

PAPER • OPEN ACCESS

## Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method

To cite this article: A Fauzi *et al* 2019 *J. Phys.: Conf. Ser.* **1192** 012025

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

# Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method

**A Fauzi<sup>1</sup>, E B Setiawan<sup>2</sup> and Z K A Baizal<sup>3</sup>**

School of Computing, Telkom University, Telekomunikasi street 01, Terusan Buah Batu, Bandung 40257, Indonesia

E-mail: <sup>1</sup>fauziachmad@student.telkomuniversity.ac.id,

<sup>2</sup>erwinbudisetiawan@telkomuniversity.ac.id,

<sup>3</sup>baizal@telkomuniversity.ac.id

**Abstract.** Twitter is one of the social media that is currently popularly used around the world. It's just that twitter has some problems that adversely affect its users. Hoax is one of the negative things that often occur in social media, news in the hoax is still doubted the truth or the fact. In this final project, the authors built a system to detect hoax news on twitter. The purpose of the research is to minimize the hoax news spread on twitter. The use of the Term Frequency Inverse Document Frequency (TF-IDF) weighting system in the system gives a weighted value to a tweet taken from the occurrence of a hoax news sent by someone on Twitter. Data classification uses the Support Vector Machine (SVM) method of the system to predict the possibility of a twitter account user spreading a hoax news based on the user's behavior. Testing data is done based on the contents of content tweets. Datasets are arranged based on attributes used such as the number of retweets, URLs, number of hashtags, provocations, feuds, anxieties, and unbalanced news. Processed data is divided into training data and testing data. The result of data using all features get the highest accuracy is 78,33%. The contribution of this research is that it can detect news that has a tendency towards hoax and can filter which is classified as hoax or not hoax.

**Keywords:** social media, twitter, hoax, TF-IDF, SVM

## 1. Introduction

Technological developments make communication easier with social media such as Facebook, Twitter, Path, Instagram, and the like. Social media is communication that is on the internet network so that users can interact without limitation of distance or time. Twitter is a microblogging-based social media site that was launched on July 13, 2006 [1]. A tweet contains a short message with a maximum of 140 characters [2]. On Twitter social media there are a lot of positive and negative things presented. Twitter besides making it easier for people to share information and interact can also have a negative impact on users, such as misuse of information (hoaxes, gossip, pornography, fraud, defamation, etc.). Hoaxes are news that is published without being based on truth or false news [3]. In general, there are hoaxes when someone tries to deceive the other party into believing in an event that did not occur [4]. Hoaxes have become a hot conversation in the mass media and social media lately because they are considered disturbing the public with information that cannot be ascertained. The purpose of the research is to minimize the hoax news spread on twitter. Thus, to be able to recognize the characteristics of users of the Twitter social networking site that have the potential to spread hoax news, a method is needed that can learn and read these characteristics.

## 2. Related Work

In this study the author conducted hoax news detection using the weighting Term Frequency Inverse Document Frequency (TF-IDF) which is the method used in weighting the appearance of words in a document [9]. For classification of data using the SVM algorithm, this algorithm uses a hyperplane to separate classes [8]. There may be some hyperplane, but SVM tries to find the best through measuring hyperplane margins and finding the maximum point. The SVM method is considered a suitable method for



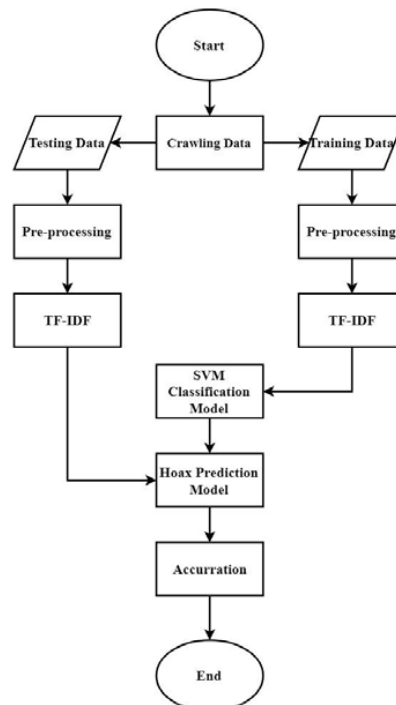
detecting user behavior that spreads hoaxes.

Support Vector Machine (SVM) is a classification method that works by finding the hyperplane with the biggest margin. Hyperplane is an inter-class data separator line. Margin is the distance between the hyperplane and the closest data in each class. The nearest data with hyperplane in each class is called support vector [5]. Basically, SVM is a method used for the classification of two classes (binary classification).

The problem that becomes the limitation of work in this final project is the user behavior that is entered into features such as the number of retweets, URLs, number of hashtags, provocation, feuds, anxiety, unbalanced news and hoax news scope (user or tweet) using Indonesian. The purpose of this final project is to implement the TF-IDF weighting and SVM method to detect hoax news on Twitter, find out how much accuracy results are generated from the system made, and find out the features that affect hoax detection.

### 3. Propose Method

The steps that must be taken to support related work are in the flowchart below in Figure 1.



**Figure 1.** Flowchart Hoax Detection System.

#### 3.1. Crawling

The data crawling phase has the purpose of obtaining data that will be used as reference data by the system. Reference data in the form of user and content. User data is obtained manually from Twitter, then given a class label based on the labelling technique. For hoax detection system, we get 10 features. These features are taken when crawling data and shown in Table 1.

**Table 1.** List of Hoax Features.

No	Feature	Description
1	Retweet	The number of retweets from the tweet
2	Urls	Tweet contains url or not
3	Hashtag	Tweet contains hashtag or not
4	Mention	Tweet contains mention or not
5	Positive Sentiment	The number of positive words in the tweet
6	Negative Sentiment	The number of negative words in the tweet
7	Provocation	Tweets contain the provocation words
8	Feuds	Tweets contain the hostility words

9	Anxieties	Tweets contain the anxiety words
10	Unbalanced News	Tweet containing news that discredit other parties

### 3.2. Pre-processing Data

Pre-processing data is the steps in processing the raw data that will be included in the classification system. This process aims to prepare data that will be used efficiently into the classification system [4]. The process in pre-processing data used in the built system is 4, namely case folding, tokenizing, stop word removal and stemming.

1. **Case folding** is the process of converting capital letters from input data to lowercase all (lowercase). This is done so that all letters in the input data become uniform.
2. **Tokenizing** is a process to separate words separated by spaces. This is done to facilitate the next pre-processing stage.
3. **Stop word removal** at this stage tweets still contain words that are considered to have no influence in determining classifications such as conjunctions. These words are entered the stop word list. If it is included in the stop word, the word will be deleted or removed from the tweet.
4. **Stemming** is the process of eliminating the word that has become a basic word. This can be done by removing the prefix or suffix from a word.

### 3.3. TF-IDF

Term Frequency Inverse Document Frequency or TF-IDF is a method used in weighting the word position in a document. TF states the number of words that appear in the document. While IDF shows the price level used in the document. On TF-IDF there is a formula for calculating the weight ( $W$ ) of each document for keys [9]. This method will calculate each token  $t$  in document  $d$  with the formula:

$$W_{ij} = tf_{ij} * IDF_j \quad (1)$$

After the weight ( $W$ ) of each document is known, then the sorting process is carried out where the greater the  $W$  value, the greater the level of similarity of the document to the keyword, and vice versa.

### 3.4. Support Vector Machine

The concept of SVM can be simply explained as an attempt to find the best hyperplane that serves as a separator of two classes in the input space [6, 7]. In linear separable data, existing data can be separated linearly. Suppose there are  $\{x_1, x_2, \dots, x_n\}$  patterns that are members of two classes, namely +1 and -1. Each pattern incorporated in each class has its own characteristics. Linear classification produces a weight vector  $w$  with the sign function ( $w^T x$ ) not with the sigmoid function. SVM does not use pure probability values, but will use margins or distance of document vectors as truth values. Farther the point from the hyper field is, the higher the probability of that point can be classified. The conditional probability formula of the regression function is defined by [10]:

$$P(y|x) = \frac{1}{1 + e^{(-yw^T x)}} \quad (2)$$

Where,  $y = \pm 1$ ,  $x$  is data,  $y$  is the class label, and  $w \in R^n$  is the vector weight,  $T$  is hyper-parameter. Vector training becomes higher or may be unlimited with the function  $\Phi$ . Then SVM separates the hyperplane with the maximum margin in a higher dimensional space.  $C > 0$  is the boundary parameter for the error value. Next,  $K(X_i, X_j) \equiv \Phi(X_i)^T \Phi(X_j)$  is called the kernel function.

## 4. Result and Analysis

In this paper, we use 1800 data tweets from crawling results with topics about 2019 presidential election using the hashtag listed in Table 2 below:

**Table 2.** List of Topics.

No	Hashtags	Amount of Tweets
	#2019GantiPresiden	445
	#TetapJokowi	254
	#PrabowoPresiden	389

2019 presidential election	#2019TetapJokowi	148
	#salam2periode	105
	#jkw2periode	114
	#PresidenBaru2019	72
	#PrabowoPresiden2019	102
	#2019PrabowoPresiden	83
	#kerjanya	88

A list of provocation words can be seen in Table 3.

**Table 3.** List of Provocation Words.

List of Provocation Words						
ayo	sebar	jangan	bagikan	bergabunglah	hindarilah	pilih
mari	ayolah	janganlah	ajak	ikut	waspada	pilihlah
viral	marilah	ganti	ajakan	ikutilah	waspadalah	berbagi
viralkan	turun	penjarakan	gabung	hindar	ancam	menang
sebar	turunkan	bagi	bergabung	hindari	ancaman	menanglah
datang	kunjungi	rapat	kumpul	kunjungilah	merapatkan	berkumpul
datanglah						

A list of hostility words can be seen in Table 4.

**Table 4.** List of Hostility Words.

List of Hostility Words					
emosi	sombong	marah	bantah	sebal	berantem
singgung	iri	hina	cekcok	dongkol	pukul
paksa	tawuran	ejek	salah	benci	sial
cengkeram	pura-pura	kelahi	salahpahaman	kecewa	
dengki	musuh	bertengkar	kesal		

A list of anxiety words can be seen in Table 5.

**Table 5.** List of Anxiety Words.

List of Anxiety Words				
stres	bercerai	gugup	gundah	gusar
khawatir	takut	galau	sedih	gelisah
bimbang	tegang	resah	bingung	cemas
trauma	panik			

Meanwhile, the total data classified as hoax and not hoax can be seen in Table 6.

**Table 6.** Hoax vs No Hoax.

Total Tweet	Hoax	No Hoax
1800	1211	589
	67,28%	32,72%

The new things that are applied are the addition of supporting corpus for hoax detection such as provocation, hostility, anxiety. The results of the corpus were obtained from expert discussions namely Indonesian Language lecturers. We use two scenarios. The first scenario changes the composition of training data and testing data from the system built and the second scenario determines the features that influence the hoax news.

#### 1. Scenario 1

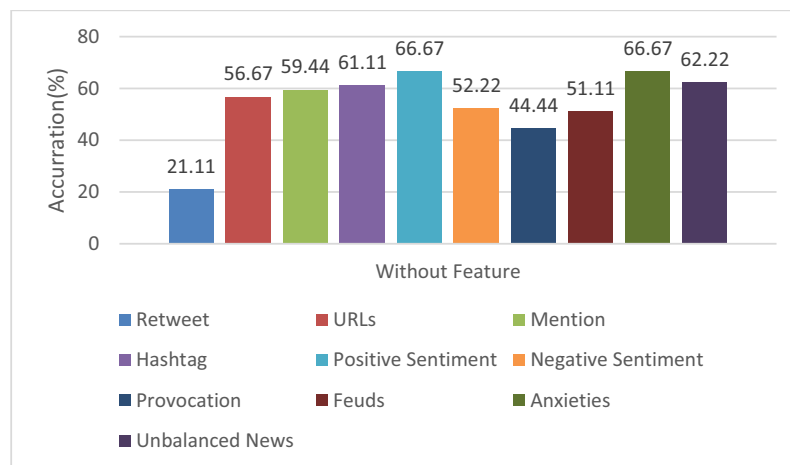
It can be seen from Figure 2 below that the highest accuracy value at 90:10 data ratio is 78,33%.



**Figure 2.** Chart of comparing data.

## 2. Scenario 2

The influential features for hoax detection can be seen in Figure 3 below, retweet and URL features that exist on Twitter and hoax detection support features is provocation and feuds.



**Figure 3.** Chart of without features.

Based on the results of the tests carried out, the interesting results are the results of the program namely feature support features of the hoax that is very beneficial in the process of labelling data to note the contents of tweets classified as hoaxes or not hoaxes, the process of pre-processing is very influential on output data when testing the system made, data that has been chopped and labelled is very beneficial on the results of accuracy.

## 5. Conclusion

Based on the results of the testing and analysis carried out by the author, the authors draw some conclusions that the TF-IDF method and SVM classification can be used to detect hoax news on twitter, the results of performance testing with analyzed accuracy of the system made obtained the best ratio at 90:10 with a value of 78,33%, and features that affect the hoax news class include retweet and URL features of Twitter and hoax detection support features, namely provocation and feuds. So that this research can be useful for the world of education, especially for teachers and students, it is deemed necessary to conduct further research based on the results of this study, namely: Suggestions for further researchers to be able to conduct research on other subjects because this research only takes one subject, namely the 2019 presidential election. Therefore there is a need for research on other subjects.

## References

- [1] Asur, Sitaran dan Bernardo A. Huberman. *Predicting The Future With Social Media*. Proceeding WI-IAT '10 Proceeding of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Volume 01. Halaman 492-499. 2010.

- [2] McCord, M dan M. Chuah. 2011. *Spam Detection on Twitter Using Traditional Classifiers*. Lehigh University, Bethlehem.
- [3] Situngkir, Hokky. (2011). *"Spread Of Hoax in Social Media"*. Bandung : Department Computational Sociology.
- [4] J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques Third Edition*, San Fransisco: Morgan Kauffman Publishers, 2012.
- [5] Yunliang, J., Qing, S., Jing, F., & Xiongtao, Z. 2010. *The Classification for E-government Document Based on SVM*. In *Web Information Systems and Mining (WISM)*, 2010 International Conference on. Vol. 2: 257-260.
- [6] James Sanger Ronen Feldman, *The Text Mining Handbook*. New York: Cambridge University Press, 2007.
- [7] Gunn, S.R. 1998. *Support Vector Machines for Classification and Regression*. *ISIS technical report*.
- [8] J.Z. Liang. 2004 "SVM Multi-Classfier And Web Document Classification", *Proceedings of the IEEE Third International Conference on Machine Learning and Cybernetics*.
- [9] J. Han dan M. Kamber, *Data Mining : Concepts and Techniques Second Edition.*: Series Editor Morgan Kaufmann, 2006.
- [10] Prasetijo, A. B., Isnanto, R. R. & Eridani, D., 2017. *Hoax detection system on Indonesian news sites based on text classification using SVM and SGD*. Semarang, IEEE, pp. 45 – 49.