SOCIAL MEDIA AND ONLINE NEWS CORRELATION ON THE PERFORMANCE OF SECTOR INDEXES IN INDONESIA

[1]YOGI YOSWARA, [2]DEDDY P. KOESRINDARTOTO

[1,2]School of Business and Management Institute of Technology Bandung, Indonesia
E-mail: [1]yogi.yoswara@sbm-itb.ac.id., [2]deddypri@sbm-itb.ac.id

Abstract- The purpose of this study is specifically identifying whether the sentiments of social media and online news have positive correlation on the performance of sector indexes in Indonesia. This study compiles public's conversations in social media and the financial articles provided in online news about several companies or stocks during 3 months (January – March 2015), and then their sentiments will be measured by using developed keyword in Bahasa. The focuses on this study are only two sector indexes in Indonesia: financial sector (JKFINA) and infrastructure, utility, and transportation sector (JKINFA). The contribution made in this study is the data of each company are pooled into one sector, so the performance of sector indexes will be the outcome variable. The results of this study are (1) sentiments from social media and online news have positive correlation on the performance of sector indexes, (2) the correlation of social media's sentiments is approximately similar with the online news' sentiments, and (3) if both sentiments (social media and online news) are combined, they generate the stronger correlation.

Keywords- social media, online news, performance of sector indexes, behavioral finance, sentiment analysis

I. INTRODUCTION

On the past decade, people chose to use traditional media (e.g., brochures, television, newspaper, and radio) to obtain information. However, the development of internet as a new information technology has changed people's tendency to gather information via online, such as websites, online news, social media, etc. Based on the research of e-Marketer published in November 2014, the netters in Indonesia are 83.7 million people during 2014. Indonesia ranks sixth of internet users in the world and e-Marketer1 predicts that the amount of them in 2017 are going to be 112 million people, beating Japan on the fifth rank. The internet growth creates the emergence of online news and social media channels (e.g., microblogging, online forums, online social networking, etc.) as the popular tools in term of gathering information. As the result of changes in people's attitude toward social media, many businesses start to concern and try to utilize the potential prospect of it. Social media channels are used in term of marketing purposes, such as to advertise products or events, directly interact with customers, reach recent and potential customers, and recognize customers' needs and wants. In addition, social media is also used to measure company and competitors' performance, generate investment decision, and forecast the future business value [1], [2]. The availability data of social media and online news has also attracted people to do research about them, not only industrial or company research but also academic research [1]. There are various topics in marketing studies that can be explored, even in this recent year researchers start to explore social media toward financial studies. That condition arouses the author to conduct a study about social media and online news in particular for cases in Indonesia. In addition, the researches about social media toward financial studies are still not common in Indonesia.

The objective of this study is identifying whether the sentiments of social media and online news have positive correlation on the performance of sector indexes in Indonesia, specifically financial sector (JKFINA) and infrastructure, utility, and transportation sector (JKINFA). The author adds online news as the other variable because companies tend to only focus on social media and ignore online news, whereas online news is major source of market information which are more stable and trustworthy [3],[4]. Also, this study does not use specific companies because the data of each company are pooled into one sector. It is conducted because the data of one company are limited, as we have known that there are still a few Indonesian people who are stock traders (about 484.688 people of total population based on the research of Pustatis2 in March 2015). The pooled data will be the contribution of this study because we can know the overall performance of the sector indexes. The outcome of this study is beneficial because top executives or businesses can consider about utilizing and maintaining company information in social media and online news as if there is a strong correlation. Also, this study can be used by businesses to determine the appropriate marketing or financial strategies in social media or online news channels.
The structure of this paper is: the related previous researches are reviewed in Section 2. Section 3 focuses on the data and methodologies used to conduct this study. Section 4 and 5 provide the results of the experiment and limitation. Finally, the conclusion of this study is discussed in the last session.

II. LITERATURE REVIEW

The emergence of big data in the digital social media channels or online news outlets has aroused the researchers to conduct many studies about them. The

research of [5] has proofed that social media contents could predict the real world outcomes. By mining Twitter's contents, they generated the prediction of box-office revenues for movies. Besides, social media data are also used to conduct financial researches with different sources and methodologies, specifically about stock or financial market. The researchers analyzed the market information to obtain the better profit [4]. This study is related to the several previous researches that utilize social media sentiments to examine financial market. The research of [6] analyzed the correlation between mood derived from Twitter contents and the value of the Dow Jones Industrial Average (DJIA). They utilized OpinionFinder to measure and classify positive or negative mood and Google-Profile of Mood States (GPOMS) to state mood into 6 mood dimensions. The result reported that there was a strong correlation between measurement of mood from Twitter's contents and the DJIA's values with an accuracy of 87.6%. The similar research was also conducted by [7] that tried to analyze Twitter's posts to make prediction on the stock market indicators. They measured investor's fear with the external benchmark Chicago Board Options Exchange Volatility Index (VIX), selected baselines (the number of tweets per day, followers per day, and retweets per day) about a certain topic, and identified how much the Twitter's conversations precedes stock market's fluctuations. The result showed that the emotional tweets have negative correlation with stock market indicators, but have positive interaction effect to VIX. The [10]'s research compared the cumulative emotional valence (positive or negative emotional sentiment) from Twitter contents about companies in the S&P 500 with the average daily stock market returns of those companies by using regression method. The result of this research showed that there was a significantly relationship between the cumulative emotional valence of Twitter posts to the companies' stock market return. The sentiment from Twitter's users that have a lot of followers affected to the return on the same day, however, the users that
have a few followers provided an impact to the future stock return. Besides using Twitter as the data source, [8]'s research tried to look for the relationship between the sentiments based on Facebook's status update and stock market return by using Facebook's Gross National Happiness Index. The result of this research showed that the sentiment of Facebook's status update had positive relation to the stock returns.

The other researches also measured the sentiments of financial news. For instance, [4]'s research experimented whether financial news article affect to the stock price return on Hong Kong Stock Exchange. They evaluated and compared six different approaches: two approaches which used sentiment dictionary (Harvard psychological dictionary and Loughran – McDonald financial sentiment dictionary), one approach which used SenticNet, one approach which used bag-of-words, and two traditional approaches which are according to the news sentiment's polarity asymmetry. The results were sentiment analysis performed better in both validation and independent testing data sets and help improve the prediction accuracy. Besides, [3] conducted a research about investigating whether the authors' words and tone in the case of financial news articles affected to stock price movements by using Arizona Financial Text (AZFinText) system. They found that the price direction was easily predicted by articles with negative sentiments (author's tone) and subjective news articles (author's words use).

From all researches about social media and online news on financial market, this study is most closely related with [1]. They compared social media and conventional media on firm equity value with the sample of 824 public traded companies in 6 industries. Naïve Bayes algorithm was used to measure the sentiments and the result was social media and conventional media had strong correlation on stock performance. Also, social media had the stronger correlation if it was compared by conventional media. However, this study creates several contributions and differences from that previous research. First, this study analyzes the performance of sector indexes as the outcome variable. Second, the data of each company are pooled into one sector. Third, this study analyzes Indonesia's indexes, while another researches used U.S.' indexes. As we known that the research about it is still not common in Indonesia. Fourth, this study uses different sentiment measurement, Semantria for Microsoft Excel3 to conduct the sentiment analysis. Fifth, this study uses different sources and media that are well-known in Indonesia.

Based on the previous researches that have been discussed, the hypotheses of this study are:

H0 = the sentiments of social media and online news do not affect the performance of sector indexes

H1 = the sentiments of social media and online news affect the performance of sector indexes

III. METHODOLOGY

3.1. Data Collection

The stocks traded in Indonesia Stock Exchange (IDX) are classified by Jakarta Exchange Industrial Classification into 10 sectors based on their industries. This study is focused on two sector indexes: financial sector (JKFINA) and infrastructure, utility, and transportation sector (JKINFA); therefore, the sample of this study are 130 companies included in those sectors.

Table I: Sample of This Study

Sector Indexes	Amount of Companies
Financial Sector	78
Infrastructure, Utility, and Transportation Sector	52
Total	130

There are three sets of data in this study obtained during 3 months (January 1st to March 31st, 2015). The first set is indexes' returns per day of each sector obtained from Yahoo Finance4. The return is calculated by using this formula:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \qquad (1)$$

where denotes the return of day t, denotes the closed price of day t, and denotes the closed price of the previous day. The second set is 1,975 Stockbit's feeds related to companies in those sectors during three months. Stockbit5 is an Indonesia's stock community or forum where traders and investors discuss or share idea about stocks traded in Indonesia. Compared with Twitter, Stockbit has a system to reduce noise by using the sign of $ before the stock's code (e.g. $BBNI), so people can search and find a specific stock easily and accurately. Also, the discussions in Stockbit are specifically about stocks, economy, or financial market. The third set is 2,000 online news feeds about companies in those sectors. The articles are obtained from several well-known financial online news channels in Indonesia, such as Kontan, Inilah, Berita Satu, MetrotvNews, detik, Bisnis, Kompas, and SindoNews.

3.2. Data Filtration

The data should be filtered first by using this formula: =CLEAN (data) in Microsoft Excel to delete the punctuation marks. Also, the data feeds of
Stockbit and online news which are not related to the companies in one sector should be removed. For example, in the case of financial sector, there is a feed "I recommend you to buy $BBNI and sell $INDF". The words of "sell $ADRO" should be removed because $ADRO is not the company which included in financial sector. (Note: BBNI is the code of "Bank Negara Indonesia" which is included in financial sector, whereas ADRO is the code of "PT Adaro Energy" which is included in mining sector). That way is conducted to make more accurate measurement of sentiment scores.

3.3. Sentiment Analysis

This study uses Semantria for Microsoft Excel to measure the sentiment of each feed in a given day. Some keywords are developed and inputted to Semantria where the score of 1 indicates positive sentiment, 0 indicates neutral sentiment, and -1 indicates negative sentiment. The keywords of positive, neutral, and negative sentiments are determined from the frequency of positive, negative, and neutral words appearing in the data compilation. Below are the tables of positive, neutral, and negative keywords inputted to Semantria (Table II – VII).

Table II: Social Media's Positive Keywords (1)

	Indonesian Words		
Increase	terbang, hijau, naik, tambah, tingkatkan, melejit, melesat, menguat, melonjak		
Nice / Good	bagus, mantap, ganteng, luar biasa, cakep, asik, maknyus		
Take / Collect	ambil, borong, tampung, koleksi, panen, akumulasi		
Buy	beli	English Keywords	Uptrend
Rebound	melambung		Golden Cross
Profit	untung		Bullish

Table III: Social Media's Neutral Keywords (0)

Hold	Tahan	Look at	Cermati, simak, perhatikan
Favorite	Favorit	Queue	Antri
Let us	Ayo	Portfolio	Portfolio

Table IV: Social Media's Negative Keywords (-1)

	Indonesian Words		
Decrease	turun, anjlok, longsor, merosot, ambrol, ambles, hancur, jeblok, terperosok, terjungkal, payah, banting, jatuh, lemah		
Discard / Dispose	buang, lepas, sayonara, pangkas		
Stuck	nyangkut, nyungsep, terjebak	English Keywords	Downtrend
Sell	jual		Cut Loss
Loss	rugi, gagal		Bearish

Table V: Online News' Positive Keywords (1)

	Indonesian Words		
Increase	menguat, apresiasi, tumbuh, naik, melejit, melesat, melonjak		
Develop	mengembangkan, membangun		
Acquisition	Akuisisi		
Expansion	ekspansi	English Keywords	Uptrend
Recommen-dation	rekomen-dasi		Rebound
Inject	suntik		Top Gainers

Table VI: Online News' Neutral Keywords (0)

Consider	Pertim-bangkan	Look at	Cermati, simak, perhatikan
Debt	Utang	Target	Target
Dividend	Dividen	Potential	Potensi

Table VII: Online News' Negative Keywords (-1)

	Indonesian Words		
Decrease	lemah, merosot, redup, lambat, turun, jeblok, anjlok, merosot, melorot		
Cut	memangkas		
Loss	rugi	English Keywords	Downtrend
Suspension	suspensi		Top Losers

The main challenge of this study is creating the keywords, specifically in Bahasa (Indonesia's language). First, there are so many words or phrases used in term of stock discussions that are not suitable with the real meaning. For example, the words of "terbang" (fly), "ganteng" (handsome), "maknyus / enak" (delicious), "hijau" (green) indicates the stock is good or growing up. Second, Bahasa has a lot of suffix or affix uses, such as "rekomendasi" (recommendation) can be written "direkomendasi", "direkomendasikan", "merekomendasikan", etc. Third, one word can be written by Indonesian people into several words (slang words) but they still have same meaning, such as "hijau" (green) can be written "ijo", "hejo", "hijo", etc. Fourth, all keywords are not developed in Bahasa because there are some words in English used by Indonesian people in their daily stock conversations, such as "golden cross", "cut loss", etc. After running the Semantria for Excel software, we obtain the sentiment data of each feed. Then, the sentiments in a given day are pooled by summing all sentiments in that day. The sentiments on Saturday and Sunday are included into Monday because sentiments on those days affect the stock return on Monday [8].

3.4. Regression Model

The sentiment measurement results from Semantria for Microsoft Excel will be the independent variables and the return of sector indexes will be the dependent variable. Then, linear regression method is conducted to test the correlation
between two variables. The tests are divided into three parts. First, test the correlation between social media's sentiments and return of sector indexes. Second, test the correlation between online news' sentiments and return of sector indexes. Third, test the correlation between the social media and online news' sentiments and return of sector indexes. The first and second test uses simple linear regression method, whereas the third test uses multiple regression models. Below are the formula of simple linear regression and multiple regression models:

Simple linear regression:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2)$$

where Y_i denotes dependent variable for observation i, β_0 denotes Y intercept for the population, β_1 denotes the slope for the population, X_i denotes independent variable for observation i, and ε_i denotes random error in Y for observation i [9].

Multiple regression models: (assuming with k independent variables)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \quad (3)$$

where β_k is the slope of Y with variable X_k, holding variables $X_1, X_2, \ldots, X_{k-1}$ constant [9].

In the regression model, we see the amount of R-square (R2) which measures the proportion of the regression model represents the linear relationship between both variables. The larger R-square indicates that there is a significantly relationship between both variables with less residual variance. Also, we see the p-value that shows the significance level of each independent variable toward the dependent variable. In addition, the classical assumption tests are necessary to be implemented as a requirement before conducting the regression model. The tests which consist of normality test, heteroscedasticity test, multicollinearity test, and autocorrelation test are processed through Statistical Product and Service Solution (SPSS).

IV. RESULTS

4.1. Social Media (SM)

The result of social media sentiments can be seen in Table VIII and Table IX. The p-value in JKINFA is 0.001 and the p-value in JKFINA is 0.006 which both of them are less than 0.05, indicates that social media sentiments in JKINFA and JKFINA affect the performance of their indexes. Then, the regression equation in JKINFA will become:

$$y = -0.001 + 0.002\ SentSMI \quad (4)$$

and the regression equation in JKFINA will become:

$$y = 0.001 + 0.001\ SentSMF \quad (5)$$

Those equations means the change in JKINFA and JKFINA returns will be affected by daily social
media sentiments. Every single change in number of social media sentiments will increase the return of JKINFA for 0.2% and JKFINA for 0.1%.

Table VIII: Summary of JKINFA's SM Sentiments

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-.001	.001		-1.657	.103
SentSMI	.002	.000	.400	3.378	.001
a. Dependent Variable: Return					

Table IX: Summary of JKFINA's SM Sentiments

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.001	.001		0.964	.339
SentSMF	.001	.000	.344	2.834	.006
a. Dependent Variable: Return					

4.2. Online News (ON)

Table X and Table XI show the result of online news sentiment toward the performance of sector indexes. The p-value in JKINFA is 0.009 and the p-value in JKFINA is 0.019 which both of them are less than 0.05, indicates that online news sentiments in JKINFA and JKFINA affect the performance of their indexes. Then, the regression equation in JKINFA will become:

$$y = -0.002 + 0.001\,SentONI \qquad (6)$$

and the regression equation in JKFINA will become:

$$y = -0.002 + 0.001\,SentONF \qquad (7)$$

So, the daily online news sentiments provide positive effect on the change in their indexes. Every single change in number of online news sentiments will increase the return of JKINFA and JKFINA for 0.1%.

Table X: Summary of JKINFA's ON Sentiments

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-.002	.001		-2.286	.026
SentONI	.001	.000	.329	2.700	.009
a. Dependent Variable: Return					

Table XI: Summary of JKFINA's ON Sentiments

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-.002	.002		-.966	.338
SentONF	.001	.000	.297	2.407	.019
a. Dependent Variable: Return					

4.3. Combination of Social Media & Online News

To be more interesting, this study combines two independent variables (social media and online news sentiments) by using multiple regression models. Table XII and Table XIII show the results of this model. The p-values in JKINFA are 0 and 0.002, whereas the p-values in JKFINA are 0.032 and 0.010. Both of them are less than 0.05 which means the combination of social media and online news affects the performance of JKFINA and JKINFA indexes. The regression equation in JKINFA will become:

$$y = -0.003 + 0.002\,SentSMI + 0.001 SentONI \qquad (8)$$

and the regression equation in JKFINA will become:

$$y = -0.002 + 0.001 SentSMF + 0.001\,SentONF \qquad (9)$$

The equations mean that the change in indexes return will be affected by daily sentiments of social media and online news. Every single change in sentiments of social media will increase the return of JKINFA for 0.2% and also increase for 0.1% every online news sentiments. Besides, return of JKFINA will be increased by 0.1% every single change in sentiments of social media. The same result is also occurred by every single change in sentiments of online news.

Table XII: Summary of JKINFA's SM & ON Sentiments

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-.003	.001		-3.285	.002
SentSMI	.002	.000	.424	3.849	.000
SentONI	.001	.000	.358	3.248	.002
a. Dependent Variable: Return					

Table XIII: Summary of JKFINA's SM & ON Sentiments

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-.002	.002		-1.337	.186
SentSMF	.001	.000	.313	2.644	.010
SentONF	.001	.000	.260	2.197	.032
a. Dependent Variable: Return					

4.4. R-square Comparison

This study also compares the R-square results from each variable toward the performance of sector indexes. In the case of social media sentiments, the regression models have 16% on JKINFA and 11.8% on JKFINA of R-square. That means 16% and 11.8% of return on JKINFA and JKFINA indexes can be represented by social media sentiments. Besides, online news sentiments represent 10.8% and 8.8% of return on JKINFA and JKFINA indexes. Online news sentiments also generate positive and approximately similar R-square if they are compared with social media sentiments. If both sentiments are combined, they generate the higher amounts, which are 28.7% on JKINFA and 18.5% on JKFINA. People perhaps feel odd with the small percentage of R-square, but those percentages are quite good. Those percentages are only influenced by sentiments of social media and online news. There are another factors that are not covered to determine the increase or decrease of sector indexes return, such as companies' financial statements, political issues, U.S. and Indonesia' economy condition, etc.

Table XIV: Summary of R-square

Indicator	JKFINA	JKINFA
Social Media	11.8%	16%
Online News	8.8%	10.8%
Both of them	18.5%	28.7%

V. LIMITATION

There are some limitations in this study. First, this study is focused on two sector indexes: financial sector (JKFINA) and infrastructure, utility, and transportation sector (JKINFA). Second, the data collection period is also limited about three months (January to March 2015). Third, this study only uses Stockbit as the source of social media data and some financial online news that exist in Indonesia as the source of online news data. Last, the majority of keywords are developed in Bahasa because Indonesian people tend to use Bahasa in generating contents in social media or online news.

CONCLUSION

This study is investigating whether the sentiments of social media and online news have positive correlation on the performance of sector indexes in Indonesia. In summary, the results of this study are: (1) sentiments from social media and online news have positive correlation on the performance of sector indexes, (2) the correlation of social media's sentiments is approximately similar with the online news' sentiments, and (3) if both sentiments (social media and online news) are combined, they generate the stronger correlation. For the future research, the researchers should develop the keywords, specifically in Bahasa, so we can create the more baskets of keywords in Bahasa. Also, the keywords should be added by bilingual languages (Bahasa and English). Keywords are the main part of this study. The measurement of sentiments can be conducted by using more advanced software, technique, or method to obtain the more accurate result. The next researches have to try to analyze the other financial outcome variable.

REFERENCES

[1]. Yu, Y., Duan, W., & Cao, Q, "The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment
[2]. Analysis Approach," Decision Support Systems, vol. 55(4), pp. 919–926, 2013.
[3]. Luo, X., Zhang, J., & Duan, W., "Social Media and Firm Equity Value. Information Systems Research, vol. 24(1), pp.146–163, 2013.
[4]. Schumaker, R.P., Zhang, Y., Huang, C., & Chen, H., "Evaluating Sentiment in Financial News Articles," Decision Support Systems, vol. 53, pp. 458-464, 2012
[5]. Li, X., Xie, H., Chen, L., Wang, J., & Deng, X., "News Impact on Stock Price Return via Sentiment Analysis," Knowledge-Based Systems, vol. 69, pp.14-23, 2014.
[6]. Asur, Sitaram, and Bernardo A. Huberman. "Predicting the Future with Social Media." arXiv preprint arXiv:1003.5699, 2010.
[7]. Bollen, J., Mao, H., & Zeng, X., "Twitter Mood Predicts the Stock Market," Journal of Computational Science, vol. 2(1), pp.1–8, 2011.
[8]. Zhang, X., Fuehres, H., & Gloor, P.A., "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear," Social and Behavioral Sciences, vol. 26, pp. 55-62, 2011.
[9]. Siganos, A., Nanos, E. V., & Verwijmeren, P., "Facebook's Daily Sentiment and International Stock Markets," Journal of Economic Behavior & Organization, vol. 107, pp. 730-743, 2014.
[10]. David M. Levine, T.C., 2013, "Business Statistics: A First Course," New Jersey: Pearson.
[11]. Sul, H. K., Dennies, A. R., & Yuan, L. I, "Trading on Twitter: The Financial Information Content of Emotion in Social Media. In: Hawaii international conference on system sciences (HICSS), pp. 10, 2014.

★★★