Scientific Journal of Informatics
Vol. 7, No. 2, November 2020

p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Fake Twitter Account Classification of Fake News Spreading Using Naïve Bayes

Heru Agus Santoso[1], Eko Hari Rachmawanto[2], Ulfa Hidayati[3]

[1,2]Informatics Engineering Department, Faculty of Computer Sciences,
Universitas Dian Nuswantoro, Indonesia
Email: [1]heru.agus.santoso@dsn.dinus.ac.id, [2]eko.hari@dsn.dinus.ac.id,
[3]111201609388@mhs.dinus.ac.id

Abstract

Twitter is a very popular microblog, where users can search for various information, current news, celebrity posts, and hot topics. Indonesia is ranked 5th for the most Twitter users. The large number of users makes Twitter used for the benefit of certain parties with bad goals, such as spreading fake news using fake accounts. Fake accounts are often used by several parties to spread fake news, therefore the spread of fake news must be immediately limited to minimize the negative impact caused by fake news. For this reason, this research is written with the aim of being able to classify fake and genuine Twitter accounts. In this study, using data mining techniques that are closely related to big data in decision making by applying the Naive Bayes method. Naïve Bayes is one of the most widely used classification methods because it has good accuracy and faster computation time. Here, we proposed Naïve Bayes to classify fake twitter account because it is very important in recognizing fake twitter which is detrimental to many parties. It is different from previous studies which required 16 parameters, in this study we only used 9 parameters and had successfully classified them with high accuracy. The classification process uses nine parameters, namely based on the Profile Created, Favorite Count, Follower Count, Following Count, Geo Enabled, Follower Rate, Following Rate, Follower Following Ratio, Verified. This study uses 210 datasets of twitter accounts that spread fake news, the result is that Naïve Bayes works very promising in the classification of fake twitter accounts and in the testing process using 5% of training set produces an accuracy of 80%.

Keywords: social media, Twitter, Fake account, Naïve Bayes

1. INTRODUCTION

Users can post short text messages called tweets which are limited by 280 characters and can be seen by followers [1]. According to Smith and Brenner in 2012, Some 15% of online adults use Twitter as of February 2012, and 8% do so on a typical day. Although overall Twitter usage has nearly doubled since the Pew Research Center's Internet & American Life Project first asked a stand-alone Twitter question in November 2010, the 15% of online adults who use Twitter as of early 2012 is similar to the 13% of such adults who did so in May 2011 [2]. Boukes [3] was investigated the uptake of current affairs knowledge as the outcome of social network usage in the context of The Netherlands, a country particularly well-suit to address this question, because of its high Internet penetration (96% of population) and the relative popularity of the social networks Facebook (62%) and Twitter (15%).

Social media plays a big role in influencing society and some people try to take advantage of this situation. Sometimes the social media manipulate information in their own way to achieve their goals [4]. There are many websites that provide false information. They try to issue propaganda, hoaxes and misinformation to make news [5]. Their main purpose is to manipulate information that can make people believe in it. Therefore, fake news affects the minds of many people [6]. Fake accounts are often used by several parties to spread fake news [7-8], therefore the spread of hoax news must be immediately limited to minimize the negative impact caused by hoax news. Twitter facilities continue to grow when viewed from the negative side, it has resulted in opportunities for manipulation and fraud, one of which is the appearance of tweets from fake accounts even though users can report spam or links directly from the attacker's profile or Tweet. Fake accounts can cause various problems such as spam and spreading untrue information [9], hate speech and some tweets that harm some parties [10]. Twitter has tried to combat the existence of fake accounts, but so far this has not been completely successful. According to [11], to help achieve this goal, we have introduced new measures to combat abuse and trolls, new policies on hate speech and violence, and use new technology and add resources to combat spam and abuse.

The emergence of fake accounts requires handling, one of which is by carrying out an account classification pattern so that information appears on the existence of real accounts and fake accounts [12]. Classification patterns require the implementation of appropriate method, for example data mining. According to research conducted by [13], the data mining model using the Gaussian medium Support Vector Machine algorithm has been tested to determine fake accounts, but fake accounts must function on the network in order to be recognized as legitimate accounts or fake, by analyzing their network of friends. The Naïve Bayes algorithm has been used by [14] in 2009 to classify uncertain data using the UCI dataset. Research on the detection of fake accounts on social media was carried out by [15] in 2018 using SVM-NN, while [16] detected fake news on twitter using machine learning. Naïve Bayes has also been used by [17] in detecting name spam on LinkedIn social media. Research conducted by [18] has compared the performance of the decision tree, naïve Bayes and the neural network on web training data classification and concluded that naïve Bayes has a good performance in the classification process.

Naïve Bayes is one of the most widely used classification methods because it has good accuracy and faster computation time. Naive Bayes uses probability and statistical methods in accordance with that put forward by a British scientist named Thomas Bayes. Then according to [19] explains that Naïve Bayes is a class of decisions, using mathematical probability calculations on the condition that the decision value is correct, based on object information. The advantage of using Naive Bayes is that this method only requires a small amount of training data to determine the parameter estimates required in the classification process [20]. According to [10] that used 16 parameters, we proposed Naïve Bayes with 9 parameters to produce a high accuracy in classification twitter fake account.

2. METHODS

Naïve Bayes is a predictive technique using probability at the limits of the Bayes theory [21] which aims to classify certain classes [22]. In this phase, a selection is made about the needs of the data in order to achieve data mining goals. The selected data are those that are relevant to the research and the operations are carried out as shown in Figure 1. In this study, data selection was carried out as follows:

1) Looking for hoax news on trusted sites, in this study using sources from TurnBackHoax.com and Cekfakta.com
2) Looking for the news on twitter.
3) Find accounts that spread the hoax.
4) There were 210 twitter accounts that spread hoax news.
5) Doing the crawling process to retrieve the account data data.
6) The process of selecting the parameters needed to classify a fake twitter account or not.
7) This study uses 9 parameters as determinants of the classification of fake or genuine Twitter accounts.
8) The 9 parameters are Profile Created, Favorite Count, Follower Count, Following Count, Geo Enabled, Follower Rate, Following Rate, Follower Following Ratio, Verified.
9) A crawling process is carried out to retrieve Created Profile data, Favorite Count, Follower Count, Following Count, Geo Enabled, Follower Rate, Following Rate, Follower Following Ratio, Verified on the twitter account that will be classified.

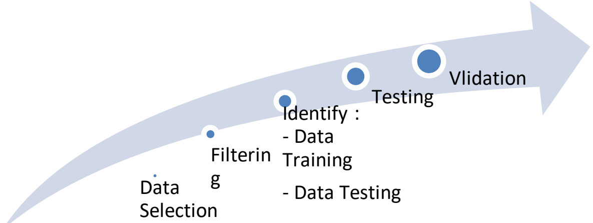


Figure 1. Proposed Method Using Naïve Bayes Classifier

3. RESULT AND DISCUSSION

The data to be used is in the form of information data from a twitter account that spreads hoax news consisting of 210 accounts. Then the required attributes to classify into the fake account category or not. Here are nine data attributes including: Profile Created, Favorite Count, Follower Count, Following Count, Geo Enabled, Follower Rate, Following Rate, Follower Following Ratio, and Verified as shown in Table 1.

Table 1. Sample of Training Data

No	PC	FC	FoC	FollC	GE	FR	FoR	FollR	V	L
1	3	0	0	0	0	0	0	1	1	1
2	1	1	1	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0
4	1	1	1	0	0	0	0	0	0	0
5	1	1	1	0	0	0	0	0	0	0
6	1	0	0	0	0	2	1.3	0	0	0
7	1	0	0	0	0	0	0	0	0	0
8	1	0	0	1	1	0	0	0	0	0
9	1	0	0	1	1	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0	0

Description : PC = Profil Created, FC = Favorite Count, FoC = Follower Count, FollC = Following, Count, GE = Geo Enabled, FR = Follower Rate, FoR = Following Rate, FollR = Follower Following Ratio, V = Verfied, L = Label

The data in Table 1 above is the normalized crawl data as follows:

1) Created profile, which is the account creation date in years.
2) Favorite Count, The number of Favorite Count is simplified into 3 categories, namely [Less than 500 = 1 | 500-1000 = 2 | More than 1000 = 3]
3) Follower Count, The number of Follower Count is simplified into 3 categories, namely [Less than 500 = 1 | 500-1000 = 2 | More than 1000 = 3]
4) Following Count, The number of following counts is simplified into 3 categories, namely [Less than 500 = 1 | 500-1000 = 2 | More than 1000 = 3]
5) Geo Enabled, Geo Enabled is divided into 2 categories, namely [Yes = 1 | None = 0]
6) Follower Rate, The total Follower Rate is simplified into 3 categories, namely [Less than 500 = 1 | 500-1000 = 2 | More than 1000 = 3]
7) Following Rate, The number of following rates is simplified into 3 categories, namely [Less than 500 = 1 | 500-1000 = 2 | More than 1000 = 3]
8) Follower Following Ratio, The number of followers following ratio is simplified into 3 categories, namely [Less than 500 = 1 | 500-1000 = 2 | More than 1000 = 3]
9) Verified, It can be ascertained that if the account is verified, it is a real account.
10) Label, Labels in the form of classification prediction results are categorized into 2, namely [Fake account = 1 | Original Account = 0]

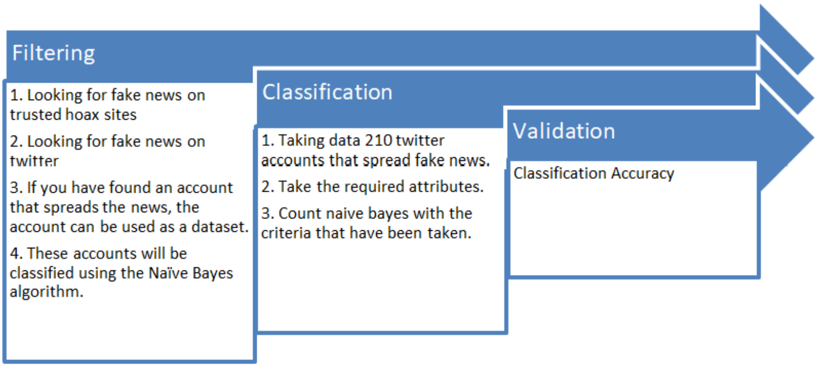
Test data is new data whose class label is not yet known and the data grouping will be sought using Naive Bayes which has been implemented into the system with training data as a reference. The test data used in this study were 10 data on hoax news spreader twitter accounts as shown in Table 2.

Table 2. Testing data

No	Twitter Username	Class
1	GunRomli	?
2	Cacienx	?
3	Muhsinlabib	?

4	MalayCyberForce	?
5	koestoer2000	?

The test data is in the form of a twitter account username, then a crawling process has been carried out to obtain data from the classification determining parameters.



Filtering

1. Looking for fake news on trusted hoax sites
2. Looking for fake news on twitter
3. If you have found an account that spreads the news, the account can be used as a dataset.
4. These accounts will be classified using the Naïve Bayes algorithm.

Classification

1. Taking data 210 twitter accounts that spread fake news.
2. Take the required attributes.
3. Count naive bayes with the criteria that have been taken.

Validation

Classification Accuracy

Figure 2. Steps of Filtering, Classification and Validation

Based on Figure 2, the filtering stage uses data taken from TurnBackHoax.com and Cekfakta.com sources, the classification stage is divided into 3 sub-stages where the last stage is operating Naïve Bayes through the process: (1) Count the amount of data, (2) Calculating the attributes, (3) Multiply all data by attributes, (4) Compare the results of the multiplication between fake and non-fake. In the validation stage, the calculation of accuracy is as in equation (1).

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{total predictions}}\ x\ 100\% \qquad (1)$$

Whereas in Table 3, there are several calculations such as Equations (2) to (4).

$$\text{Follower Rate} = \frac{\text{Follower Count}}{\text{Profil Age}} \qquad (2)$$

$$\text{Following Rate} = \frac{\text{Following Count}}{\text{Profil Age}} \qquad (3)$$

$$\text{Following Following Ratio} = \frac{\text{Following Count}}{\text{Follower Count}} \qquad (4)$$

Table 3. Sample of Training Data

No	PC	FC	FoC	FollC	GE	FR	FoR	FollR	V	L
1	3	3	3	3	0	0	0	1	0	1
2	8	1	3	1	1	0	0	1	0	1
3	6	1	1	2	1	0	0	1	0	1
4	10	1	3	3	1	0	0	1	1	0
5	8	1	3	1	1	0	0	1	1	0

6	0	1	1	1	0	0	0	1	0	1
7	8	1	1	1	1	0	0	1	0	1
8	5	1	3	1	1	0	0	1	1	0
9	9	1	3	1	1	0	0	1	1	0
10	8	2	3	1	1	0	0	1	1	0
11	9	3	1	1	1	0	0	1	0	1
12	0	1	1	1	0	0	0	1	0	1
13	9	1	3	1	1	0	0	1	1	0
14	9	1	3	1	0	0	0	1	1	0
15	10	1	3	1	1	0	0	1	1	0
	0	1	1	1	0	0	0	1	0	**?**

Description : PC = Profil Created, FC = Favorite Count, FoC = Follower Count, FollC = Following, Count, GE = Geo Enabled, FR = Follower Rate, FoR = Following Rate, FollR = Follower Following Ratio, V = Verfied, L = Label

The results in Table 3, can be obtained from the calculation:

1) Calculate the total probability of each class of events P (Ci)
   P (Label) = 1 = 7/15 = 0.46
   P (Label) = 0 = 8/15 = 0.53
2) Calculating the Probability of Variable Details in Class
   P (x | Ci) = Profile Variable Created
   P (Profile Created = 0 | Label = 1) = 2/7 = 0.28
   P (Profile Created = 0 | Label = 0) = 0/8 = 0
   Variable Favorite Count
   P (Favorite Count = 1 | Label = 1) = 5/7 = 0.71
   P (Favorite Count = 1 | Label = 0) = 7/8 = 0.87
   Variable Follower Count
   P (Follower Count = 1 | Label = 1) = 5/7 = 0.71
   P (Follower Count = 1 | Label = 0) = 0/8 = 0
   Variable Following Count
   P (Following Count = 1 | Label = 1) = 5/7 = 0.71
   P (Following Count = 1 | Label = 0) = 7/8 = 0.87
   Geo Variable Enabled
   P (Geo Enabled = 0 | Label = 1) = 3/7 = 0.42
   P (Geo Enabled = 0 | Label = 0) = 1/8 = 0.12
   Follower Rate Variable
   P (Follower Rate = 0 | Label = 1) = 7/7 = 1
   P (Follower Rate = 0 | Label = 0) = 8/8 = 1
   Variable Following Rate
   P (Following Rate = 0 | Label = 1) = 7/7 = 1
   P (Following Rate = 0 | Label = 0) = 8/8 = 1
   Variable Follower Following Ratio
   P (Follower Following Ratio = 1 | Label = 1) = 7/7 = 1
   P (Follower Following Ratio = 1 | Label = 0) = 8/8 = 1
   Variable Verified
   P (Verified = 0 | Label = 1) = 7/7 = 1
   P (Verified = 0 | Label = 0) 0/8 = 0

3) Multiplying All the Variables Class =
   Calculating P (x | Label = 1) = 0.28 * 0.71 * 0.71 * 0.71 * 0.42 * 1 * 1 * 1 ** 1 = 0.042
   Calculating P (x | Label = 0) = 0 * 0.75 * 0 * 0.87 * 0.12 * 1 * 1 * 1 * 0 = 0
4) Comparing the results between classes
   P (X | Ci) * P (Ci) = P (X | Label = 1) * P (Label = 1) = 0.057 * 0.46 = 0.019
   P (X | Label = 0) * P (Label = 0) = 0 * 0.53 = 0
   Because P (1) <P (0) then the decision taken is 1.
   Then it is included in label 1 (Fake)
   Description: 1 = Fake
                0 = Non-Fake

Then performed testing on 10 twitter accounts using the Naïve Bayes algorithm calculation and produce the classification results as in Table 4 and Table 5 below.

Table 4. Classification Results

No	Twitter Username	Class
1	DaeIm85	Fake
2	detikinet	Non-Fake
3	SANTRISALFY	Fake
4	eramuslim	Fake
5	kompasiana	Non-Fake
6	DKIJakarta	Non-Fake
7	KariYan38483426	Fake
8	voaindonesia	Non-Fake
9	kewlfee	Non-Fake
10	sulasman_aditya	Fake

Table 5. Accuration Results

No	Twitter Username	Prediction	Results
1	ridwanaedhy	Non- Fake	Fake
2	IwanDar26033245	Fake	Fake
3	NarangNarangi	Fake	Fake
4	HeriSenoAji	Fake	Fake
5	alif2000nur	Fake	Fake

By using 5 data samples, 4 data are suitable so that you get an accuracy of 80%.

4. CONCLUSION

The application of fake twitter account classification for hoax news spreaders using the Naïve Bayes algorithm uses 9 data attributes including: Profile Created, Favorite Count, Follower Count, Following Count, Geo Enabled, Follower Rate, Following Rate, Follower Following Ratio, and Verified to produce classification results the good one. The results of the classification using the Naïve Bayes algorithm use 210 datasets of hoax news spreader twitter accounts and use 5 twitter accounts for accuracy calculations and produce an accuracy of 80%.

So that the results obtained by classification using the Naïve Bayes algorithm show that the Naïve Bayes algorithm provides a good level of accuracy, which is 80% in the classification of fake twitter accounts that spread hoax news.

The application of the classification method using the Naïve Bayes algorithm in this study can be increased accuracy if a larger number of datasets are used. It is hoped that this research can be developed using more parameters to increase the accuracy of the classification.

5. REFERENCES

[1] Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. (2011, December). Twitter trending topic classification. In *2011 IEEE 11th International Conference on Data Mining Workshops* (pp. 251-258). IEEE.

[2] Smith, A., & Brenner, J. (2012). Twitter use 2012. *Pew Internet & American Life Project*, *4*.

[3] Boukes, M. (2019). Social network sites and acquiring current affairs knowledge: The impact of Twitter and Facebook usage on learning about the news. *Journal of Information Technology & Politics*, *16*(1), 36-51.

[4] Krishnamurthy, B., Gill, P., & Arlitt, M. (2008, August). A few chirps about twitter. In *Proceedings of the first workshop on Online social networks* (pp. 19-24).

[5] Aphiwongsophon, S., & Chongstitvatana, P. (2018, July). Detecting fake news with machine learning method. In *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 528-531). IEEE.

[6] Lillie, A. E., & Middelboe, E. R. (2019). Fake news detection using stance classification: A survey. *arXiv preprint arXiv:1907.00181*.

[7] Conti, M., Poovendran, R., & Secchiero, M. (2012, August). Fakebook: Detecting fake profiles in on-line social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1071-1078). IEEE.

[8] Sowmya, P., and Chatterjee, M., (2019) Detection of Fake and Cloned Profiles in Online Social Networks. *Proceedings 2019: Conference on Technologies for Future Cities (CTFC),* (pp. 1-5). SSRN

[9] Kontaxis, G., Polakis, I., Ioannidis, S., & Markatos, E. P. (2011, March). Detecting social network profile cloning. In *2011 IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops)* (pp. 295-300). IEEE.

[10] Kontaxis, G., Polakis, I., Ioannidis, S., & Markatos, E. P. (2011, March).

Detecting social network profile cloning. In *2011 IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops)* (pp. 295-300). IEEE.

[11] Alsaedi, N., Burnap, P., & Rana, O. (2017). Can we predict a riot? Disruptive event detection using Twitter. *ACM Transactions on Internet Technology (TOIT), 17*(2), 1-26.

[12] Gurajala, S., White, J. S., Hudson, B., & Matthews, J. N. (2015, July). Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach. In *Proceedings of the 2015 international conference on social media & society* (pp. 1-7).

[13] Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M. (2018). Identifying fake accounts on social networks based on graph analysis and classification algorithms. *Security and Communication Networks, 2018.*

[14] Ren, J., Lee, S. D., Chen, X., Kao, B., Cheng, R., & Cheung, D. (2009, December). Naive bayes classification of uncertain data. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 944-949). IEEE.

[15] Khaled, S., El-Tazi, N., & Mokhtar, H. M. (2018, December). Detecting Fake Accounts on Social Media. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 3672-3681). IEEE.

[16] Helmstetter, S., & Paulheim, H. (2018, August). Weakly supervised learning for fake news detection on Twitter. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 274-277). IEEE.

[17] Freeman, D. M. (2013, November). Using naive bayes to detect spammy names in social networks. In *Proceedings of the 2013 ACM workshop on Artificial intelligence and security* (pp. 3-12).

[18] Xhemali, D., J Hinde, C., & G Stone, R. (2009). Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages. *D. Xhemali, Cj Hinde And Roger G. Stone," Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages", International Journal of Computer Science Issues, IJCSI, Volume 4, Issue 1, pp16-23, September 2009, 4*(1).

[19] Narayanan, V., Arora, I., & Bhatia, A. (2013, October). Fast and accurate sentiment classification using an enhanced Naive Bayes model. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 194-201). Springer, Berlin, Heidelberg.

[20] Frank, E., Hall, M., & Pfahringer, B. (2002, August). Locally weighted naive

bayes. In *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence* (pp. 249-256). Morgan Kaufmann Publishers Inc..

[21] udibyo, U., Astuti, Y., & Kurniawan, A. (2017). High School Major Classification towards University Students Variable of Score Using Naive Bayes Algorithm. *Scientific Journal of Informatics, 4*(2), 191-198.

[22] Trihanto, W. B., Arifudin, R., & Muslim, M. A. (2017). Information Retrieval System for Determining The Title of Journal Trends in Indonesian Language Using TF-IDF and Na? ve Bayes Classifier. *Scientific Journal of Informatics*, *4*(2), 179-190..