Advances in Economics, Business and Management Research, volume 152
Proceedings of the 5th Padang International Conference On Economics Education, Economics, Business and Management, Accounting and Entrepreneurship (PICEEBA-5 2020)

Twitter Sentiment on Mispricing in Indonesia Stock Market
(Long / Short Strategies Following Sentiment)

Lusiana Indra*, Zaafri Ananto Husodo

Universitas Indonesia, Depok and Indonesia
Corresponding author. Email: lusianaindra90@gmail.com

ABSTRACT
This paper examines the relationship between twitter sentiment on mispricing in Indonesia listed firms over period 2013 – 2017. This study uses machine learning method to classify sentiments based on Naïve Bayes, Support Vector Machine and Decision Tree algorithm. The results show that Decision Tree is the best method to classify sentiment in Indonesian. To measure mispricing, we use mispricing score method from Stambaugh Yu and Yuan 2012 associated with 9 long/short anomalies. The results of this experimental studies show that sentiment exhibits significant relation to stock returns on the long/short strategies. The short-leg strategy is more profitable following low or positive sentiment.

Keywords: *twitter sentiment, anomaly, long-short strategies, mispricing, machine learning*

1. INTRODUCTION

Twitter is a micro-blogging platform which allows users to send real-time messages as many as 140 characters and has widely used for financial forecast by (Graham et al. 2014; Java et al. 2007; Nguyen, Varghese, and Barker 2013). Sentiment analysis refers to a broad field of natural language processing, linguistic computing and text mining. (Zhang et al. 2011) consider that language can be an ambiguity in the use of words, absence of intonation, and the development of the language itself. (Bhuvaneswari, Parimala, and Periyar 2017; Ibrahim and Salim 2016) has applied Naïve Bayes (NB) and Support Vector Machine (SVM) classification method to find the best classification results. (Laeeq 2017) used K-NN, NB, and Decision Tree (DT) and found 77.50%, 80%, and 78%, respectively. Classification methods used in this study are NB, SVM and DT using RapidMiner version 9.

The potential of investor's sentiment to influence asset pricing that has proven by many researchers before and found that rational investor are disturbed by their cognitive biases and irrational behaviour that are systematically correlated and result in asset pricing error that cannot be eliminated by rational arbitrage (Baker et al. 2007; Baker and Wurgler 2006; Barberis, Shleifer, and Wurgler 2005; Kumar and Charles 2006; Long et al. 1990) and lead to investing decisions based on noise as if their basis information will let them take abnormal return consistently as described on (Peress and Schmidt 2017). However (Hu et al. 2014) argue that the traditional finance theory confirms that asset pricing is not affected by investor sentiment because rational investor is ready to arbitrage independently and rationally to compensate those mispricing, but (Shleifer and Vishny 1997) consider mispricing might continue because arbitrage is risky and expensive which finally will limit arbitration request to return asset to its fair value.

Based on those analysis, the idea behind this work is to classify sentiment on twitter data in Indonesian and its effect on mispricing. The hypothesis of this study is that stock returns are influenced by sentiment conditions. Besides that, twitter sentiment in Indonesian still very limited, some previous research did compare classification method in Indonesian, but not yet found linkage to financial data. This study uses 11 anomalies scoring system from (Stambaugh et al. 2012; 2015). Each anomaly will be examined based on the long/short strategies. This method has widely used to measure mispricing in the US market (Green and Zhang 2014; Griffin and Kelly 2010; Hou, Xue, and Zhang 2014), and the first measure mispricing score outside US stock market was (Jacobs 2016).

2. METHODS

The methods related to this work consist of two aspects: sentiment classification and mispricing scoring. Twitter data used are Indonesian tweet period 2013 – 2017 which contain keywords “ekonomi”, “harga pokok”, “suku bunga”, “inflasi”, “nilai tukar rupiah”, “produk domestik bruto”, “devisa”, “dana asing”, “harga saham”, “net sell” and “net buy” with total 166.895 tweets obtained from twitter crawling using python programming language. To compute mispricing score, we use 9 of 11 cross-sectional individual anomalies developed by (Stambaugh et al. 2012; 2015). These anomalies are o-score bankruptcy probability, net stock issues, accruals, net operating assets, price momentum, gross profitability, asset growth, return on assets, and investment to assets. We use only 9 out of 11 anomalies because the other 2 anomalies (financial distress and composite equity issues) have almost the same characteristics with o-score bankruptcy probability and net stock issues. Based on (Jacobs 2016), it takes at least five individual anomalies to build composite index. Observation data taken from data stream for listed company in Indonesia period 2013 – 2017 considering data availability.

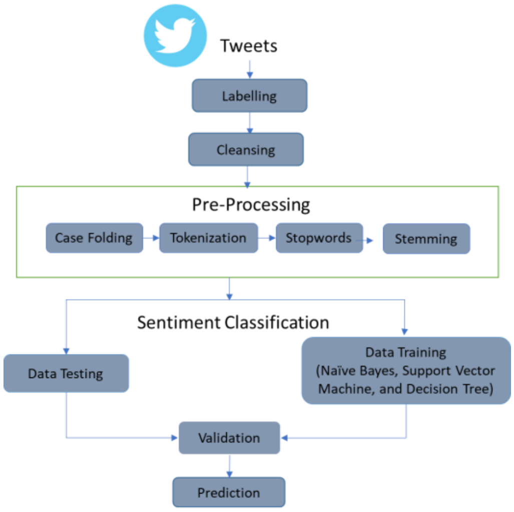

Figure 1. Sentiment Analysis Stages

In analysing the sentiment, there are some procedure needed to be done to get the best test results:

1. Labelling: process to give label for training data by person who are competent in this field.
2. Pre-processing
    a. Cleansing: process to eliminate irrelevant tweet such as redundant or meaningless tweet.
    b. Case folding: process to transform text into the same variety of writing.
    c. Tokenization: process to split sentences into certain parts and used to reduce words that are too short or too long which has no meaning for sentiment processing.
    d. Stopwords: process to eliminate words that are often used but do not have significance effect on sentiment analysis.
    e. Stemming: process to transform affix into its root.

3. Classification:
    a. Naïve Bayes (NB)
    Conditional probability theory that describes the likelihood of an event based on prior knowledge about conditions that may be associated with the event. NB is a powerful algorithm. It only needs to scan one single data to build the model.

    b. Support Vector Machine (SVM)
    This method was first developed to solve limited case of separating training data without error (optimal hyperplane). This approach was later developed by (Cortes and Vapnik 1995) with the idea of soft margins that can minimize the errors arise when it is not possible to separate training data vectors without error.
    c. Decision Tree (DT)
    DT is a greedy algorithm, using top-down recursive method to build a decision tree structure for classification. DT trees are based on the attribute of the data and the attribute itself is used as a basis for dividing the different classes.
4. Validation: After cleansing and pre-processing, the data are ready to be validated with the algorithm of each model to determine classification performance of each method.

3. RESULTS AND DISCUSSION

3.1 Sentiment Classification

The result of the classification process is in the form of a percentage of success and failure in predicting positive and negative sentiments. A measure of the success of a model is based on the amount of accuracy, precision and recall that occur in the model. The best method to classify Indonesian twitter sentiment in this study is DT classifier with 84% of accuracy.

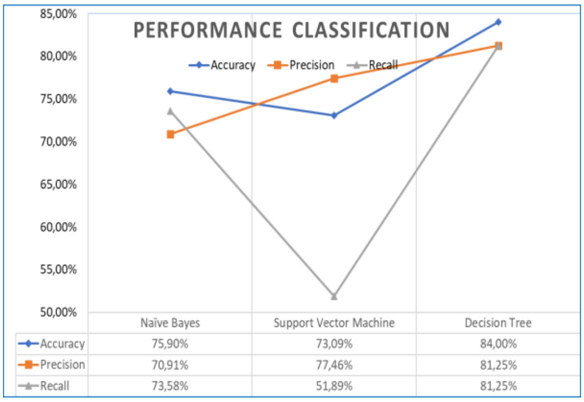

Figure 2 performance classification

3.2 Long/Short Strategies

The long/short strategies are an investment strategy that takes long position on stocks that are considered to be increasing (optimistic) and takes short position on stocks that are considered to be declining (pessimistic). For each anomaly obtained by value weighted portfolio returns. We then construct long and short strategy using the extreme decile, 1 and 10, with decile 1 as the long leg strategy with higher performing decile and decile 10 as the short leg strategy with lower performing decile.

Table 1 reports correlations of weekly returns on long/short strategies in the sample period. The highest correlation on the long leg strategy is total accrual to return on assets anomalies of 83 basis points (bps). High correlation that also appear on the short leg strategy are net operating assets to asset growth and combination of 97 bps and 90 bps, respectively. The strategies with high correlation aren't omitted because they don't appear on both sides. The lowest correlation is ohlson's and asset growth of – 1 bps. Overall, the strategies are not highly correlated.

3.3 Excess Returns and Sentiment (High/Low)

To figure out the relationship between sentiment and returns, we classify sentiments into 2 categories that are high and low sentiment based on the number of sentiments in the previous week. It is classified into high if the total sentiment is above median value, and into low if lower. Subsequently, this study performed excess return regression of each anomaly with (Fama and French 1993) model. (Baker and Wurgler 2006) found that when the sentiment is high, stocks that are speculative and difficult to arbitrage become overvalued and stocks that are safe and easy to arbitrage become undervalued because there is limitation to do short selling so arbitration becomes riskier and more expensive. Based on that argument, the short leg strategy which is consisting of lower performing stocks should have higher spread returns following high sentiment because the arbitrators pull out themselves from arbitration, otherwise for the short leg strategy.

The results in table 2 report 5 out of 10 anomalies in the long leg and only 2 out of 10 anomalies in the short leg strategy are significant. In the long leg strategy, the average return following high sentiment earns 0.07 bps which is lower than low sentiment that earns 0.41 bps. In the short leg strategy, the average return following high sentiment earns 0.04 bps which is higher than low sentiment that earns -0.34 bps (more profitable). Both strategies are significant of the t-test at 0.05 and 0.01 significance level, respectively. The results support the hypothesis that stock returns are influenced by sentiment conditions. This results also support the previous research which state when the sentiment is high, stocks that are speculative become overvalued and safe stocks become undervalued.

Tabel 1 Return anomalies
The table reports return correlations among each anomaly in sample period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Correlations: long leg										
(1) Ohlson's	1.0000									
(2) Net stock issues	0.2392	1.0000								
(3) Total accruals	0.2328	0.1840	1.0000							
(4) Net operating assets	0.3218	0.1439	0.2876	1.0000						
(5) Momentum	0.1333	0.1083	0.1292	0.1817	1.0000					
(6) Gross profitability	0.0801	0.0814	0.0480	0.0742	0.0921	1.0000				
(7) Asset growth	0.3684	0.2105	0.4069	0.2530	0.2444	0.0301	1.0000			
(8) Return on assets	0.2843	0.1572	0.8268	0.3093	0.0853	-0.0595	0.4743	1.0000		
(9) Investment to assets	0.1825	0.1116	0.2266	0.1432	0.1091	0.0595	0.3036	0.2655	1.0000	
(10) Combination	0.4521	0.5597	0.5272	0.4548	0.4881	0.5682	0.4995	0.4708	0.4449	1.000
Panel B. Correlations: short leg										
(1) Ohlson's	1.0000									
(2) Net stock issues	0.1722	1.0000								
(3) Total accruals	0.3651	0.4069	1.0000							
(4) Net operating assets	0.0562	0.6823	0.1551	1.0000						
(5) Momentum	0.0620	0.0847	0.1602	0.0331	1.0000					
(6) Gross profitability	0.3304	0.3005	0.6150	0.0666	0.1361	1.0000				
(7) Asset growth	-0.0124	0.6724	0.1238	0.9735	0.0320	0.0363	1.0000			
(8) Return on assets	0.3237	0.2839	0.5324	0.1689	0.1191	0.6602	0.1529	1.0000		
(9) Investment to assets	0.2015	0.4509	0.4624	0.2141	0.1638	0.4481	0.1503	0.3869	1.0000	
(10) Combination	0.2523	0.7948	0.4267	0.9047	0.2853	0.3454	0.8855	0.4181	0.4425	1.000

Tabel 2 Excess returns anomali on the long/short strategies following high/low sentiment period

Anomaly	Long leg			Short leg		
	High Sentiment	Low Sentiment	High - Low	High Sentiment	Low Sentiment	High - Low
Ohlson's	-0.0010	-0.0004	-0.0006	-0.0013	0.0002	-0.0002
	(-0.3890)	(-0.1898)	(-0.1992)	(-0.8668)	(0.0927)	(-0.0927)
Net stock issues	-0.0008	0.0035	-0.0044	-0.0010	-0.0055	0.0055
	(-0.3367)	(1.7241) *	(-2.0607)	(-0.2369)	(-0.7161)	(0.7161)
Total accruals	-0.0008	-0.0008	0.0001	-0.0006	-0.0023	0.0023
	(-0.7383)	(-0.8406)	(0.1023)	(-0.2271)	(-0.9096)	(0.9096)
Net operating assets	0.0036	0.0098	-0.0062	0.0016	-0.0026	0.0026
	(0.4163)	(1.6150)	(-1.1987)	(0.5893)	(-1.1205)	(1.1205)
Momentum	0.0039	0.0049	-0.0010	0.0013	-0.0011	0.0011
	(1.0273)	(1.0908)	(-0.0635)	(0.2169)	(-0.2442)	(0.2442)
Gross profitability	-0.0019	0.0003	-0.0022	0.0064	-0.0123	0.0123
	(-2.3581) **	(0.2774)	(-2.6355)	(1.1804)	(-1.1780)	(1.1780)
Asset growth	0.0021	0.0149	-0.0128	-0.0006	-0.0026	0.0026
	(0.2001)	(1.9962) **	(-1.7961)	(-0.5298)	(-2.0474) **	(2.0474)
Return on assets	0.0019	0.0041	-0.0022	-0.0024	-0.0034	0.0034
	(1.3821)	(2.6127) ***	(-1.2305)	(-1.0014)	(-1.6959) *	(1.6959)
Investment to assets	-0.0003	0.0009	-0.0012	0.0025	-0.0027	0.0027
	(-0.1760)	(0.5070)	(-0.6830)	(0.5093)	(-1.0569)	(1.0569)
Combination	0.0007	0.0042	-0.0034	-0.0014	-0.0018	0.0018
	(0.2990)	(2.2620) **	(-1.9630)	(-0.8237)	(-1.1823)	(1.1823)
Average	0.0007	0.0041	-0.0034	0.0004	-0.0034	0.0039
T-test	-2.0760**			3.0800***		

This table reports excess returns for each anomaly based on regression:

$$Ri,t = a_h d_{h,t-1} + a_l d_{l,t-1} + bMKT_t + cSMB_t + dHML_t + e_{i,t}$$

Where Ri,t is excess return strategy in week t. $d_{h,t-1}$ and $d_{l,t-1}$ are dummy variables indicate sentiment period $t-1$. $d_{h,t-1}$ represent high sentiment and $d_{l,t-1}$ represent low sentiment. MKT is the excess return on the stock market. SMB is a return spread between small and large firms. HML is a return spread between stocks with high and low book-to-market ratios. All t-statistics are based on the heteroskedasticity-consistent standard errors of White (1980). Statistical significance at the 10%, 5%, and 1% level indicated by *, **, ***, respectively.

3.4 Excess Returns and Sentiment (Positive/Negative)

To robust the relationship between sentiment and returns, we investigate positive and negative sentiment and regress on long/short strategies as reported in table 3. (Stambaugh et al. 2012) argue that an optimistic views has the greatest effect on stock prices because the views are not affected by other less optimistic views, so it should be easier for optimistic sentiment to raise price than pessimistic to reduce price. If optimistic sentiment indeed stronger drive the price than pessimistic then mispricing should be stronger following positive sentiment. Based on that argument, long/short strategies should have spread returns lower following positive sentiment because they are easier to arbitrage than negative sentiment.

The results in table 3 report 6 out of 10 anomalies are significant on the long leg strategy and only 3 out of 10 anomalies are significant on short leg strategy. In the long leg strategy, the average return following positive sentiment earns 0.09 bps which is lower than negative sentiment that earns 0.37 bps. In the short leg strategy, the average return following positive sentiment earns -0.34 bps which is lower and more profitable than negative sentiment that earns 0.01 bps. Both strategies significance of the t-test at 0.05 significance level. The results support the hypothesis that stock returns are influenced by sentiment conditions. This results also support the previous research that the optimistic sentiment (positive) stronger drive the price up than pessimistic sentiment (negative) to drive it down.

Tabel 3 Excess returns anomaly on the long/short strategies following positive/negative sentiment

Anomaly	Long leg			Short leg		
	Positive	Negative	Positive - Negative	Positive	Negative	Positive - Negative
Ohlson's	-0.0002	-0.0011	0.0009	0.0002	-0.0011	0.0013
	(-0.0855)	(-0.5290)	(0.4436)	(0.1109)	(-0.8535)	(0.9644)
Net stock issues	-0.0016	0.0035	-0.0051	-0.0076	0.0004	-0.0080
	(-0.5751)	(1.8887) *	(-2.4638)	(-0.9924)	(0.0812)	(-1.0736)
Total accruals	-0.0017	-0.0001	-0.0016	-0.0016	-0.0013	-0.0003
	(-1.6102)	(-0.1284)	(-1.4818)	(-0.4898)	(-0.5509)	(0.0611)
Net operating assets	0.0038	0.0089	-0.0052	0.0018	-0.0022	0.0040
	(0.3813)	(1.8918) *	(-1.5104)	(0.6051)	(-1.0284)	(1.6335)
Momentum	0.0016	0.0068	-0.0053	-0.0108	0.0089	-0.0197
	(0.3775)	(1.6274)	(-1.2499)	(-2.1114) **	(1.7322) *	(-3.8436)
Gross profitability	-0.0018	0.0000	-0.0019	-0.0096	0.0030	-0.0126
	(-2.1729) **	(0.0115)	(-2.1844)	(-1.5766)	(0.3131)	(-1.8897)
Asset growth	0.0058	0.0105	-0.0047	-0.0023	-0.0010	-0.0013
	(0.4729)	(1.9814) **	(-1.5084)	(-1.7317) *	(-0.8678)	(-0.8639)
Return on assets	0.0029	0.0032	-0.0003	-0.0023	-0.0033	0.0010
	(1.9809) **	(2.1503) **	(-0.1694)	(-0.9200)	(-1.6836) *	(0.7636)
Investment to assets	-0.0008	0.0012	-0.0020	-0.0002	-0.0001	-0.0002
	(-0.4887)	(0.6965)	(-1.1851)	(-0.0557)	(-0.0148)	(-0.0409)
Combination	0.0009	0.0037	-0.0028	-0.0012	-0.0019	0.0007
	(0.3149)	(2.3893) **	(-2.0744)	(-0.6241)	(-1.3215)	(0.6973)
Average	0.0009	0.0037	-0.0028	-0.0034	0.0001	-0.0035
T-test	-2.0446**			-2.1052**		

This table reports excess return for each anomaly based on regression:

$$Ri,t = a_p d_{p,t} + a_n d_{n,t} + bMKT_t + cSMB_t + dHML_t + e_{i,t}$$

Where Ri,t is *excess return* for each anomaly. $d_{p,t}$ dan $d_{n,t}$ are dummy variables indicate sentiment period t. $d_{p,t}$ represent positive sentiment and $d_{n,t}$ represent negative sentiment. All t-statistics are based on the heteroskedasticity-consistent standard errors of White (1980). Statistical significance at the 10%, 5%, and 1% level indicated by *, **, ***, respectively.

3.5 *Predictive Regressions*

Table 4 reports predictive regression to investigate whether tweet volume and level of sentiment predict returns consistently with the previous finding. Tweet volume is the number of tweets on a given week. The higher the number of tweets, the higher the sentiment. Mood sentiment is the level of sentiment. The value of mood sentiment can be between -1 and 1 where 1 represents a 100% positive mood in sentiment of tweets for a given week and -1 would determine the opposite. Mood sentiment is conducted from:

$$Mood\ sentiment_t = \frac{TPositive_t - TNegative_t}{total\ number\ of\ tweets_t}$$

Where $TPositive_t$ and $TNegative_t$ indicate positive and negative sentiment period t.

Regression results on high/low sentiment with excess return (the average return is lower following high sentiment on the long leg strategy and higher following high sentiment on the short leg strategy) predict negative (positive) relation between tweet volume and the long leg (short leg) strategy. The higher the sentiment, the lower (higher) the return generated on the long leg (short leg) strategy. Consistent with the prediction, in table 4 reports for the long leg strategy, all anomalies negatively correlated with tweet volume while 3 out of 10 anomalies are significant. In the short leg strategy, 7 out of 10 anomalies are positively correlated with tweet volume while none anomaly is significant. The short leg strategy is not significant statistically but economically, which earns -0.3 bps per week at 0.01 t-test significance level.

Regression results on positive/negative sentiment with excess return (the average return is lower following positive sentiment on both strategies) predict negative relation between sentiment level and long/short strategies. The more positive the sentiment then the lower the return generated on both strategies. In table 4 reports for the long leg strategy, all anomalies negatively correlated with mood sentiment while 4 out of 10 anomalies are significant. In the short leg strategy, 3 out of 10 anomalies are negatively correlated with mood sentiment while none anomaly is significant. The short leg strategy is not significant statistically but economically, which earns -0.3 bps per week at 0.01 t-test significance level.

Tabel 4 Tweet volume and mood sentiment: predictive regression for excess returns on the long/short strategies.

Anomaly	Tweet volume				Mood sentiment			
	Long leg		Short leg		Long leg		Short leg	
	β	t-stat	β	t-stat	β	t-stat	β	t-stat
Ohlson's	-0.0000	(-0.5020)	-0.0000	(-0.5476)	-0.0000	(-0.5099)	-0.0000	(-0.4710)
Net stock issues	-0.0000	(-1.5156)	-0.0000	(-0.0189)	-0.0000	(-1.8435) *	-0.0000	(-0.4392)
Total accruals	-0.0000	(-0.7637)	0.0000	(0.4343)	-0.0000	(-0.6363)	0.0000	(0.5444)
Net operating assets	-0.0000	(-2.1131) **	0.0000	(1.0484)	-0.0000	(-2.2445) **	0.0000	(1.1194)
Momentum	-0.0000	(-0.5674)	-0.0000	(-0.7063)	-0.0000	(-0.4802)	-0.0000	(-0.8938)
Gross profitability	-0.0000	(-1.5892)	0.0000	(0.3529)	-0.0000	(-1.5175)	0.0000	(0.4458)
Asset growth	-0.0000	(-2.1051) **	0.0000	(0.5965)	-0.0000	(-2.2555) **	0.0000	(0.5670)
Return on assets	-0.0000	(-0.8587)	0.0000	(0.8205)	-0.0000	(-0.3649)	0.0000	(0.7136)
Investment to assets	-0.0000	(-0.2586)	0.0000	(0.8631)	-0.0000	(-0.3261)	0.0000	(1.0311)
Combination	-0.0000	(-2.3859) **	0.0000	(0.4243)	-0.0000	(-2.5905) **	0.0000	(0.4430)

The table reports b estimation in the regression:

$$Ri,t = a + bVtweet_t + bMood_t + cMKT_t + dSMB_t + eHML_t + e_{i,t}$$

Where Ri,t is excess return for each anomaly. bVtweet is the number of tweets. bMood is the level of sentiment. All t-statistics are based on the heteroskedasticity-consistent standard errors of White (1980). Statistical significance at the 10%, 5%, and 1% level indicated by *, **, ***, respectively.

4. CONCLUSIONS

The purposes of this study are to find the best method to classify sentiment in Indonesian and to investigate the influence of twitter sentiment on mispricing in Indonesia. The results show that decision tree is the best method to classify sentiment in Indonesia. We find that sentiment through twitter influences the profits of Indonesian stock market that can be obtained by investors. Many investors try to identify mispricing stocks and expected to earn higher return (Arffa. 2001), as theoretically, arbitrage opportunities can be obtained by investor with do maths, observe price error and take opposite position with noise trader (Black 1985). We also find that investor can get higher profit by taking long position on overvalued stocks following low sentiment or when the sentiment moves positively.

According to the limitation of this study, the suggestions for further research are as follows:

3. Using other social media as an indicator of sentiment such as google search or news platform.
1. In addition to macroeconomic, there are some other hashtags such as government policies, infrastructure, foreign outlook to Indonesia, and political can be added.
2. Adding other research models to support the influence of sentiment in twitter on mispricing in Indonesia.

ACKNOWLEDGMENTS

I would thank God for being able to finish this research. Then I would thank to my husband, my parents in heaven, my whole family, my classmates and friends who helped me with their valuable suggestions and guidance to complete this research. Thank to my mentor Mr. Zaafri Ananto Husodo, whose valuable guidance has served as major contributor towards the completion of this research. Last but not least, I would thank to Miss. Jap Vicky and Mr. Helmi Satria for helping me to get the data.

REFERENCES

Arffa., Ed. by Robert C. 2001. 16 *Expert Financial Planning ; Investment Strategies from Industry Leaders* . John Wiley & Sons.

Baker, Malcolm, and Jeffrey Wurgler. (2006). American Finance Association Investor Sentiment and the Cross-Section of Stock Returns Author (MALCOLM BAKER and JEFFREY WURGLER): Malcolm Baker and Source : The Journal of Finance , 61(4), 1645-1680

Baker, Malcolm, Jeffrey Wurgler, Malcolm Baker, and Jeffrey Wurgler. (2007). "Investor Sentiment in the Stock Market t the Go-Go Years of the Late 1960s , the Nifty Fifty Bubble of the Early 1970s ,." 21(2): 129–51.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. 75 *Article in Press*.

Bhuvaneswari, K, R Parimala, and E V R Periyar. 2017. "Sentiment Reviews Classification Using Hybrid Feature Selection." 10(7): 1–12.

Black, Fischer. 1985. "Fischer Black President of the American Finance Association 1985." *The Journal of Finance* XLI(July 1986).

Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." 297: 273–97.

Fama, F, and R French. 1993. "Common Risk Factors in the Returns Stocks and Bonds *." 33: 3–56.

Graham, Mark et al. 2014. "Where in the World Are You ? Geolocation and Language Identification in Twitter Identification in Twitter ∗." (October): 37–41.

Green, Jeremiah, and X Frank Zhang. 2014. "The Remarkable Multidimensionality in the Cross-Section of Expected U . S . Stock Returns." : 1–55.

Griffin, John M, and Patrick J Kelly. 2010. "Inferences ? A Comparison of Developed and Emerging Markets."

Hou, Kewei, Chen Xue, and Lu Zhang. 2014. "Digesting Anomalies : An Investment." 28(614): 650–705.

Hu, Changsheng, Wei Sun, Yongfeng Wang, and Yangchun Chi. 2014. "Investor Sentiment and the Predictability of Asset Returns : Evidence from China." 6(6): 577–85.

Ibrahim, Mnahel Ahmed, and Naomie Salim. 2016. "Sentiment Analysis of Arabic Tweets : With Special Reference Restaurant Tweets." 4(3): 173–79.

Jacobs, Heiko. 2016. "Market Maturity and Mispricing R." *Journal of Financial Economics* 122(2): 270–87. http://dx.doi.org/10.1016/j.jfineco.2016.01.030.

Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. "Why We Twitter : Understanding Microblogging Usage and Communities Why We Twitter : Understanding Microblogging." (January 2014).

Kumar, Alok, and M.C. Lee Charles. 2006. "American Finance Association Retail Investor Sentiment and Return

Comovements Author (s): Alok Kumar and Charles M . C . Lee Source : The Journal of Finance , Vol . 61 , No . 5 (Oct ., 2006), Pp . 2451-2486 Published by : Wiley for the American Financ." 61(5): 2451–86.

Laeeq, Farhan. 2017. "Sentimental Classification of Social Media Using Data Mining." 8(5): 2015–18.

Long, J Bradford De et al. 1990. "Noise Trader Risk in Financial Markets Andrei Shleifer." 98(4): 703–38.

Nguyen, Vu Dung, Blesson Varghese, and Adam Barker. 2013. "The Royal Birth of 2013 : Analysing and Visualising Public Sentiment in the UK Using Twitter A." (ii).

Peress, Joel, and Daniel Schmidt. 2017. "Noise Traders Incarnate : Describing a Realistic Noise Trading Process."

Shleifer, Andrei, and Robert W Vishny. 1997. "American Finance Association The Limits of Arbitrage Author (s): Andrei Shleifer and Robert W . Vishny Source : The Journal of Finance , Vol . 52 , No . 1 (Mar ., 1997), Pp . 35-55 Published by : Wiley for the American Finance Association Stable URL :" 52(1): 35–55.

Stambaugh, Robert F et al. 2012. "The Short of It : Investor Sentiment and Anomalies $." 104: 288–302.

Stambaugh, Robert F, Jianfeng Yu, and Y U Yuan. 2015. "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle." LXX(5).

Zhang, Lei et al. 2011. "Sentiment Analysis Combining Lexicon-Based and Learning-Based Methods for Twitter Sentiment Analysis."