Twitter Buzzer Detection for Indonesian Presidential Election

Andi Suciati
Faculty of Computer Science
Universitas Indonesia
Depok, Indonesia
andi.suciati@ui.ac.id

Ari Wibisono
Faculty of Computer Science
Universitas Indonesia
Depok, Indonesia
ari.w@cs.ui.ac.id

Petrus Mursanto
Faculty of Computer Science
Universitas Indonesia
Depok, Indonesia
santo@cs.ui.ac.id

***Abstract*—The campaign that was done in social media has high correlation to the supporters who disseminating the information deliberately, which called as buzzer. However, data that were generated by buzzer accounts can be considered as noise and need to be removed. In this research we performed task for detecting the buzzer accounts in Twitter by observing the impact of features we used which we selected based on their Mutual Information scores. We examined the performance of four machine learning algorithms which are Ada Boost (AB), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Histogram-based Gradient Boosting (HGB). The algorithms were evaluated using 10 folds cross validation and the results show that the best accuracy and precision achieved by AB which are 62.3% and 61.3% respectively with 25 features while the recall attained by XGB (67.9%) which the score same with its recall result with 20 features.**

Keywords—social media, twitter, buzzer detection, mutual information

I. Introduction

Presidential Election 2019 was becoming hot topic for the citizens in Indonesia for these past few months. Even people started to talk about this topic since 2018 especially on the internet. Social media such as Twitter [2] and Facebook[3], have important role as they become tools for politics activities. This happened because by using social media, the information that were written can spread speedily. One example that can show social media role in politics was the Presidential Election in United States in 2008. According to the study from Kiyohara [6], almost all candidates had own Facebook profile pages and Myspace during election campaign period. Besides, new communication technology had important role for Obama victory by gaining young voters through it. Another example was Japan election. Since amendment of Japan's Public Officers Election Act approved the utilization of internet for election campaigning in 2013, the political parties use social media such as Twitter and Facebook to campaign and advertise the candidates [5]. In addition, by using social media, the information that retrieved from them, which cannot be earned from media or traditional polling results, can be used to predict the election result [1].

In Indonesia, both competitors of president election also used social media as platforms for selling the excellences for each candidate. In 2015, the study explored by Satria et. al [9] shows that there was correlation between social media and presidential campaign 2014 in Indonesia by measuring correlation scores between the uses of social media of the voters and voting decision. Besides, there was a work conducted by Ibrahim et. al [2] related to Indonesian Presidential Election in 2014 and the data was retrieved form Twitter as well. The campaign that was done in social media has high correlation to the supporters who disseminating the information deliberately, which later we called as buzzer or influencer. Juzar and Akbar [4] mentioned that buzzer is social media user that has influence on other users and considered as central point of friendship in social media. However, data that were generated by buzzer accounts, in case Twitter, can be considered as noise while we want to conduct other study such as sentiment analysis and need to be removed. For that reason, detecting the buzzers can be said as important task. In this research we performed task for detecting the buzzer accounts in Twitter by observing the impact of features which were chosen by counting their Mutual Information scores. Then, those were examined by applying four machine learning algorithms which are AdaBoost (AB), Extreme Gradient Boosting (XGB), Gradient Boosting (GB), and Histogram-based Gradient Boosting (HGB). Then, we compared the precision, recall, and accuracy from all algorithms. By this work, we hoped it can give new buzzer detection framework reference to people while they want to conduct sentiment analysis using Twitter data.

The rest of this paper is constructed as follows: In section 2, we mention several related to our work. Then in section 3, we describe the research methods that applied in this work and the result. Furthermore, in last section, we present the conclusion from our work.

II. Related Work

[2]https://twitter.com
[3] https://www.facebook.com

978-1-7281-4610-2/19/$31.00 ©2019 IEEE

TABLE I. RELATED WORK

Author	Title	Objective	Data	Method	Result	Future Works
Juzar M. T., Akbar S. (2018) [4]	Buzzer Detection on Twitter Using Modified Eigenvector Centrality	Detecting the buzzer account by considering dynamics data from Twitter such as user interactions, activeness, influence level, interaction within users, level of activeness users, and influence level of users.	Data were retrieved from Twitter by using Twitter API	Modified eigenvector centrality. The modification was in adjacency matrix and removing the eigenvalue calculation.	By the evaluation of execution time, eigenvector centrality was slower than degree centrality. In evaluation from spread of information, eigenvector centrality got 6, 15, and 18 for threshold 0.75, 0.5, and 0.25 respectively.	Use sentiment value to make weight value in graph relation much better.
Ibrahim et. al. (2015) [2]	Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in A Twitter Nation	Predicting the results of Indonesian Presidential Election in 2014 by removing the buzzer accounts and analyzing the sentiment polarity from the data.	10 million data were collected by using Twitter Streaming API 10 during the campaign period (May 1st - July 6th, 2014)	Classifier based automatic buzzer detection and compute the sentiment polarity using Indonesian Sentiment Lexicon	The method was successfully attained 0.6% for mean absolute error (MAE) score.	
Kantepe M. , Ganiz M. C. (2017) [3]	Preprocessing framework for Twitter bot detection	Detecting social bot on Twitter by using large number features in preprocessing step.	The data collection contains 1800 users on Twitter which were divided to 1200 non-suspended accounts and 600 suspended accounts	Machine learning methods which are Logistic Regression, Multinomial Naive Bayes, SVM, and Gradient Boosted Trees.	Logistic Regression obtained 75% accuracy while Multinomial Naive Bayes, SVM, and Gradient Boosted Trees obtained 78%, 82%, and 86% accuracy respectively.	Improving models by adding social networks analysis features and detecting which bot accounts are managed by same software agents.
Ping H., Qin S. (2018) [8]	A Social Bots Detection Model Based on Deep Learning Algorithm	Detecting social bots by using Deep Learning.	The data that were used was from [10]	The deep learning model was DeBD which uses CNN-LSTM algorithm to detect social bots.	The DeBD achieved 0.999 F1 score in first test and 0.995 F1 score in second test	Using the user tweets and information to detect social bots.

III. RESEARCH METHODS

In this section, we will explain about the methods steps that we used in this research. The methods consist 4 steps, which are data collection, feature selection, experiment, and evaluation.

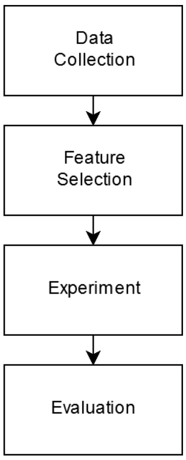

Fig. 1. Diagram of research methods.

A. Data Collection

We collected 2500 tweet about politic using Twitter search API within the election campaign period, which were since 29th September 2018 until 13th April 2019. For retrieving those tweets, we used four keywords which portrayed two candidates and one keyword for neutral tweets. The keywords selection was based on trending topic hashtags on Twitter which represent those parties. From many keywords that may occurred, we randomly selected five keywords. The keywords we selected can be seen at Table II.

After gathering 2500 tweets from three parties, we extract the features from the accounts that wrote those tweets. We found 1017 accounts, then, we labelled those accounts into 'buzzer' and 'non-buzzer' manually. The definition of buzzer that we applied was the accounts that tweeted about specific candidate continuously. The border between buzzer and non-buzzer is their daily activities. If there are accounts that talk about other the topic outside topic about the candidates, we labelled those account with non-buzzer. However, there were 24 accounts that have deactivated by the owners. So, we removed those accounts

TABLE II. KEYWORDS FOR COLLECTING DATA

Parties	Keywords
Joko Widodo	"saya01", "jokowi2periode"
Prabowo Subianto	"2019gantipresiden", "01maincurang"
Neutral	"pemiludamai"

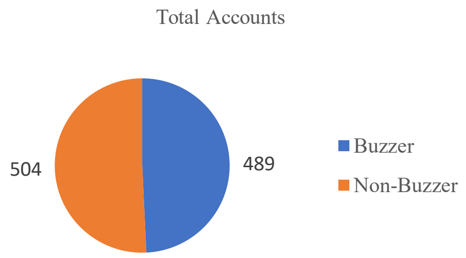

Fig. 2. Comparison of the total number of Buzzer and Non-Buzzer accounts.

from our dataset and that made our data consists 993 accounts that we used for detecting buzzer in Twitter.

B. Feature Selection

The features that we used were based on research by Kantepe and Ganiz [3] which is a study about bot detection in Twitter. We chose 11 features randomly as our initial features. We later extracted the final features for our study by using these initial features. The features which have nominal type were encoded into several new features. For instance, the 'month_created' feature was encoded into 12 months (e.g. month_1, month_2, month_3). For the features with numerical type, we scaled them using z-score. After encoding and scaling all the features, our total final features are 64 features. Furthermore, still following [3], we used Mutual Information (MI) as our feature selection method for measuring the mutual dependence between features and label.

Feature selection was method for avoiding curse of dimensionality which was done by selecting few features from original features. Those selected features become new feature subset [11]. In addition, the mutual information is a feature correlation discriminated methods [7], which is measured the correlation scores between features and label. There are other techniques for selecting features such as information gain, chi square statistic, document frequency, etc. [12]. However, for our scenario, we used MI and selected 5, 10, 15, 20, and 25 features with highest MI scores from total 64 features. The features we used after extraction as can be seen at Figure 3.

C. Experiments

In this study, we applied four boosting algorithms which are AdaBoost (AB), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Histogram-based Gradient Boosting (HGB). The reason we used boosting-based algorithm was because in the previous work by Kantepe and Ganiz [3], GB achieved best scores while using MI as one of methods to choose best features.

TABLE III. INITIAL FEATURES

No	Features Name	Types of Data	Description
1	*Most_active_days*	nominal	The day of a user tweet frequently.
2	*Most_active_hours*	nominal	The hours of a user tweet frequently.
3	*Freq_of_most_hashtag*	numerical	The frequency of hashtags in 500 newest tweets.
4	*Location*	nominal	Determine whether the users included location in their tweets or not.
5	*Verified*	nominal	Checking if the users are verified users.
6	*Followers_count*	numerical	Counting the number of followers from an account.
7	*Friends_count*	numerical	Counting the number of following from an account.
8	*Favorites_count*	numerical	Counting number of favorite tweets.
9	*Statuses_count*	numerical	Counting the number of tweets that have been posted.
10	*Month_created*	nominal	Retrieved the month when the account was created
11	*Year_created*	nominal	Retrieved the year when the account was created.

Besides, boosting-based algorithms combine multiple weak classifiers to build strong classifier. So, by comparing the boosting-based algorithms, we want to see which algorithm perform best than other.

D. Evaluation

For the evaluation, we used 10 folds cross validation to avoid overfitting. Then we compared the scores of accuracies, precision, and recall for each algorithm after average the results of every fold.

From the figures of the performances of all algorithms, we can see that best scores for every scenario was achieved by different model. The best scores while classifying with 5 features was obtained by XGB following by GB. However, AB classified better than XGB while using 10 features. For 15, 20, and 25 features, the best scores for precision and recall were led by AB when XGB got highest scores for recall. By observing the results, it seems that all algorithms accomplished best scores in fourth scenario, which was 20 features except for AB.

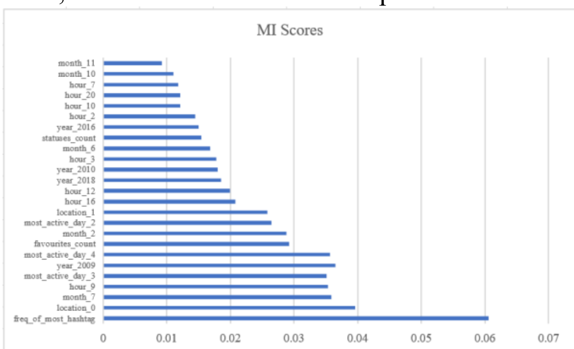

Fig. 3. 25 features with highest Mutual Information scores.

This is possibly due to the 21th until 25th features did not give many information that can increase the metric scores and can be categorized as unimportant features and do not make the features more variative However, the highest accuracy and precision achieved by AB which are 62.3% and 61.3% respectively with 25 features while the recall attained by XGB (67.9%) which the score same with its recall result with 20 features.

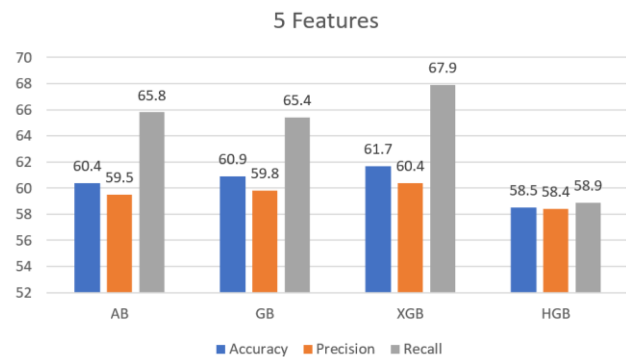

Fig. 4. The result of classifiers with 5 features.

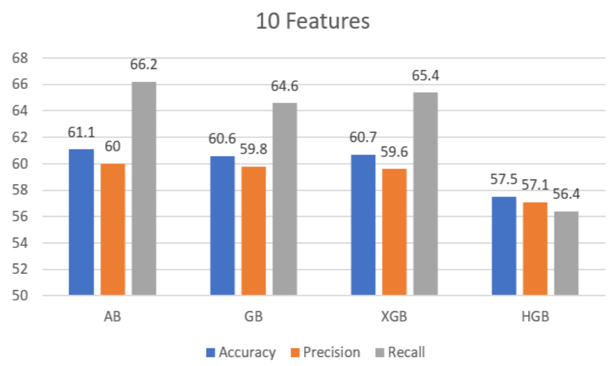

Fig. 5. The result of classifiers with 10 features.

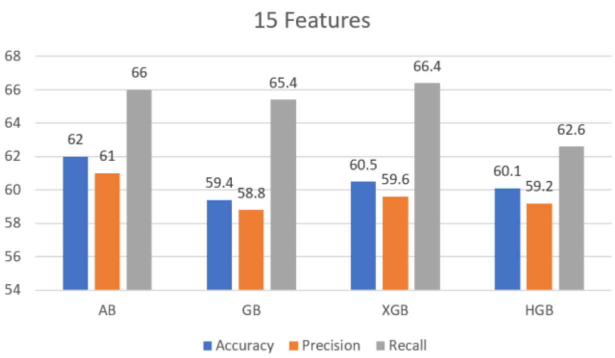

Fig. 6. The result of classifiers with 15 features.

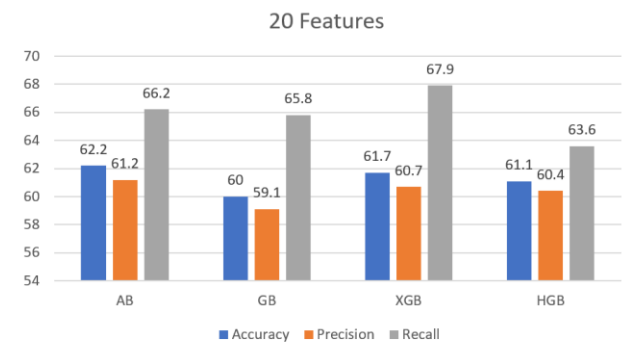

Fig. 7. The result of classifiers with 20 features.

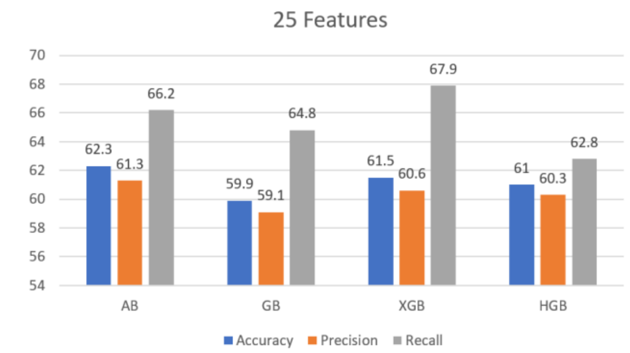

Fig. 8. The result of classifiers with 25 features.

Furthermore, the difference of the scores that obtained by from all algorithms were not too far. The highest difference only happened on Recall scores between XGB and HGB which about 3.8% to 9%. In addition, the best scores were only in range 60% - 68%. This may be highly caused by the data that we used were limited, so, the scores were not too great. Besides, the human error when labelling the data also could affect the scores because the boundaries between buzzer and non-buzzer are difficult to identify. Most of the campaign parties in social media not only talking about the superiority of candidates, but also various focuses such as about Islamic topic (refer to candidate number 01) or nationalism topic (refer to candidate number 02).

IV. Conclusion and Future Works

In this study, we have examined the performance of four machine learning algorithms which are AdaBoost (AB), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Histogram-based Gradient Boosting (HGB). We conducted buzzer detection by doing five scenarios by using, 5, 10, 15, 20 and 25 features which we selected based on their Mutual Information scores. The algorithms were evaluated using 10 folds cross validation, then we compared the accuracy, precision, and recall of each algorithm. The results show that the best accuracy and precision achieved by AB which are 62.3% and 61.3% respectively with 25 features while the recall attained by XGB (67.9%) which the score same with its recall result with 20 features. This maybe resulted by the algorithm of AB that updated the weight of every misclassified datapoint and use it to build the next classifier. The higher the weight of classifier, the influence of final accuracy will also high. Consider, our data were small, there are high chance that our model classified the data incorrectly. So, by updating the weight of data that wrongly classified and combining these weak classifiers, AB become stronger and can classify better with high accuracy.

However, by seeing the results, it can be concluded that there are several factors that may affect the metric scores of the algorithms, such as, human error while labelling the amount of data that we used, and the number of features that were examined. For the future study, the addition of data should be considered, so, the features that can be used will have more variance.

ACKNOWLEDGMENT

The authors acknowledge the PIT 9 research grant NKB-0011/UN2.R3.1/HKP.05.00/2019 from Directorate Research and Community Services, Universitas Indonesia. The authors also would like to thank Alif Ahsanil Satria, Fari Qodri Andana, Febrian Fikar Farras, M. At Thoriq, and Valerysa Regita for their help in making this work possible.

REFERENCES

[1] L. Wang and J. Q. Gan, "Prediction of the 2017 French election based on Twitter data analysis," *2017 9th Computer Science and Electronic Engineering (CEEC),* Colchester, 2017, pp. 89-93. doi: 10.1109/CEEC.2017.8101605.

[2] M. Ibrahim, O. Abdillah, A. F. Wicaksono and M. Adriani, "Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, 2015, pp. 1348-1353. doi: 10.1109/ICDMW.2015.113.

[3] M. Kantepe and M. C. Ganiz, "Preprocessing framework for Twitter bot detection," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, pp. 630-634. doi: 10.1109/UBMK.2017.8093483.

[4] M. T. Juzar and S. Akbar, "Buzzer Detection on Twitter Using Modified Eigenvector Centrality," 2018 5th International Conference on Data and Software Engineering (ICoDSE), Mataram, Lombok, Indonesia, 2018, pp. 1-5. doi: 10.1109/ICODSE.2018.8705788.

[5] M. Yoshida and F. Toriumi, "Analysis of Political Party Twitter Accounts' Retweeters during Japan's 2017 Election," *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI),* Santiago, 2018, pp. 736-739. doi: 10.1109/WI.2018.000-2.

[6] S. Kiyohara, "A Study on How Technological Innovation Affected the 2008 U.S. Presidential Election: Young Voters' Participation and Obama's Victory," *2009 Ninth Annual International Symposium on Applications and the Internet*, Bellevue, WA, 2009, pp. 223-226. doi: 10.1109/SAINT.2009.51.

[7] L. Ting and Y. Qingsong, "Spam Feature Selection Based on the Improved Mutual Information Algorithm," *2012 Fourth International Conference on Multimedia Information Networking and Security*, Nanjing, 2012, pp. 67-70. doi: 10.1109/MINES.2012.203.

[8] H. Ping and S. Qin, "A Social Bots Detection Model Based on Deep Learning Algorithm," *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, Chongqing, 2018, pp. 1435-1439. doi: 10.1109/ICCT.2018.8600029.

[9] I. Satria, A. Kurnia and Y. Nurhadryani, "Influence of presidential candidates e-campaign towards voters in 2014 presidential election in Bogor City," *2014 International Conference on Advanced Computer Science and Information System*, Jakarta, 2014, pp. 90-95. doi: 10.1109/ICACSIS.2014.7065874.

[10] S. Cresci, R. DiPietro, M.Petrocchi, et al. “The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race,” Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 2017, pp. 963-972.

[11] X. Su and F. Liu, "A Survey For Study of Feature Selection Based on Mutual Information," 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, Netherlands, 2018, pp. 1-4. doi: 10.1109/WHISPERS.2018.8746913.

[12] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," 2016 International Conference on Wireless Communications, Signal Processing and Networking WiSPNET), Chennai, 2016, pp. 2264-2268.