TWITTER BUZZER DETECTION SYSTEM USING TWEET SIMILARITY FEATURE AND SUPPORT VECTOR MACHINE

Ahmad Mustofa1), Fitrah Maharani Humaira2), Myrna Ermawati3), Peni Sriwahyu Natasari4), Akhmad Arif Kurdianto5), Aries Alfian Prasetyo6), and A Labib Fardany Faisal7)

1, 5, 6, 7 Teknik Listrik Industri, Politeknik Negeri Madura
Jl. Raya Camplong KM.4, Taddan, Camplong, Sampang Madura
2) Departemen Informatika dan Komputer, Politeknik Elektronika Negeri Surabaya
Jl. Raya ITS, Keputih, Sukolilo Surabaya
3) Hexagon Mining Balikpapan
Sudirman Office Tower Lt.10 Jl. Jenderal Sudirman No.7, Klandasan Ilir, Balikpapan, Kalimantan Timur
4) Data Science Indonesia
Jl. Mampang Prapat Raya Kav. 100, Pancoran, Duren Tiga, Jakarta
e-mail: mustofahmad@poltera.ac.id1), fitrah@pens.ac.id2), myrnaerma.nina@gmail.com3), penitasari93@gmail.com4), aak@poltera.ac.id5), aries@poltera.ac.id6), a.labib@poltera.ac.id7)

ABSTRACT


Kata Kunci: sosial media, deteksi buzzer, pengolahan text, support vector machine.

ABSTRACT

Over the past few years, more people have turned to social media sites like Twitter as a source of news. This phenomenon may incur many positive and negative effects on society. One of the negative impacts is that public opinion becomes easier to be influenced towards certain opinions that benefit some person. Buzzer accounts usually try to convince people to agree on certain opinions. A system that can determine if an account is a buzzer account or not is required to reduce the negative impact caused by buzzer accounts. The main characteristic of a buzzer account is repeatedly uploading the same content within a certain period. This study uses text processing and classification methods to build a buzzer detection system. The tweet similarity, obtained by applying Cosine Similarity to the Term Frequency - Inverse Document Frequency (TF-IDF), is used as the main feature for the buzzer detection system. In addition, other features such as the number of followers, the number of followings, the intensity of tweets, the ratio of retweets, and the ratio of tweets that contain links are also used as additional features in this study. This study uses these features as input to the Support Vector Machine model to determine whether an account is a buzzer. This system has promising results with 89% accuracy, 86.67% precision, 70.91% recall, and 78% F1 score.

Keywords: social media, buzzer detection, text processing, support vector machine.
I. INTRODUCTION

The significant increase in the use of social media in Indonesia is a good opportunity to seek out a wide variety of information. Social media makes it easier for people to express their opinions and feelings, thus the data become abundant and become one of the primary data resources. By processing those data, public opinion can be disclosed and concluded easily in real time. For example, tweets data from Twitter was utilized to make a prediction system for the winner of the Indonesian presidential election in 2014 and 2018 [1] [2]. Twitter data is also used for demographic analysis of Candidates Supporter during Indonesian Presidential Election in 2019 [3] and to analyze public sentiment toward candidates of US Presidential Election in 2012 [4]. The use of twitter data is not only in political domains, but others also use twitter data to analyze customer’s perception about some products, brands, or events by applying sentiment analysis to user’s tweets [5] [6] [7] [8] [9].

However, at the same time, the abundant information available online which can be accessed freely is risky for those who cannot distinguish which information is fake or genuine. Especially, if the society has a big event such as a public election, people are very sensitive with other opinions especially affected by false information. Buzzer account usually sway public opinion by spreading false information, provoking, insulting, and harassing other accounts. Thus, to get the neutral public opinion, buzzer detection is usually embedded in a system that utilizes data from social media.

From the research conducted by Suciati et al., several classifiers are used to categorize whether a Twitter account is a buzzer or non-buzzer. The research used eleven basic features such as most active days, most active hours, and frequency of most used hashtag resulting the performance of 67.9% [10]. Another promising feature is text-similarity which was used in bot detection system by implementing Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity, having best accuracy of 86% [11]. TF-IDF is commonly used in processing a text-based data, like in [12], TF-IDF is used to make a rumor detection system on twitter data. Other research in text-based data analysis often use cosine similarity together with TF-IDF to calculate similarity between 2 texts [13] [14] [15] [16] [17]. Cosine similarity and TF-IDF can also be used to classify text-based data [18] [19]. An Arabic text classification system is build using TF-IDF, cosine similarity, and some classifier methods in [20] having best result obtained by Support Vector Machine (SVM) classifier. To classify the buzzer and non-buzzer account, Support Vector Machine (SVM) was used in research conducted by Panatra et al. so that neutral information can be maintained for the 2019 presidential election in Indonesia [21]. The use of SVM in a binary class classification system also produces a fairly good performance in detecting glaucoma disease from retinal fundus of eye represented in images [22]. SVM also has a good performance by having 81% accuracy in classifying spam opinions on Steam Review [23]. Others also use SVM for object detection having result 82% accuracy [24].

In this research, we use SVM as the classifier to differentiate buzz and non-buzz accounts. We use the tweet similarity of each user, number of followers, number of followings, intensity of tweets, ratio of retweets, and ratio of tweets that contain links as features led to the SVM classifier. The rest of this paper is arranged as follows: In section 2, we explain the methods used in this research. Then in section 3, we describe the result and analysis of our experiment. Finally, we conclude our work in section 4.

II. RESEARCH METHOD

A. Data Collection

The first task in our research was collecting relevant and appropriate amount of observation data for our classification model approach. We use Twitter public Application Programming Interface (API) to collect data needed in this research. First, we need to collect many twitter users account whose nationality is Indonesian. Since twitter public API didn’t have endpoint which can collect users account data directly, we used tweets collection by keywords endpoint using “Indonesia” as keyword. We also gave limitation to collect tweet that were tweeted by users located in Indonesia. We got 10000 tweets; each tweet contains user’s id information. Then Then, we extracted 1096 distinct users from those tweet data. Endpoint of twitter public API used in this research is provided in Table. 1.

<table>
<thead>
<tr>
<th>No</th>
<th>Endpoint</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tweets/search/all</td>
<td>Endpoint to retrieve tweets with specific keyword and location of user when the tweet is posted</td>
</tr>
<tr>
<td>2</td>
<td>users/id/timelines</td>
<td>Endpoint to retrieve tweets of specific user_id</td>
</tr>
</tbody>
</table>

To detect buzzer, we performed observation on our data and found that Twitter buzzers has specific characteristics such as high frequency of tweets in a short period since buzz accounts actively post tweets to sway other users’ opinion. We add those characteristics as features in our dataset. Later, we will explain these features in next section. To get these features, we collected more tweet from Twitter user accounts. For each user, we retrieved user’s tweet in the last 3 month. If the user’s tweet in the last 3 month is less than 10, we continue to retrieve user’s tweet up to 10 tweets.
To apply supervised machine learning model to detect Twitter buzzer account, we need to obtain labeled data for each user accounts. Labeling data to provide dataset for our classification model approach usually requires human effort. In the next step, we observed Twitter users and manually labeled those accounts as ‘buzzer’ or ‘non-buzzer’. Twitter user that continuously promoting or discrediting the same things multiple times are labeled as ‘buzzer’. We involved three annotators to label those Twitter user by manual observation. Final label of each user is decided using a voting system. We found that 395 from the total of 1096 users from our data are labeled as buzzer.

B. Buzzer Classifier Features

To differentiate buzzers among normal Twitter users, we implemented a machine learning approach to build our detection model. We defined a buzzer detection problem as a classification problem that will be able to classify a Twitter user as a buzzer or not. In this research, we used statistical property of tweet similarity as the main feature together with other features like follower count and tweet per day. Table 2 and Table 3 displays the list of Twitter user-based features and tweet-based features for our classification approach.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1</td>
<td>Follower count</td>
</tr>
<tr>
<td>No. 2</td>
<td>Following count</td>
</tr>
<tr>
<td>No. 3</td>
<td>Tweet per day</td>
</tr>
<tr>
<td>No. 4</td>
<td>Retweet ratio</td>
</tr>
<tr>
<td>No. 5</td>
<td>Link ratio</td>
</tr>
<tr>
<td>No. 6</td>
<td>Tweets similarity</td>
</tr>
</tbody>
</table>

Table 2 List of Proposed Features

<table>
<thead>
<tr>
<th>No. 1</th>
<th>Follower count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>count</td>
</tr>
<tr>
<td></td>
<td>number of users who follow each user</td>
</tr>
</tbody>
</table>

User-based features like follower count and following count is retrieved directly from data provided by twitter public API. Twitter public API also provide accounts’ creation time and total number of tweets. These two data used to calculate tweet per day feature.

C. Tweet Based Feature

To obtain tweet-based feature used in this research such as retweet ratio, link ratio, and tweets similarity, we need to retrieve tweets data from each user and perform several text preprocessing steps as follows:

1. Tweet Retrieval
   For each user, we use twitter public API to retrieve user’s tweet in the last 3 months. If the user’s tweet in the last 3 months is less than 10, we continue to retrieve user’s tweet up to 10 tweets.
   Tweets data retrieved from twitter public API contains information whether a tweet is a retweet or not. For each user, we count how many tweets is a retweet and divide the result with total tweet retrieved to get retweet ratio feature.
   Tweet content is also provided by twitter public API which value is used to calculate link ratio feature. For each user, we count number of tweets that contain links and divide the result with total number of tweets retrieved.

2. Tweet Preprocessing
   After user’s tweets retrieved, we applied a filtering process, the stop word removal, to remove irrelevant words or characters. This preprocessing stage removed all words in stop words lists from tweets, including Twitter-specific words like RT. We also performed non-letter specific character removal, link removal, and mention removal in this preprocessing phase.

3. Tweet Vectorizing
   In the next preprocessing stage, we need to vectorize each tweet to represent tweet in a vector space and calculate its similarity. We use a term frequency – inverse document frequency (TF-IDF) method to vectorize each tweet. The vectorization step is processed locally for each user.
   This vectorizer takes tweets and tokenize them to obtain terms, and then computing term frequency – inverse document frequency (TF-IDF) as weighting method. The term frequency (TF) value is obtained by counting the number of occurrences of a term in one tweet, while the inverse document frequency value is obtained by (1) [25].
   \[ \text{idf}_i = \log_2 \frac{N_t}{df_i} \]  

4. Tweet Similarity Feature
   To extract tweet similarity feature of each user, we need to build a matrix of tweet similarity. Suppose we have a user with 8 tweets, the matrix of tweet similarity of this user will look like Fig. 1.
Fig. 1 Matrix of Tweet Similarity

Similarity between tweet is calculated using cosine similarity in (2) [26] [27].

\[ S_{i,j} = \frac{\sum_{k=1}^{n} w_{i,k} \times w_{j,k}}{\sqrt{\sum_{k=1}^{n} w_{i,k}^2} \times \sqrt{\sum_{k=1}^{n} w_{j,k}^2}} \]  

After matrix of tweet similarity is formed, we can extract the similarity feature of the matrix using (3) and (4):

\[ f_1 = \frac{\sum_{i=1}^{n} \sum_{k=1}^{n} S_{i,k} - 1}{N(N-1)} \]  

\[ f_2 = \frac{\sum_{i=1}^{n} \sum_{k=1}^{n} (S_{i,k} - f_1)^2}{N-1} \]

D. Classifier: SVM

Support Vector Machine (SVM) was used as the classifier to differentiate buzzer and non-buzzer accounts. Our research used six features, as listed before, as features for the binary classification using SVM.

After feature extraction process is done, we combine all features extracted to represent each user in the buzzer detection system. Then all the user training data used to train classification model. Finally, we used all the user testing data to validate the trained model. Fig. 2 describes our proposed buzzer detection system.

III. RESULT AND DISCUSSION

A. Dataset

As the result of tweet data collection and user account extraction, we have 1096 distinct user accounts that had been labeled manually. Buzzer characteristic features from these user account data and the label are then used as our research dataset. To build a classification model for buzzer detection, we split our dataset into training data and testing data. Total dataset for training is 896 user account data that will be used to train our model. The rest of the dataset will be used as testing data to validate our model. Fig. 3 describes the distribution of our data.

B. Result

We use Support Vector Machine (SVM) as the classification method in our proposed buzzer detection system as SVM works well on a binary classification case. In this research, each user is represented by a 1x7 vector which will be used as features in SVM. Before training our model, we analyze the value range of each non-statistical feature to decide whether a normalization process is necessary or not.

Table 4 Value Range

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>Value Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Follower count</td>
<td>0-289588</td>
</tr>
<tr>
<td>2</td>
<td>Following count</td>
<td>0-6849</td>
</tr>
<tr>
<td>3</td>
<td>Tweet per day</td>
<td>0.0005-83.333</td>
</tr>
<tr>
<td>4</td>
<td>Retweet ratio</td>
<td>0.0-0.92</td>
</tr>
<tr>
<td>5</td>
<td>Link ratio</td>
<td>0-1</td>
</tr>
</tbody>
</table>

From Table. 3, we can see that the value range is variable. We decide to do normalization process for feature 1-4 so that we have same value range for all 5 features (0-1).

We experiment using 3 different kernels on SVM such as linear, radial basis function (rbf), and polynomial kernel. The result of our experiment is shown in Table 4.

Table 5 Experiment Result

<table>
<thead>
<tr>
<th>No.</th>
<th>Kernel</th>
<th>Accuracy (%)</th>
<th>Recall (%)</th>
<th>F1 Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear</td>
<td>86.5</td>
<td>60</td>
<td>70.97</td>
</tr>
<tr>
<td>2</td>
<td>Polynomial</td>
<td>87.5</td>
<td>65.45</td>
<td>74.23</td>
</tr>
<tr>
<td>3</td>
<td>RBF</td>
<td>89</td>
<td>70.91</td>
<td>78</td>
</tr>
</tbody>
</table>

Considering our research plan, we need a buzzer detection system that has a good recall and f1 score. From
Table 4, we achieve the best result when using RBF kernel.

IV. CONCLUSION

Data from social media can be used to achieve many purposed, like predicting election result, analyzing product review, etc. Buzzer detection is needed to get an objective and valid result. In the next step of our research plan, we will build a system that rely on people’s opinion in social media, so we need a good buzzer detection system.

Our experiment shows that our proposed buzzer detection system can distinguish buzzer user from normal user well. Experiments using other classification method can be done to get a better result. Another approach that we assume can improve the performance of the buzzer detection system is implementing similarity threshold on the tweet similarity feature extraction.

NOTATION

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_t$</td>
<td>the total number of tweets</td>
</tr>
<tr>
<td>$d_i$</td>
<td>number of tweets containing term-i</td>
</tr>
<tr>
<td>$n$</td>
<td>number of unique terms in the bag of words (BOW) of the corresponding user</td>
</tr>
<tr>
<td>$S_{ij}$</td>
<td>similarity of tweet i and tweet j</td>
</tr>
<tr>
<td>$w_{kj}$</td>
<td>the weight of term k in tweet i</td>
</tr>
<tr>
<td>$w_{j,k}$</td>
<td>the weight of term k in tweet j</td>
</tr>
<tr>
<td>$N$</td>
<td>number of tweets of the corresponding user</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENT

We would like to thank Direktorat Riset dan Pengabdian Kepada Masyarakat Republik Indonesia for funding this research, and Pusat Penelitian dan Pengabdian kepada Masyarakat Politeknik Negeri Madura for the support to this research.

REFERENCES


