Sentiment Analysis in Twitter on Stock Performance in Indonesia

Selvi Lesmana Putri[1]*, Arga Hananto[1]

[1]*Department of Management, Faculty of Economics and Business, Universitas Indonesia, Indonesia*

ABSTRACT

This paper is aimed at investigating the correlation between social media Twitter and stock market performance by looking at industry-level perspective to specific companies incorporated in the. In the context of investment, Twitter has become an alternative source of information to answer the needs of investors to be able to keep abreast of news factually and practically. Therefore, we are interested in applying the same research objectives to the stock of several companies in Indonesia that are incorporated in the IDX30 index. This stock index is the illiquidity sub-category of the Headline Index, which contains companies that are used as a reference to describe the stock market conditions that meet the main criteria of having a high level of liquidity. To support this research, we collected some opinions obtained from Twitter as a source of streaming data using Python programming, and Thomson Reuters to obtain information on company performance in the form of stock prices, volumes, and market capitalization. Research models are built based on the Amihud Illiquidity method and volatility calculation to measure the correlation between sentiment analysis and stock performance. This research shows that sentiment analysis of statements uploaded on Twitter has insignificant correlation to the liquidity and volatility of IDX30 stock in Indonesia. Nevertheless, this research has not been able to separate between tweets which are generated based on user opinion and tweets which are made based on requests from certain market participants to influence the value of shares by spreading biased information to provoke a public reaction.

Keywords: sentiment analysis; social media; risk investment; stock investment; volatility; liquidity; Amihud Illiquidity.

1. INTRODUCTION

Rothman (2019) states that Twitter's short format allows for informal collaboration in the process of disseminating information. This is aligned with the use of Twitter for scientific studies to examine the relationship or influence of Twitter sentiment analysis on stock movements. Specifically, Mao *et al.* (2012) through their study using Twitter as a research component because they see the character of social media which is used widely and in real-time. The results of their study found a correlation between statements on Twitter on daily trading in several companies belonging to the Standard & Poor's 500 (S&P 500) group. Furthermore, this study also shows that data obtained from Twitter can be used accurately to predict the rise or fall of S&P 500 stock prices. In a separate study, Ranco *et al.* (2015) through his research revealed a significant correlation between Twitter sentiment on the stock prices of several companies incorporated in the Dow Jones Industrial Average (DJIA). The study looked at the relevance of Twitter sentiment with the movement of the return value both under normal conditions and when entering a certain period through the implementation of the statistical event study method.

In Indonesia, there are a number of journals that raise research on the relationship between social media sentiment on stock movements. Mufidah (2018) examines the influence of Stockbit social media sentiments on companies incorporated in LQ45 through the Ordinary Least Square (OLS) and Generalized Linear Model (GLM) methods which the results of the study show that positive and negative sentiments have an influence on the value of stock returns and trading volume. According to the existing research, we are interested in

* *Corresponding author:selvi.mmui2018@gmail.com*

applying similar research objectives with focus on social media Twitter and the stock of several companies in Indonesia that are incorporated in the IDX30 index. This stock index is a sub-category of the illiquidity of the headline index containing companies that describe the state of the stock market that meets the main criteria that have a high level of liquidity. In order to support this research, we created machine learning that is able to recognize a number of words in Indonesian and define them into groups of words that have positive or negative sentiments to be able to produce sentiment analysis values that are close to accurate results.

2. LITERATURE REVIEW

2.1 Liquidity and Volatility

Sarr *et al.* (2002) mention several main characters that need to be considered in assessing stock liquidity, namely the relatively low stock transaction costs, the number of shares requested, the availability of shares in the market, how fast the buying and selling process occurs, and the level of elasticity in maintaining balance stock request. Lipson *et al.* (2009) also highlight that transaction costs as an aspect of liquidity reflected by the capital adequacy of the company. Chordia *et al.* (2000) mention that there is a significant influence between the value of market equity, short-term interest rates, market volatility on liquidity, and stock trading activities. Domowitz *et al.* (2001) also explain that the high level of volatility of a stock can have an impact on decreasing company turnover which indirectly affects the transaction costs as well as the value of the stock returns themselves.

2.2 Twitter Sentiment Analysis

Twitter is a form of social media that is present in the form of microblogging that allows users to express expressions, opinions, information along 280 characters including links from other sites that are used as a reference. Ortega (2017) examines that Twitter is used as a journal publication media because it can increase the visibility of articles and increase opportunities for their journals to become a reference for other academics. This statement indirectly supports several studies that use Twitter as a source of data to study the relationship of social media to market behavior. The type of users is not only limited to individuals, but also company organizations in various industrial sectors and even academics. (Ranco *et al.*, 2015) through their study find out that Twitter data in aggregate in a certain period has strong implications for financial markets which can later be developed to predict the evolution of the market direction.

Referring to Liu (2012), sentiment analysis is a study that outlines the opinions, sentiments, evaluations, behavior, and emotions of a person expressed in written sentence form. Audrino *et al.* (2020) specifically studied the impact of sentiment analysis on stock market volatility and found that attention and sentiment variables can significantly increase the predictions of volatility. I have also done a similar study, which measures Twitter sentiment analysis of S&P 500 shares, which is a stock index consisting of several companies with large capital in the United States. This study concluded that tweets from social media users with even the smallest number of followers turned out to have a significant influence on changes in the value of stock returns in the next period.

In addition, Sul *et al.* (2017) have also conducted a similar study that measures Twitter sentiment analysis of S&P 500 company shares which is a stock index consisting of a number of companies with large capital in the United States. This study inquires that tweets from social media users with even the smallest number of followers turned out to have a significant influence on changes in the value of stock returns in the next period. Investors who use the stock trading strategy using this information have proven to have achieved significant achievements which is around 11- 15% in a year.

3. RESEARCH METHODOLOGY

3.1 Research Design

In this research, we use non-probability sampling data to retrieve a number of company stock data that represent the industrial sectors incorporated in the IDX30 index with the highest capitalization value in 2019. We apply the same filter condition to retrieve tweets and stocks by using stock codes and within the period of September to December 2019. The results of processing from each data group produce Twitter sentiment analysis as an independent variable that affects the value of liquidity which is calculated using the Amihud Illiquidity method and volatility which is calculated using a standard deviation. The results of the processing are analyzed with a descriptive method to determine the distribution of data. Subsequently, we conducted a panel data regression analysis to determine the correlation between the independent variables and the dependent variable that has been set. In this study we use panel data because the types of data sources are used in the form of cross sections and time series.

In order to get the proper regression model which is suitable for this research, we evaluate three types of modeling: Common Effect Model, Fixed Effect Model, and Random Effect Model. The initial step is comparing the results of the Common Effect Model with the results of Fixed Effect Model using Chow test on both models. Then, the results are compared with the Random Effect Model using the Hausman test. After obtaining the relevant regression model, we implement the model in all data for regression testing purposes before proceeding to the hypothesis test. This test includes F-test and t-test to measure significance of independent variables to the dependent variables so that effective conclusions can be drawn from this study.

3.2 Research Scope

This study aims to look for the relationship between statements on social media with stock liquidity and volatility in several industrial sectors in Indonesia so that the scope of the study includes:

1. Information on Twitter statement related to the company to be investigated
2. History of stock transaction activities in several industrial sectors to be investigated which includes basic information regarding daily stock price closing, daily stock transaction volume, and market value

3.3 Research Model

This study uses a model that combines the studies of Hu *et al.* (2019) and Guijarro *et al.* (2019) who examine the effect of sentiment analysis on stock liquidity. Liquidity calculation uses Amihud Illiquidity in which the ILLIQ variable becomes the dependent variable which is influenced by several independent variables Twitter sentiment analysis (TSA), while the control variables are including Returni, t-1 which is a calculation of return between the current period and the previous period, ln(MVi,t-1) which is a calculation of market capitalization between the current period and the previous period, TVi,t-1 which is a calculation of transaction volume between the current period and the previous period, ln(MVi,t-1)*ln(TSAi,t/TSAi,t-1) which is a variable to control the effect of sentiment analysis on market capitalization, and t which is a time variable. The research model between dependent and independent variables can be described as follows.

$$ILLIQ_{i,t} = \alpha + \beta_1 \ln(TSA_{i,t}/TSA_{i,t-1}) + \beta_2 Return_{i,t-1} + \beta_3 \ln MV_{i,t-1} + \beta_4 \ln(TV_{i,t-1}) + \beta_5 (\ln(MV_{i,t-1})*\ln(TSA_{i,t}/TSA_{i,t-1})) + \beta_6 (t-1) + \varepsilon_{i,t} \quad (1)$$

While the research model to determine the effect of sentiment analysis on stock volatility can use the standard deviation of the value of stock returns. In this model, the standard deviation becomes the dependent variable that is influenced by the independent variable sentiment analysis (TSA), TVi, t which is the transaction volume, Returni, t which is the value of stock returns, and t is the time variable.

$$Volatility_{i,t} = \alpha + \gamma_1 \ln(TSA_{i,t}/TSA_{i,t-1}) + \gamma_2 \ln(TV_{i,t}) + \gamma_3 \ln(Return_{i,t}) + \gamma_4 t + \varepsilon_{i,t} \quad (2)$$

3.4 Research Sample

In this study, the sample data refers to the type of non-probability sampling of stock movement data of some industries in which its companies are listed in the Top 10 IDX30 Companies with The Highest Market Capitalization. This list will be released by Indonesia Stock Exchange Handbook in December 2019. The highlighted industries are Finance, Infrastructure, Utilities & Transportation, Miscellaneous Industry, Consumer Goods Industry, and Basic Industry & Chemicals.

The definition and criteria regarding the grouping of companies refer to the classification and sub-classification of indices determined by the Indonesia Stock Exchange. The main category consists of Headline Index, Sector Index, Thematic Index, and Factor Index. Based on the stock index grouping, we focus the research on the shares of companies incorporated in IDX30 which are part of the Headline Index category and Liquidity sub-category because the types of companies in them include a number of companies with significant transaction trading volumes that meet the target criteria companies with large levels of liquidity.

3.5 Stock Data Collection Techniques

In order to gather stock information on each company sample, we utilize the Thomson Reuters application by entering the company's stock code. Then we choose Stock Price History and filter stock price data in the period from September to December 2019. This filtration will produce some information including the transaction date, the closing price of shares, and the volume of shares per day in that period.

3.6 Twitter Data Collection and Processing Techniques

Twitter social media data collection is downloaded from the official Twitter account through the API (Application Program Interface) which is indeed provided by Twitter for the benefit of the public who will conduct research. When authentication is successful, the next process is defining several parameters such as the company's stock code, the range of tweet date periods, and the type of language to filter Twitter data based on research needs. All of these filters become parameters to call the tweepy.cursor function, the results of which are extracted and stored in an output file in CSV (Comma-separated values) format. In the data extraction process, the program will also perform a data cleansing to remove the retweet, hyperlink, hashtag, mention, and non-ASCII characters. Based on the training data obtained, the next step is tokenization which aims to break sentences into words and standardize the language.

The next step is processing the contents of the tweet using TfidfVectorizer which aims to obtain the value of weighting the word on a tweet. Term Frequency (TF) calculates the frequency of occurrence of a word, while Inverse Document Frequency (IDF) calculates the word distribution in tweet sentences. The results are used to train existing models using the binary cross-entropy loss error function to calculate the average difference between the actual probability distribution and the predicted results, and the Stochastic Gradient Descent (SGD) optimizer to obtain an output weighting. The model that has been tested with the training data is then used to call the predict_proba function in the library Keras to provide a probability value from some input data.

3.7 Data Analysis Technique

The data analysis process aims to see the correlation between the results of Twitter sentiment analysis that has been obtained on the value of liquidity and stock volatility. We apply the Amihud Illiquidity calculation method based on the return value and stock volume to obtain the value of stock liquidity. Amihud (2002) states that stock liquidity can be seen through the calculation of the ratio between the values of the return to the volume of shares on the same day. Calculation of the Amihud Illiquidity method can be described in the following formula.

$$ILLIQ = \frac{1}{N}\sum_{t=1}^{T}\frac{|rt|}{\$Vt} \quad (3)$$

While the value of stock volatility is obtained by calculating the standard deviation of stock returns.

$$Volatility = \sqrt{\sum}(Pavg - Pt)^2/n \quad (4)$$

Referring to Gujarati *et al.* (2012), to estimate the regression model with panel data of this study using the Fixed Effect Model. This method is suitable for panel data types where the amount of time series data is greater than the number of cross-sections. For the purpose of data analysis, we apply descriptive statistical methods, hypothesis testing, and goodness of fit. Descriptive statistical analysis is the initial stage to study the characteristics of the data by looking at the distribution of data. Hypothesis testing includes t-tests to test how significantly each independent variable influences the dependent variable, and F-test that tests how significantly the overall independent variable influences the dependent variable.

4. RESULTS

4.1 Sentiment Analysis

Twitter sentiment analysis results are decimal values in the range of 0 to 1. Therefore, we need to categorize it into two groups. If the analysis sentiment score is >= 0.5, then we put label 1 to indicate positive sentiment. Conversely, if the analysis sentiment score is < 0.5, then we put label 0 to indicate negative sentiment.

Table 1. Sample of Positive Sentiment Analysis

Tweet	Sentiment Analysis
asing terus melakukan net buy pada saham bbca bbri dan bmri	0.708699942
sudah beberapa hari ini asing terus memborong bank papan atas sehingga membuat bbca dan bbri terus menguat.	0.760607481
sektor kedua favorit saya: bank. ekonomi modern berkembang sejak adanya 'kredit' dari bank. terbayang laba yang didapat kalau sudah mengumpulkan bbca atau bbri dari dulu.	0.728387952

Table 2. Sample of Negative Sentiment Analysis

Tweet	Sentiment Analysis
saya sudah bilang beli hmsp salah	0.029455524
kenaikan harga jual — penurunan harga saham hmsp	0.357681394
hmsp jadi bea cukai rokok akan naik tahun 2020 tapi yang di jual perusahaan rokok berbeda dengan produk lainnya. yang di jual produk adiktif. jumlah perokok muda juga tambah setiap waktunya. mahal per bungkus? yah paling beli satuan eceran.	0.226297751

4.2 Descriptive Analysis Result

Descriptive statistical test results in each industry sector describe the average, mean, maximum, minimum value, and standard deviation of ILLIQ variables that represent the results of Amihud Illiquidity calculation, TSA for the results of sentiment analysis, Market Value containing market capitalization, return which is the calculation of the difference between the current stock price and the previous day's stock price, the square of the return used to measure stock volatility, volatility as a variable of stock volatility, and the trading value for total stock transactions in units of currency.

Referring to the descriptive analysis result as shown below, we take a sample on Basic Industry & Chemicals as a sample of interpretation where the average value of the ILLIQ variable is 0.0000043 and the standard deviation is 0.0000043. The smaller the value of ILLIQ means that the stock is more liquid which also indicates that the frequency of trading of these shares rotates quickly accompanied by low transaction costs. The TSA variable as a result of Twitter sentiment analysis has an average of 0.077514 and a standard deviation of 0.181503. While the Volatility variable as a volatility reference has an average of 2.916434 and a standard deviation of 0.889574.

Table 3. Descriptive Analysis Result of Stock on Each Industry

Industry	Variable	Mean	Median	Maximum	Minimum	Std. Dev.
Basic Industry and Chemicals	ILLIQ	0.000043	0.000033	0.000322	0.000000	0.000043
	TSA	0.065630	0.000000	1.000000	0.000000	0.248103
	MARKET VALUE	78.97503	74.61288	134.4141	32.68910	26.24784
	RETURN	0.303552	0.000000	11.55378	-9.863946	2.828927
	RETURN_KUADRAT	0.095023	0.017569	0.966346	0.000003	0.152199
	VOLATILITY	2.916434	3.057195	5.373744	0.451672	0.889574
	TRADING VALUE	58.44654	49.51399	217.3893	8.231400	35.86476
Consumer Goods Industry	ILLIQ	0.000021	0.000016	0.000128	0.000001	0.000020
	TSA	0.083333	0.000000	1.000000	0.000000	0.276746
Infrastructure, Utilities, and Transportation	MARKET VALUE	158.7422	116.9450	354.7948	64.75563	96.39604
	RETURN	-0.053318	0.000000	6.908378	-5.654049	1.645015
	RETURN_KUADRAT	0.088896	0.017678	0.982145	0.000001	0.142874
	VOLATILITY	1.681991	1.578382	2.855330	0.211077	0.574538
	TRADING VALUE	67.34494	57.18434	334.9252	7.299684	44.88652
	ILLIQ	0.000035	0.000013	0.000259	0.000001	0.000052
	TSA	0.093750	0.000000	1.000000	0.000000	0.292243
	MARKET VALUE	0.164900	0.050907	0.430920	0.033314	0.170606
	RETURN	-0.046488	0.000000	8.415842	-13.52459	2.413493
	RETURN_KUADRAT	0.158896	0.116458	0.771210	0.002239	0.160965
	VOLATILITY	2.150925	2.065341	4.686278	0.152481	1.068623
	TRADING VALUE	158.1675	63.57723	1,084.670	8.509815	182.9928
Finance	ILLIQ	0.000010	0.000037	0.000124	0.000001	0.000018
	TSA	0.175000	0.000000	1.000000	0.000000	0.380562
	MARKET VALUE	0.351779	0.323977	0.824396	0.018819	0.266820
	RETURN	0.138729	0.000000	5.882353	-5.434783	1.615984
	RETURN_KUADRAT	0.118829	0.036241	0.951770	0.000003	0.167415
	VOLATILITY	1.675526	1.864126	3.269349	0.296414	0.637868
	TRADING VALUE	267.4367	246.9354	1,228.086000	11.520860	196.135200
Miscellaneous Industry	ILLIQ	0.000138	0.000012	0.001312	0.000002	0.000231
	TSA	0.234375	0.000000	1.000000	0.000000	0.425272
	MARKET VALUE	0.137625	0.130269	0.282373	0.004990	0.132673
	RETURN	-0.107112	0.000000	4.580153	-4.137931	1.593952
	RETURN_KUADRAT	0.206170	0.182751	0.620653	0.015576	0.130372
	VOLATILITY	1.518000	1.583810	2.079149	0.543865	0.372929
	TRADING VALUE	93.624360	79.365340	502.706000	1.445677	101.361600

4.3 Hypothesis Testing

The purpose of hypothesis testing is to test whether the obtained regression coefficients are significant or not. There are two types of hypothesis tests used in this study that are F-test and t-test. Below are the results of hypothesis testing based on regression analysis for each research model of Amihud Illiquidity (ILLIQ) and Volatility. The interpretation of both research models will take samples on Basic Industry & Chemicals.

Based on the regression analysis on ILLIQ model on Basic Industry & Chemicals, the F-test result shows p-value of 0.000000 and F-statistic of 8.937807 which means that all independent variables jointly influence the dependent variable and the model is valid. The result t-test on constant shows p-value of 0.000000 and t-statistic of 7.963521. TSA variable shows p-value of 0.1216 and t-statistic of -1.552552. From these results, we conclude that the independent variables individually affect the ILLIQ variable at α = 10%. Market capitalization as a control variable has a significant effect based on the probability value (0.0131) at α = 5%. The adjusted R-squared value of 0.121075 can be interpreted that the independent variable may impact the changes of ILLIQ variable by 12.1075%.

Table 4. Descriptive Analysis Result of Stock on Each Industry

Industry	Variable	Coefficient	Std. Error	t-Statistic	Prob.
Basic Industry and Chemicals	C	0.000017	0.000014	7.963521	0.000000
	TSA	-0.000010	0.000007	-1.552552	0.121600
	RETURN	0.000001	0.000001	0.886374	0.376100
	MV	0.000000	0.000000	-1.886374	0.063000
	TV	0.000000	0.000000	-1.587079	0.113500
	MV*TSA	0.000000	0.000000	-0.123670	0.901700
	T	0.000000	0.000000	-1.431578	0.153300
	Adjusted R-squared	0.217574			
	F-statistic	8.937807			
	Prob(F-statistic)	0.000000			
Consumer Goods Industry	C	0.000048	0.000018	2.636574	0.008700
	TSA	-0.000001	0.000003	-0.455096	0.649300
	RETURN	-0.000001	0.000000	-1.116108	0.265100
	MV	0.000000	0.000000	-0.958254	0.338600
	TV	0.000000	0.000000	-2.176279	0.030200
	MV*TSA	0.000000	0.000000	-0.723350	0.469900
	T	0.000000	0.000000	-2.848626	0.004600
	Adjusted R-squared	0.121075			
	F-statistic	5.327757			
	Prob(F-statistic)	0.000000			

Table 4. Descriptive Analysis Result of Stock on Each Industry (continued)

Infrastructure, Utilities, and Transportation	C	0.000058	0.000011	5.467539	0.000000
	TSA	0.000000	0.000001	0.321542	0.748200
	RETURN	0.000000	0.000000	0.558871	0.576900
	MV	-0.000014	0.000055	-1.884546	0.061100
	TV	0.000000	0.000003	-0.042688	0.966000
	MV*TSA	0.000000	0.000000	-0.343382	0.731700
	T	0.000058	0.000011	5.467539	0.000000
	Adjusted R-squared	0.446270			
	F-statistic	19.939440			
	Prob(F-statistic)	0.000000			
Finance	C	0.000019	0.000006	3.157403	0.001800
	TSA	-0.000001	0.000000	-1.287412	0.198900
	RETURN	0.000000	0.000000	0.803115	0.422500
	MV	-0.000013	0.000018	-0.698939	0.485100
	TV	0.000000	0.000000	-1.243719	0.214600
	MV*TSA	0.000000	0.000001	-0.740215	0.459700
	T	0.000000	0.000000	0.238282	0.811800
	Adjusted R-squared	0.335001			
	F-statistic	15.380130			
	Prob(F-statistic)	0.000000			
Miscellaneous Industry	C	0.0000216	0.000053	4.059508	0.000100
	TSA	-0.000002	0.000007	-0.321489	0.748400
	RETURN	0.000001	0.000003	0.299613	0.765000
	MV	-0.000564	0.0000466	-1.208881	0.229100
	TV	0.000000	0.000000	-0.711275	0.478300
	MV*TSA	-0.000004	0.000020	-0.192716	0.847500
	T	0.000001	0.000000	4.007618	0.000100
	Adjusted R-squared	0.470433			
	F-statistic	12.991900			
	Prob(F-statistic)	0.000000			

According to a regression analysis on the Volatility model of Basic Industry & Chemicals, the F test result shows p-value of 0.000000 and F-statistic of 26.69822 which means that all independent variables jointly influence the dependent variable and the model is valid. The t-test result shows p-value of 0.000000 and t-statistic of 19.89860. TSA variable shows p-value of 0.0035 and t statistic of -2.946843. From these results, we conclude that the independent variables individually affect the Volatility variable at $\alpha = 5\%$. The adjusted R-squared value of 0.391901 can be interpreted that the independent variable may impact the changes of Volatility variable by 39.1901%.

Table 5. Regression Sentiment Analysis of Volatility on Each Industry

Industry	Variable	Coefficient	Std. Error	t-Statistic	Prob.
Basic Industry and Chemicals	C	2.613067	0.131319	19.898600	0.000000
	TSA	-0.209283	0.071019	-2.946843	0.003500
	TV	0.001743	0.000532	3.278945	0.001200
	RETURN	0.005342	0.007856	0.680015	0.497000
	T	0.006572	0.003482	1.887152	0.060100
	Adjusted R-squared	0.391901			
	F-statistic	26.698220			
	Prob(F-statistic)	0.000000			
Consumer Goods Industry	C	1.397313	0.048777	28.647240	0.000000
	TSA	0.067213	0.026538	2.532650	0.011700
	TV	0.000964	0.000223	4.317227	0.000000
	RETURN	0.005214	0.005340	0.976413	0.329500
	T	0.006599	0.001302	5.068988	0.000000
	Adjusted R-squared	0.794111			
	F-statistic	165.135700			
	Prob(F-statistic)	0.000000			
Infrastructure, Utilities, and Transportation	C	1.649877	0.046523	35.463900	0.000000
	TSA	0.042393	0.046598	0.909773	0.364100
	TV	0.000198	0.000091	2.169798	0.031300
	RETURN	0.012966	0.011649	1.113062	0.267100
	T	0.014348	0.001277	11.233130	0.000000
	Adjusted R-squared	0.848495			
	F-statistic	179.280800			
	Prob(F-statistic)	0.000000			
Finance	C	1.511644	0.035813	42.209900	0.000000
	TSA	0.008635	0.028891	0.298895	0.765200
	TV	0.000038	0.000069	0.551390	0.581800
	RETURN	0.020898	0.008562	2.440802	0.015200
	T	0.004593	0.000946	4.853372	0.000000
	Adjusted R-squared	0.792397			
	F-statistic	148.381700			
	Prob(F-statistic)	0.000000			
Miscellaneous Industry	C	1.150407	0.074328	15.477480	0.000000
	TSA	-0.001317	0.053910	-0.024426	0.980600
	TV	0.000745	0.000331	2.248456	0.026300
	RETURN	0.025408	0.015996	1.588467	0.114800
	T	0.009258	0.001881	4.921725	0.000000
	Adjusted R-squared	0.310164			
	F-statistic	12.420340			
	Prob(F-statistic)	0.000000			

5. DISCUSSION

Analysis and testing that have been done show that Twitter sentiment analysis only affects the liquidity of company shares incorporated in the IDX30 stock index in certain industrial sectors, namely Basic Industry & Chemicals, and Miscellaneous Industry. However, the company's shares in the industrial sector Consumer Goods Industry, Infrastructure, Utilities & Transportation, and Finance, Twitter sentiment analysis has no significant effect on the company's stock liquidity on the same stock index. Therefore overall it can be concluded that Twitter sentiment analysis does not significantly influence the liquidity of company shares incorporated in the IDX30.

Meanwhile, analysis and testing conducted on volatility show that Twitter sentiment analysis only affects the volatility of company shares incorporated in the IDX30 stock index in certain industrial sectors, namely Basic Industry & Chemicals, and Consumer Goods Industry. However, in the company's shares in the Infrastructure, Utilities & Transportation, Finance, and Miscellaneous Industry sectors, Twitter sentiment analysis has no significant effect on the volatility of company shares in the same stock index. Roundly, it can be concluded that the Twitter sentiment analysis also does not significantly influence the volatility of company shares incorporated in IDX30.

6. CONCLUSION

Based on the discussion above, we conclude that the result of the descriptive analysis and regression test show that the sentiment analysis of Twitter has insignificant correlation on the liquidity and volatility of IDX30 index stock. This conclusion comes from the insufficient number of tweets which contains information about stock liquidity in Indonesia. Even though the data processing already uses a machine learning model to define the sentiment score at word level, extracting information from tweets in Bahasa Indonesia may give biased results. In order to cope with this limitation, we suggest to any similar study in the future to use a simple method such as bid-ask spread formula instead of ILLIQ method to measure the impact of tweets to the stock liquidity. Furthermore, we also hope that further research will include a process to filter out tweets which are indicated as buzzers. In addition, regulators are expected to work together to create regulations that can anticipate information manipulation that results in a decrease in stock transactions due to biased information.

REFERENCES

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets, 5*(1), 31-56.

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting, 36*(2), 334-357.

Chordia, T., Roll, R., & Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of financial economics, 56*(1), 3-28.

Domowitz, I., Glen, J., & Madhavan, A. (2001). Liquidity, volatility and equity trading costs across countries and over time. *International Finance, 4*(2), 221-255.

Guijarro, F., Moya-Clemente, I., & Saleemi, J. (2019). Liquidity Risk and Investors' Mood: Linking the Financial Market Liquidity to Sentiment Analysis through Twitter in the S&P500 Index. *Sustainability, 11*(24), 7048.

Gujarati, D., Ekonometrika, D.-d., Terjemahan Mangunsong, R., & Empat, S. (2012). buku 2, Edisi 5. In: Jakarta.

Hu, S., Zhong, M., & Cai, Y. (2019). Impact of Investor Behavior and Stock Market Liquidity: Evidence from China. *Entropy, 21*(11), 1111.

Lipson, M. L., & Mortal, S. (2009). Liquidity and capital structure. *Journal of financial markets, 12*(4), 611-644.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.

Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). *Correlating S&P 500 stocks with Twitter data.* Paper presented at the Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research.

Mufidah, N. (2018). Pengaruh Media Sosial Berbasis Trading Terhadap Pergerakan Harga Saham: Pendekatan Content Analysis.

Ortega, J. L. (2017). The Presence Of Academic Journals On Twitter And Its Relationship With Dissemination (Tweets) And Research Impact (Citations). *Aslib Journal Of Information Management.*

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects Of Twitter Sentiment On Stock Price Returns. *PloS one, 10*(9), e0138441.

Rothman, T. (2019). Trading the Dream: Does Social Media Affect Investors' Activity-The Story of Twitter, Telegram and Reddit. *International Journal of Financial Research, 10*(2).

Sarr, A., & Lybek, T. (2002). *Measuring Liquidity In Financial Markets* (Vol. 2): International Monetary Fund.

Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences, 48*(3), 454-488.