pendorong	positif	pilihan	pertumbuhan
ekspansi	penopang	penggerak	bullish	pendongkrak
dividen	melaju	mover	borong	menggenjot
meningkat	melesat	akuisisi	kuat	melampaui

TABEL II: Negative Keywords

Negative (-1)				
negatif	longsor	bearish	dibuang	jual
anjlok	dilepas	melemah	cut loss	dijual
menyusut	menekan	pelemahan	loss	sell
perlahan	penekan	downtrend	terendah	menurunkan
melorot	kurang	losers	turun	memerahkan
merosot	merah	tekor	menurun	melambat
utang	pemberat	kerugian	penurunan	kehilangan

TABEL III: Neutral Keywords

Neutral (0)						
Netral	tunggu	dicermati	program	awards	kapitalisasi	mengumumkan
hold	sideways	dipertimbangkan	proyek	penghargaan	layanan	partisipasi
tahan	perjanjian	diperhatikan	kegiatan	transaksi	fasilitas	memberikan

Most of the keywords are in Bahasa, but some of them are in English. The keywords used, be it in Bahasa or in English, are the exact same words used in the contents such that the prefixes and suffixes cannot be omitted, even though two or more keywords have the same root word which means they have similar meaning. For instance, the word 'naik' and 'kenaikan' have the same root word; therefore both of them have to be input in order to be analyzed by Semantria,

3. Linear Regression

The dates of sentiment data from Semantria have to match with the dates of trading days. The sentiment on public holidays or on the weekends should not be, and hence, should be deleted. Then, the data is pooled per five days in this research 1.1. since the number of trading days is 245 days, and being done so means that there is no remaining days. The result from Semantria, which is sentiment, is the independent variable (X), whereas the excess return is the dependent variable (Y). This research uses four data models of linear regression as the method for analyzing the correlation between the variables for each company. The models are sum X – sum Y, cumulative X – cumulative Y, average X – sum Y, and cumulative average X – cumulative Y. The formula for each model using linear regression:

$$\boldsymbol{Yi = a + \beta Xi + \varepsilon i} \quad (2)$$

The test of this method is assisted by using SPSS program. The result can be seen from p-value of each X variable. The standard value for this model is the same as the level of alpha, 0.05, with the confidence level of 0.95. Besides, it is also shown by the result of R-square that represents the influence of independent variable to dependent variable. The data must pass the assumption classic test first before measuring the correlation.

4. Result

Each company has been tested using linear regression method with four models in defining the most appropriate model. However, the cumulative and cumulative average models do not pass one assumption classic

test for all company. The rest of the models pass and have same score of p-value and R-square for each company. Due to that, this research will show the result of sum model only. This model totals up both the sentiment score and excess return per five days trading days.

Below are the movement graphs of sentiment and excess return each company that have been pooled 5 days trading days (Figure 1-4).

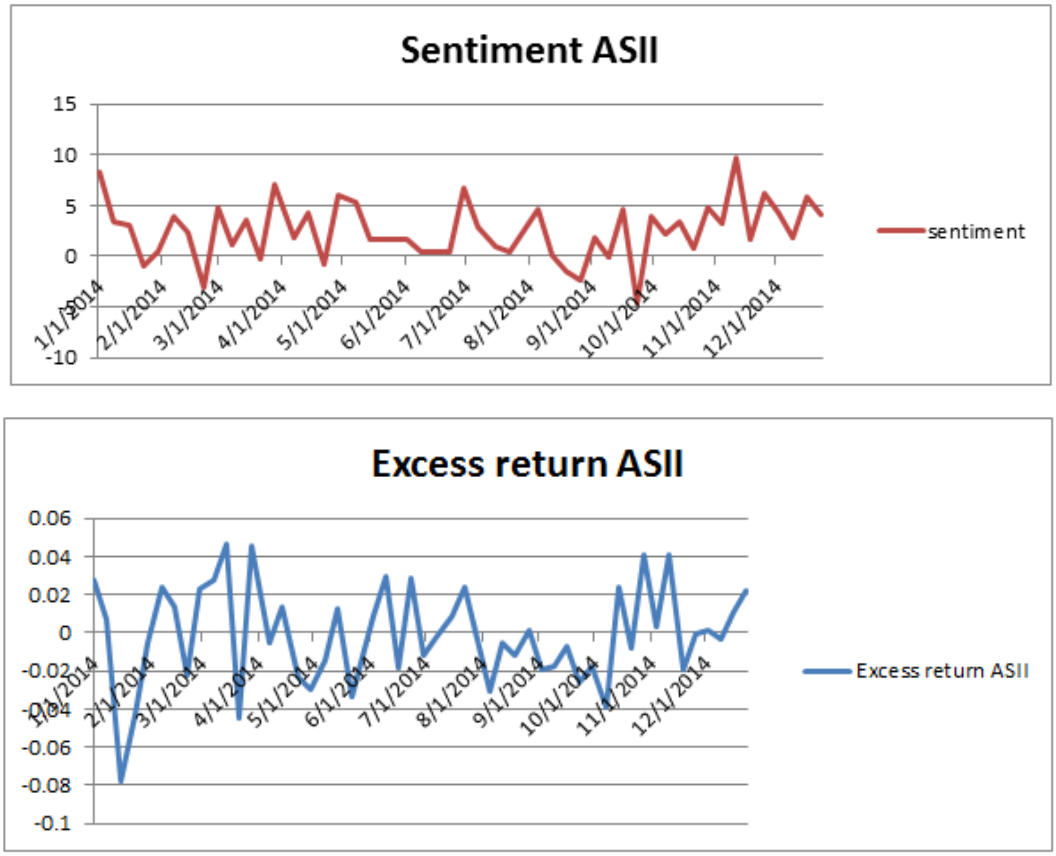

Fig. 1: Astra International Tbk. (ASII)

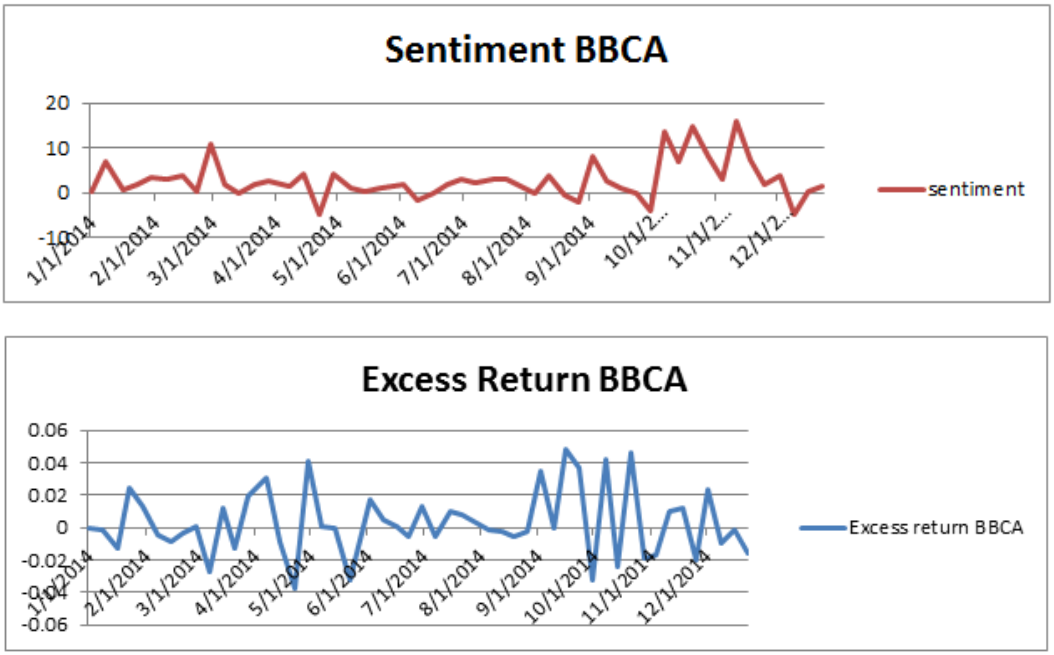

Fig. 2: Bank Central Asia Tbk. (BBCA)

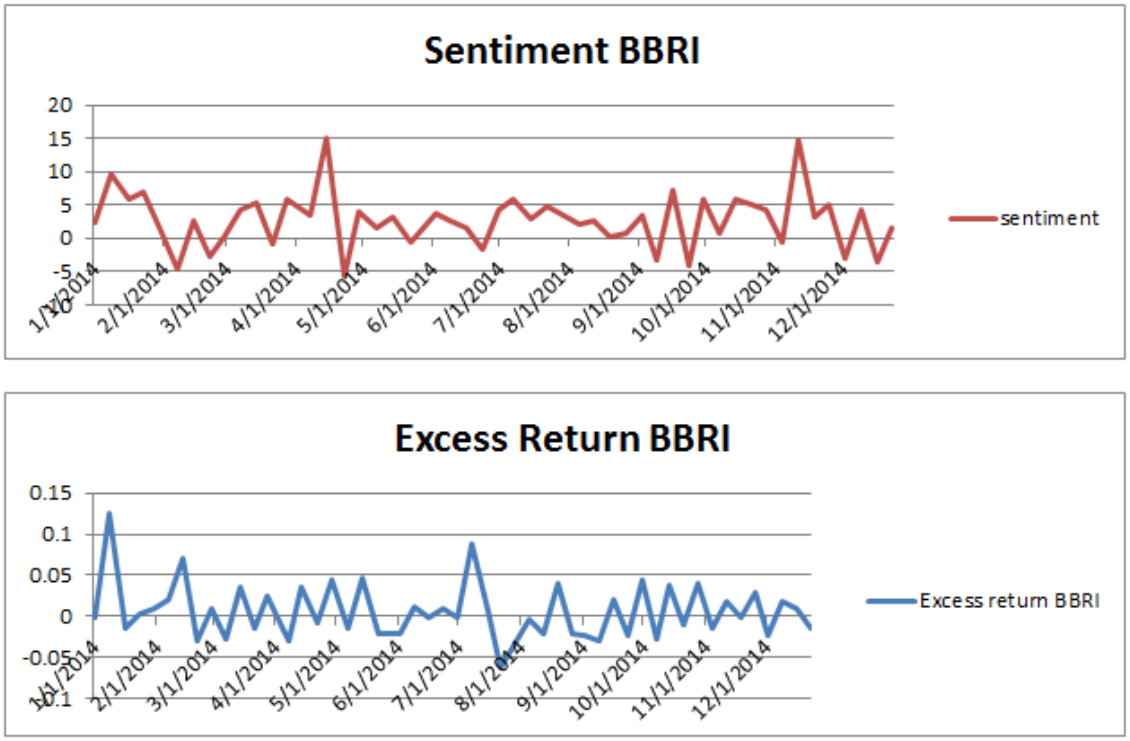

Fig. 3: Bank Rakyat IndonesiaTbk. (BBRI)

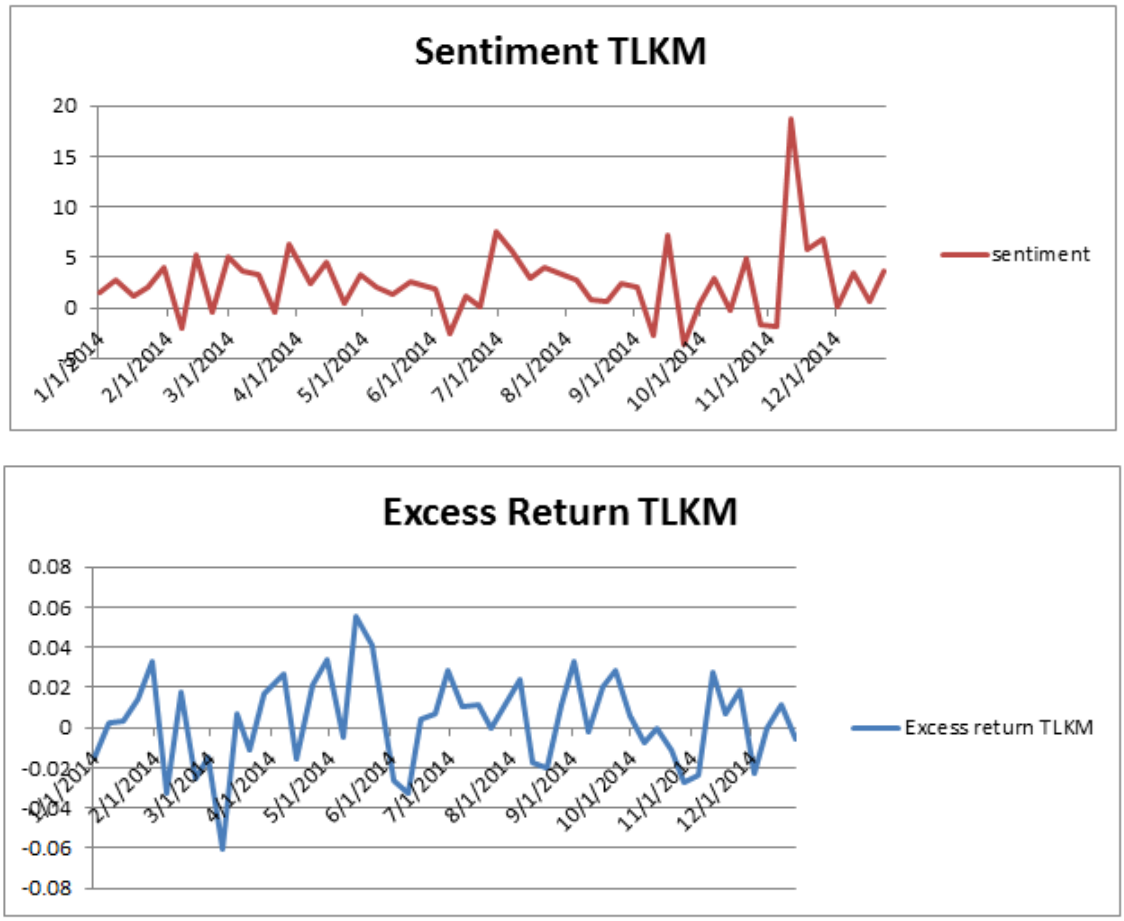

Fig. 4: Telekomunikasi Indonesia Tbk. (TLKM)

- Analysis Summary

The correlation between sentiment on online news and excess return can be seen from the result of linear regression, which is represented by the score of p-value and R-square. The higher the score of R-square, the stronger the correlation between variables, which means that the independent variable gives an impact on the dependent variable.

TABLE IV: The Result of Regression

No	Stock Ticker	Company	P-Value	R-Square
1	ASII	Astra International Tbk.	0,010370867	0,131741238
2	BBCA	Bank Central Asia Tbk.	0,004628614	0,158372313
3	BBRI	Bank Rakyat Indonesia Tbk.	0,002888975	0,173738845
4	TLKM	Telekomunikasi Indonesia Tbk.	0,019447569	0,11078636

From the table above, it can be concluded that there is correlation between the variables, with the level of confidence 0.95 on the population. The highest score of R-square is BBRI. It means that the sentiment on online news for BBRI influence the excess returns 17.37%, and the remaining 82.63% is affected by other factors.

5. Conclusion

The objective of this research is to analyze the correlation between social media, especially on online news, and the excess return of each stock. This research uses linear regression as the methodology with four different models, and the most appropriate one is sum model. The result of linear regression can be seen from R-square and p-value using 0.95 level of confidence. It shows that all big capitalization companies, represented as the mover of market index (JKSE), are proven to have a correlation with the sentiment on online news. Overall, the score of R-square is less than 20%, which means that the online news gives impact on excess return of 20%. The remaining 80% is affected by other factors, like politics and economics issues. Among the researched companies, Bank Rakyat Indonesia Tbk. has the highest correlation with the sentiment on online news of 17.37%.

For more reliable results, it is recommended to clean the data from online news since online news may have a lot of topics and it contains multiple-ticker. The most crucial factor is to get the analysis of the sentiments correct because there may be errors in processing the data using Semantria program. Other factors, such as politic and economic issues, can also be put into consideration as they can affect the excess return and hence it is better for these factors to be analyzed as well.

6. Acknowledgements

The author wants to thank my supervisor, Deddy P. Koesrindartoto, for his comments and guidance. He is an Assistant Professor at the School of Business and Management, Institut Teknologi Bandung. (SBM ITB).

7. References

[1] Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter Mood Predicts The Stock Market. Computational Science, 1-8. http://dx.doi.org/10.1016/j.jocs.2010.12.007

[2] Boyd, D. M., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. Computer-Mediated Communication, 210-230. http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x

[3] Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2011). Customers as Advisors: The Role of Social Media in Financial Markets. 3rd Annual Behavioural Finance Conference.

[4] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal, 1093-1113. http://dx.doi.org/10.1016/j.asej.2014.04.011

[5] Oh, C., & Liu Sheng, O. (2011). Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement. Knowledge Management and Business Intelligence.

[6] Zhang, X., Fuehres, H., & A. Gloor , P. (2010). Predicting Stock Market Indicators through Twitter "I hope it is not as bad as I fear". Procedia - Social and Behavioral Sciences, 55 - 62.