International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Hoax News Classification using Machine Learning Algorithms

Sy.Yuliani, Shahrin Sahib, Mohd Faizal Bin Abdollah, Fariska Z. Ruskanda

Abstract*: Hoax news on social media has had a dramatic effect on our society in recent years. The impact of hoax news felt by many people, anxiety, financial loss, and loss of the right name. Therefore we need a detection system that can help reduce hoax news on social media. Hoax news classification is one of the stages in the construction of a hoax news detection system, and this unsupervised learning algorithm becomes a method for creating hoax news datasets, machine learning tools for data processing, and text processing for detecting data. The next will produce a classification of a hoax or not a Hoax based on the text inputted. Hoax news classification in this study uses five algorithms, namely Support Vector Machine, Naïve Bayes, Decision Tree, Logistic Regression, Stochastic Gradient Descent, and Neural Network (MLP). These five algorithms to produce the best algorithm that can use to detect hoax news, with the highest parameters, accuracy, F-measure, Precision, and recall. From the results of testing conducted on five classification algorithms produced shows that the NN-MPL algorithm has an average of 93% for the value of accuracy, F-Measure, and Precision, the highest compared to five other algorithms, but for the highest Recall value generated from the algorithm SVM which is 94%. the results of this experiment show that different effects for different classifiers, and that means that the more hoax data used as training data, the more accurate the system calculates accuracy in more detail.

***Keywords* : Hoax News, Text classification, Machine Learning, Support Vector Machine, Naïve Bayes. Decision Tree, Logistic Regression, Stochastic Gradient Descent, Neural Network –MLP.**

I. INTRODUCTION

Hoax or lies are intentional news made intentionally for emerging as truth, often referred to as a hocus to trick [1], the hoax is currently developing and is at an alarming rate. This situation can cause anxiety and panic in the community. Although hoaxes sometimes do not affect threats, however, new perceptions of spread news can affect a country's social and political conditions. Hoax is deliberate news that is intentionally made to appear as truth, often referred to as hocus to trick [2], Hoax is currently developing and is at an alarming rate. This situation can cause anxiety and panic in the community . Although deception sometimes does not affect threats, however, new perceptions about the spread of news can affect a country's social and political conditions [3]. Hoax news can also result in financial losses, namely by making hoax news on a product, which destroys the company's credibility.

News hoaxes have unique characteristics [4], such as, all in capital letters, underscores, lots of exclamation marks, ask to be distributed, have no realistic date that can be verified, URL addresses are invalid or known. Therefore we can categorize several categories of news hoaxes: a. The possibility of information created without intending to harm, but potentially deceptive. b. Misinformation to frame a problem or an individual. c. Information from sources original, d. Information in the form of new content that is 100% wrong and intentionally designed to deceive and harm, e. information that contains the title, picture, or description does not match the content, f. original information that is related to the context of the misinformation, g. Original information or images and deliberately manipulated to deceive.

A. Category Of Hoax

Hoax are often distributed using sensational titles. They are provocative by utilizing hot news and taking news from official sites [5] and then changing the contents according to the perception desired by the hoax spreader. Following are the Hoax categories :

- Virus Hoax

Messages and emails circulating on e-mails and social media that contain false or misleading information about suspected virus threats

- Giveaway Hoaxes

Messages disseminated via email, social media posts and websites, which contain information that claims that someone gets a gift, voucher or some money.

- Charity Hoaxes

Messages that disseminated via e-mail and social media that falsely claim to be donated are shared and expect that the person receiving the message trust and forward the e-mail.

- Hacker Hoaxes

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
SY.Yuliani, Information Security and Networking Research Group (In FORSNET), Faculty of Information Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia. Email: sy.yuliani@widyatama.ac.id
Shahrin Sahib, Information Security and Networking Research Group (InFORSNET), Faculty of Information Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia. Email: shahrinsahib@utem.edu
Mohd Faizal Bin Abdollah, Information Security and Networking Research Group (InFORSNET), Faculty of Information Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia. Email: faizalabdollah@utem.edu.my
Fariska Z. Ruskanda, School of Electrical Engineering and Informatics, School of Electrical Engineering and Informatics, Indonesia. Email: fariska.zr@informatika.org

A News Hoax that provides a Warning about Fake Security and wrong advice about hackers and other computer security issues.

- Chain Letters

Messages on social media that predict dire consequences for people who 'break the chain' and not share them with others.

B. Hoax News Issue

Social media is issue very vulnerable and is often used as a place to publish hoax news. The number of active users is elementary to spread hoax news. Channels or information containing the highest hoax content came from social media consisting of the highest order of 92.40%, chat applications of 62.80%, and web sites of 34.90%, Mastel 2019, as well as survey results for fake acceptance news behaviour of people. Respondents who argued for checking hoax news were 69.3%, this indicates that "Not Always" the community explores the truth of the information first because community maturity responds to the commotion and inhibition of deception which shows that deception is increasingly vague to be recognized as a hoax [6], hoax of this phenomenon states that technological developments can be a global threat to detection automatically needed [7] to make it easier for people to recognize the news facts, and where the hoax news.

II. RELATED STUDY

Some of the Literature Study that discuss Hoax news, both characteristics, detection and screening techniques are presented below:

Srijan Kumar, discusses the credibility of data from a hoax located on Wikipedia, Kumar discusses how (a) Measures a story and how long they last before being denied, How many page views do they receive, how much referenced by documents in Web. A small number of hoaxes last a long time by web (b) create a characteristic or characteristic of differences in the structure of email articles and content, and editor features hoaxes made, (c) Applying the findings to answer a series of classifications, primarily to determine whether a news is a hoax (d) classification automatically using evaluations that distinguish between Hoax and non-Hoax [8]

Eugenio Tacchini discusses social networking sites (SNS), Tacchini makes hoax detection systems automatically, this study shows that Facebook posts can classify with high accuracy as hoaxes or non-hoaxes based on users who "like" them. The study presents two classification techniques, the first based on logistic regression, the second uses the boolean crowdsourcing technique. In the dataset consisting of 15,500 Facebook posts and 909,236 users, the results of the study resulted in classification accuracy values exceeding 99%. The results of this study indicate that mapping the diffusion pattern of information can be a useful component of the automatic Hoax detection system. [8][9]

Yoke Yie Chen stated that hoaxes made by getting personal data by sending the news to victims. Yoke Chan thinks Hoax is different from spam by the way they disguise through email addresses of people who are related directly or indirectly. Most hoaxes appear as messages that are forwarded and by using a legitimate company name. Then Yoke Yie chan developed a Hoax detection system using the Levenshtein Distance method. The proposed model is used to identify the text-based hoax e-mail. Sensitivity and specificity are used to evaluate the accuracy of the system in identifying e-mail hoaxes [9][10].

Errissya Rasywir conducts hoax news classification in Indonesian using a statistical approach based on machine language, with application based on text categorization, where the proposed system consists of preprocessing, feature extraction, feature selection and execution of classification models. The machine learning algorithm technique chosen in the hoax news classification system is Naïve Bayes (NV), Support Vector Machine (SVM). The results of the conclusions Research results from The best experimental results achieved with naïve Bayes algorithms with unigram features where feature selection uses union operations between information gain and mutual information [11]

A. Classification Technique Detection Of Hoax

Classification Technique Text is a process to find a model or function to describe the class or concept of data. Text classification is used to improve data retrieval using reduction in search time in document locations for summarizing text and reduce comparison time in hoax detection [12]. Here are some algorithms used in the text classification:

- Naive Bayes

Naive Bayes is a simple probabilistic classification algorithm that calculates a set of probabilities by summing frequencies and combinations of values from a given dataset [13]

$$\hat{c} = \underset{c \epsilon C}{\operatorname{argmax}} P(c|d) \quad (1)$$

An email document d consisting of many words x1, x2,…, x..., its posterior probability value is

$$\hat{c} = \underset{c \epsilon C}{\operatorname{argmax}} P(x_1, x_2, \dots, x_n|c)\, P(c) \quad (2)$$

Naïve Bayes classification uses two simplifying assumptions. The first assumption is the container of words, that is, email represents by words from a collection of non-sequential words that do not take a position, only storing the frequency of words in the email. The problem is. Regarding the probability P (xi | c) independently consider class c and can be multiplied complete as follows:

$$P(x_1, x_2, \dots, x_n|c_j) = \prod_i P(x_i|c_j) \quad (3)$$

- Support Vector Machine

Support Vector Machine (SVM), is one of the supervised learning techniques, which is a combination of linear learning machines and kernel functions [14]. There is a training data set; each labelled as a class; this algorithm maximizes the margin between the closest hyperplane and data points by assigning weights to the feature vector. This Makes it a non-probabilistic binary linear classifier. The advantage of this algorithm for hoax detection is its reliability and ability to handle large feature spaces. This algorithm does not try to minimize the error rate, but separates the pattern in high-dimensional space, making it insensitive to class size. as

$$a + b = \gamma \quad (4)$$

- Decision Tree (DT)

Classification using decision trees is a popular method used for classification because it is easily understood and applied well in calculating its accuracy, each branch represents the result from the test, and the leaf node represents class or Distribution class Decision trees are one of the most numerous popular classification methods because it's easy interpreted by many people [15]

$$Entr(S) = - PJ \log 2 Pj \quad (5)$$

Where:
S = Set of cases.
K = Number of partition S.
Pj = Probability obtained from the number (yes/no) divided by the total of the cases.

From the Entropy calculation, the Gain value can be calculated using the following :

$$Gain\ (S,A) = E\ (S) - \sum_i^n = 1 - \frac{Si}{S} X\ E\ (Si) \quad (6)$$

- Logistic Regression (LR)

Logistic regression is an approach to making predictive models of the probability of occurrence of an event by matching data and predicting dichotomous scale dependent variables. The dichotomous scale in question is a nominal data scale with two categories, for example: Yes and No, Good and Bad or High and Low, hoaxes or fact. The logistic regression model learns a weight the probability pi that a post i is non-hoax is then given. Intuitively > 0 indicates that u likes mostly non-hoax (News) posts. [10]

- Stochastic Gradient Descent (SGD)

The Stochastic Gradient Descent (SGD) algorithm is a drastic simplification. Instead of calculating gradient correctly [11], note that when calculating, we need to do the sum for all existing samples because the derivative still contains the sum. If a sample is large, this addiction can take a long time. SGD overcomes this problem by updating guessed values only with derivative values in just one sample, that contains only one sample. In this way, the process of updating guesses can occur faster.

- Neural Network – MLP

Multilayer Perceptron (MLP) is the development of artificial neural networks that emphasize using one or more hidden layers in hidden layers in the network and the use of non-linear connections as a transformation process. This network calls Feedforward because it carried information from the input layer (input layer) to carrie and transformed forward to the output layer (output layer). To reduce problems Single-layer perceptron (SLP) [12] and MLP has several stages, and the first is calculating the value of hidden layers[13]. In Neural Network network topology design, some things must be determined in advance, wrong one of which is determining how many neurons input to be used in the network. Amount Input neurons follow the number of input variables used.

III. METHODS IN HOAX NEWS DETECTION

The Hoax News Detection Process that is proposed consisting of three main components: pre-process text, feature extraction, and classification process. Pre-processing text collects hoax news that is tested to produce hoax or News. The system uses the six different classification method to identify potential hoaxes from contents and compare them to the hoax dataset. The hoax dataset consists of hoaxes collected from various sources such as web hoax-slayer.net, www.sophos.com and www.truthorfiction.com. This News hoax detection uses a classification method. This process is done to classify Hoax (h) and News based on news structure: title, URL, publish date, author, and text. The stages of the Hoax news detection process are carried out through the following stages.

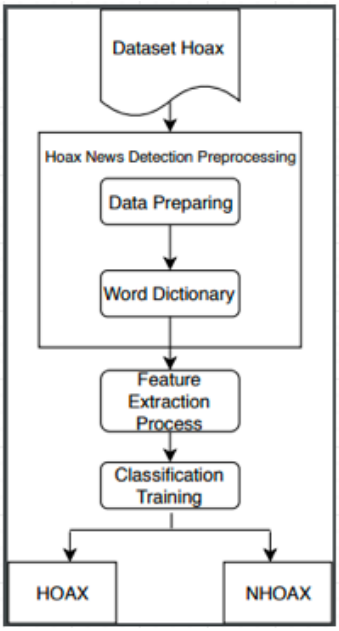

Figure 1: The Stages of The Hoax News Tetection

The data collection used divided into two parts, namely training sets and test sets, each containing 463 News and 260 news. Both of these data sets contain hoaxes and non-hoaxes that are equally distributed. Hoax news label with names containing the phrase "h". and News that given the phrase "nh."

Data set hoax that contains many words do not contribute to information and meaning introduce hoax news, such as punctuation, stop-words, numbers. The steps taken to process the data set used are:

- Remove stop-words

Stop-words that omit are those that are not letters and whose words are only one

- Lemmatization

At this stage, a process of grouping words carried out that has the same root words into one group. For example, the word "include", "includes", "included", is grouped into the word "include".

- Eliminate non-word characters

At this stage, it is necessary to delete punctuation or other characters that do not represent the word.

A. Build a Word Dictionary

The first stage of building a dictionary make a list of words contained in the e-mail and the frequency of their appearance. The next step is to eliminate words that only contain 1 character. The last step is to cut the dictionary into only 3000 words, which often appear. Bles at the top and bottom of columns. Avoid placing them in the middle of the columns. Significant figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence. Examples of Hoax contained in the data set are as follows.

> IT IS OFFICIAL. IT WAS EVEN ON THE NEWS. FACEBOOK WILL START CHARGING. DUE TO BEING A PUBLICLY TRADED ENTITY. IF YOU COPY THIS ON YOUR WALL YOUR ICON WILL TURN GOLD AND FACEBOOK WILL BE FREE FOR YOU FOREVER. PLEASE PASS THIS MESSAGE ON, IF NOT YOUR ACCOUNT WILL BE DISABLED WHEN YOU DO NOT PAY

Figure 2: Example of Hoax

The full function can be seen in the program listing function def make_Dictionary. Example of the contents of the dictionary are as follows.

> ('language', 661) ('university', 407) ('linguistic', 294) ('one', 198) ('edu', 159) ('english', 150) ('discourse', 144) ('conference', 138) ('georgetown', 133) ('interest', 124) ('send', 117) ('word', 116) ('address', 110) ('please', 110) ('form', 105) ('speaker', 104) ('linguist', 104) ('example', 100) ('linguistics', 98) ('information', 97) ('department', 96) ('two', 96) ('analysis', 95) ('list', 95) ('work', 93) ('research', 91) ('between', 90) ('follow', 89) ('include', 89) ('de', 89) ('us', 89) ('number', 87) ('seem', 86) ('student', 84) ('point',

Figure 3: The contents of the dictionary

After the dictionary is ready, the next step is feature extraction, which is a word count vector. Word count vector contains the frequency of occurrence of 1435 hoax news dataset, in each hoax news in the training set. For example, the contents of the training file are "Get the work done, work done." Then the contents of the word count are: [0,0,0,0,0, …… .0,2,0,0,0, ……, 0,0,1,0,0, ... 0, 0,1,0,0,0,0,0,0,0,0,0].

B. Training in Classification

The training process for classification for hoax news filtering hoaxes by using a sci-kit-learn library. The classification method used is Naïve-Bayes Classifier and Support Vector Machines (SVM). Naïve Bayes Classifier is an algorithm that widely used for text classification. This method supervised probabilistic classifier based on the Bayes theorem, which assumes no dependence between each pair of features. Whereas SVM is a supervised binary classifier that is very effective when the number of features is vast. The purpose of SVM is to separate the subset from training data to support vector. The decision function on the SVM model predicts the class from the test data based on the support vector. Classification using decision trees is a method of calculating accuracy, branches represent the results of tests, and leaf nodes represent classes or Distribution classes, and Logistic regression method is an approach to create a probability prediction model where this produce two categories of hoaxes or non-hoaxes, and Stochastic Gradient Descent (SGD) algorithm is a simplification by calculating the gradient of the derivative function of the data, its deficiencies If a large sample of data, can calculate the length, overcome this problem with a comparison of derived estimates in only one sample, the last is the Neural Network - MLP is the development of artificial neural networks that emphasize the use of one or more hidden layers in hidden layers in the network

IV. EVALUATION

In the search for text patterns from hoax news that is the closest to the truth (information retrieval), To evaluate the system performance, the following stages are carried out for testing: precision and recall calculations, precession and recall are two calculations that are widely used to measure the performance of the system or method use [20], To evaluate the system performance, the following stages are carried out for testing.

A. Precision

Precision is the level of accuracy between the information requested by the user and the answer given by the system [21], Precision is used with recall, the percent of all relevant documents that is returned by the search

Calculation of precision can write as follows :

$$\text{Precision} = \frac{tp}{tp + fp} \tag{7}$$

Where:

TP = True Positives the number of relevant data that is correctly classified as matches data by the system.

FP = False Positives the number of irrelevant data, but classified as matches data by the system

B. Recall

Recall is the system's success rate in rediscovering information. In other words, the recall shows how complete the relevant results are displayed by the system. Calculation of recall values can be written in form. Recall is the level of success of the system in rediscovering information.

$$\text{Recall} = \frac{tp}{tp + fn} \tag{8}$$

Where:

TP = True Positives the number of relevant data that is correctly classified as matches data by the system.

FN = False Negatives the number of relevant data, but isn't classified as matches data by the system.

C. F-measure

F-Measure is one of the evaluation calculations in information retrieval that combines recall and precision .where TP, FP, T N, FN represent true positive, false positive, true negative and false negative, respectively. The following is the result of the experiment by using precision-recall and F-measure [21],

The following is the formula from F-Measure:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

D. Accuracy

Accuracy defined as the level of closeness between predictive value and actual value. Accuracy is the level of closeness between predictive value and actual value. Assessment to be verified

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (10)$$

Where:

TP = True Positives the number of relevant data that is correctly classified as matches data by the system.

TN = True Negatives the number of irrelevant data and correctly classified as unmatched data by the system.

FP = False Positives the number of irrelevant data, but classified as matches data by the system.

FN = False Negatives the number of relevant data, but was not classified as matches data by the system.

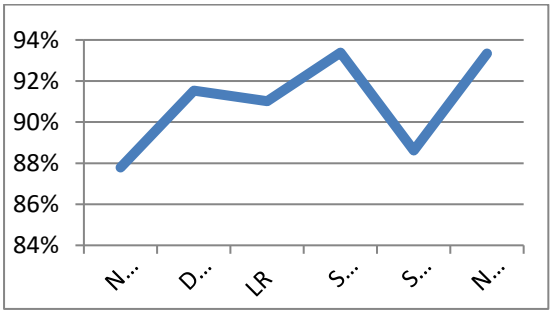

Figure 4: The Accuracy of Results

Experimental result (Figure 4) shows the fact that the NN-MLP classifier has the highest accuracy value of 93%. Whereas the second-best SVM in terms of accuracy yielded 93% each but had a 94% advantage in the Recall category. The results, the precision algorithm is explained by Table 2: Naïve Bayes algorithm result a precision value of 85.42%, Decision Tree algorithm gets 90.56%, and the Logistic Regression algorithm, result 92.94%, then Support Vector Machine algorithm ,result a precision value of 91, 47%, while the Stochastic Gradient Descent algorithm gets 90.70%. The algorithm that has the highest precision value in this study is the Neural Network - MLP algorithm with a precision value of 93.33. Then the algorithm that has the highest precision value in this study is the Neural Network - MLP algorithm with a precision value of 93.33

V. EXPERIMENT RESULTS

Experiments for this classification process is done using Python 3.7 with Anaconda 3 and PyCharm as IDE. Library used: os, numpy, and sklearn. Experiments were carried out on 1435 hoax datasets (Table I). The results obtained are in the form of a confusion matrix. Then the accuracy for both classification training is as shown in Fig. 4.

Table 1: Dataset Hoax News

Hoax Dataset		Number of News	Average word count Hoax
Training set	Hoax	1000	716.9
Test set	NHoax	435	473.9
	Hoax	435	706.8

The experiment (Table 2) shows the fact that in SVM classifiers stemming measures give the best results. While in NN-MLP.

Table 2: Precision Value On Testing Data

Algorithm	Precision
Naïve Bayes	85,42%
Decision Tree	90,56%
Logistic Regression	92,94%
Support Vector Machine	91,47%
Stochastic Gradient Descent	90,70%
Neural Network – MLP	93,33%

From the research that we have done, we obtain different recall values from the algorithm used by Table 3. The Naïve Bayes algorithm gets a precision value of 92.13%, the Decision Tree algorithm gets 90.07%, and the Logistic Regression algorithm gets 88.09%. The Neural Network - MLP algorithm gets a precision value of 93.33%, while the Stochastic Gradient Descent algorithm gets 85.89%. Then the algorithm that has the highest precision value in this study is the Support Vector Machine algorithm with a precision value of 93.36%.

Table 3: Recall Values On Testing Data

Algorithm	Recall
Naïve Bayes	92,13%
Decision Tree	90,07%
Logistic Regression	88,09%
Support Vector Machine	93,36%
Stochastic Gradient Descent	85,89%
Neural Network – MLP	93,33%

Based on the precision and recall values described by Table 4, the Naïve Bayes algorithm obtained F-Measure Score of 88,55%, and Decision Tree of 90,62%, the Logistic Regression algorithm obtained a value of 90,35%. Stochastic Gradient Descent algorithm obtained a value of 87,94 %, then the Neural Network - MLP algorithm with the highest value of 93,33%, and then followed by a Support Vector Machine algorithm of 92,39%. Comparison of the F-Measure Score the five algorithms can be seen in the table 4

Table 4: F-Measure Score On Testing Data

Algorithm	F-Measure
Naïve Bayes	88,55%
Decision Tree	90,62%
Logistic Regression	90,35%
Support Vector Machine	92,39%
Stochastic Gradient Descent	87940%
Neural Network - MLP	93,33%

Based on the results described in Table 5, the Naïve Bayes algorithm received an accuracy value of 88.10%, then the Decision Tree algorithm got a value of 90.61%, and the Logistic Regression algorithm received an accuracy value of 90.63%, followed by a Stochastic Gradient Descent with accuracy the lowest among the six algorithms compared to us, which is 88.40%, the Neural Network algorithm - MLP gets the highest accuracy value of 93.33%, and the last one is the Support Vector Machine algorithm with an accuracy value of 92.31%

Table 5: Accuracy Values On Testing Data

Algorithm	Accuracy
Naïve Bayes	88105%
Decision Tree	90,61%
Logistic Regression	90,63%
Support Vector Machine	92,31%
Stochastic Gradient Descent	88,40%
Neural Network - MLP	93,33%

The results of the text test were carried out on the news hoax dataset which numbered around 1486 and compared the accuracy of the text by using 6 algorithms, naive Bayes, decision tree, logistic regression, support vector machine, and Stochastic Gradient Descent, NN-MPL. Test results obtained, with values of accuracy, precision, recall, f-measure NN-MLP algorithms which have the highest accuracy value compared to the other 5 algorithms where, for precision 93%, recall 94%, f-measure 93 %, accuracy 93%.

Table 6: Experiment Result Comparison Among Precision, Recall, F-Measure Score and Accuracy Values

Algorithm	Precision	Recall	F-Measure	Accuracy
NB	85%	92%	88%	88%
DT	91%	92%	92%	92%
LR	93%	89%	91%	91%
SVM	93%	94%	93%	93%
SGD	91%	87%	88%	89%
NN-MLP	93%	93%	93%	93%

VI. CONCLUSION

In general, hoaxes, identification can be identified with four criteria [22]. First, hoax information usually has the characteristics of chain letters. Second, hoax information usually does not include the date of the event or does not have a real-time or can be verified, statements that do not show clarity. Thirdly, hoax information usually does not have an expiry date on the information alert. Fourth, no identifiable organization is cited as a source of information or includes the organization but is usually not linked to data.

Hoax news detection research is carried out using classification techniques, to hoax news accuracy testing using six algorithms, accuracy is generated from the number of words that come out of hoax news , the algorithm used is, Naive Bayes, Decision Tree, Logistic Regression, Support vector machines, Stochastic Gradient Descent and, Neural Network-MLP. Hoax news detection research conducted by classification techniques with machine learning methods, there are six algorithms that are used to compare the accuracy of the results of hoax news texts, accuracy is generated from the number of words that come out of hoax news, these algorithms include, Naive Bayes, Decision tree, Logistic Regression, Support vector machines, Stochastic Gradient Descent and, Neural Network-MLP. This stage of the method is carried out by pre-preparation for the detection of hoax news. Studying methods using appropriate pre-qualification methods for each classification results in better accuracy, and produces hoaxes and not hoaxes. The purpose of this research is to produce the most frequent words out of hoax news texts that often disseminated. Further improvements made in this study are more significant and more considerable dataset sizes to produce better hoax news detection. The results of experiments that have been carried out from this study produce, the highest value of many news hoaxes that come out are the NN-MPL algorithm and the SVM algorithm. The experimental results concluded that the more hoax news used as training data, the more accurate the system performance detection results.

ACKNOWLEDGEMENTS

This work been supported under Universiti Teknikal Malaysia Melaka research grant Gluar/CSM/2016/FTMKCACT/100013. The Author would like to thanks University Teknikal Malaysia Melaka, CyberSecurity Malaysia and all member of INFORSNET research group for their incredible support in this project.

REFERENCES

1. S. Volkova and J. Y. Jang, “Misleading or Falsification ? Inferring Deceptive Strategies and Types in Online News and Social Media,” pp. 575–583, 2018.
2. Y. Y. Chen, S. Yong, and A. Ishak, “Email Hoax Detection System Using Levenshtein Distance Method,” vol. 9, no. 2, pp. 441–446, 2014.
3. A. B. Prasetijo et al., “Hoax Detection System on Indonesian News Sites Based on Text Classification using SVM and SGD,” pp. 45–49, 2017.
4. D. Harley, :“ : Common Hoaxes and Chain Letters,” vol. 1.
5. S. Y. Yuliani, S. Sahib, M. F. Abdollah, M. N. Al-mhiqani, and A. R. Atmadja, “Review Study of Hoax Email Characteristic,” vol. 7, pp. 778–782, 2018.
6. B. Fitnah et al., “Heboh HOAX Nasional.”
7. Y. Chen and V. L. Rubin, “Towards News Verification : Deception Detection Methods for News Discourse Towards News Verification : Deception Detection Methods for News Discourse,” vol. 2015, 2015.
8. E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some like it Hoax: Automated fake news detection in social networks,” CEUR Workshop Proc., vol. 1960, pp. 1–12, 2017.
9. P. Kumar, H. Kumar, and R. Joseph, “A Framework for Email Clustering and Automatic Answering Method,” Int. J. Adv. Res. Comput. Eng. Technol., vol. 1, no. 9, pp. 52–59, 2012.
10. P. Kumar, H. Kumar, and R. Joseph, “A Framework for Email Clustering and Automatic Answering Method,” Int. J. Adv. Res. Comput. Eng. Technol., vol. 1, no. 9, pp. 52–59, 2012.
11. S. Volkova and J. Y. Jang, “Misleading or Falsification ? Inferring Deceptive Strategies and Types in Online News and Social Media,” pp. 575–583, 2018.
12. Y. Y. Chen, S. Yong, and A. Ishak, “Email Hoax Detection System Using Levenshtein Distance Method,” vol. 9, no. 2, pp. 441–446, 2014.

13. A. B. Prasetijo, R. R. Isnanto, D. Eridani, Y. Alvin, A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax Detection System on Indonesian News Sites Based on Text Classification using SVM and SGD," pp. 45–49, 2017.
14. D. Harley, ": : Common Hoaxes and Chain Letters," vol. 1.
15. S. Y. Yuliani, S. Sahib, M. F. Abdollah, M. N. Al-mhiqani, and A. R. Atmadja, "Review Study of Hoax Email Characteristic," vol. 7, pp. 778–782, 2018.
16. B. Fitnah, J. Ajak, K. Milenial, B. Hoax, M. Pembangunan, H. Bentuk, L. Teror, P. Masyarakat, T. B. Informasi, J. M. Terprovokasi, K. Hitam, P. B. Politik, M. C. Menjadi, and A. M. Hoax, "Heboh HOAX Nasional."
17. Y. Chen and V. L. Rubin, "Towards News Verification : Deception Detection Methods for News Discourse Towards News Verification : Deception Detection Methods for News Discourse," vol. 2015, 2015.
18. S. Kumar, R. West, and J. Leskovec, "Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes," Www, pp. 591–602, 2016.
19. E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some Like it Hoax: Automated Fake News Detection in Social Networks," 2017.
20. P. Kumar, H. Kumar, and R. Joseph, "A Framework for Email Clustering and Automatic Answering Method," Int. J. Adv. Res. Comput. Eng. Technol., vol. 1, no. 9, pp. 52–59, 2012.
21. E. Rasywir, A. Purwarianti, and K. Kunci, "Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin," J. Cybermatika, vol. 3, no. 2, 2015.
22. V. Mitra, C. Wang, and S. Banerjee, "Text classification : A least square support vector machine approach," vol. 7, pp. 908–914, 2007.
23. A. Bhowmick and S. M. Hazarika, "Machine Learning for E-mail Spam Filtering: Review,Techniques and Trends," 2016.
24. A. Fauzi, E. B. Setiawan, and Z. K. A. Baizal, "Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method," Proceeding 2nd Int. Conf. Data Inf. Sci., pp. 1–6, 2018.
25. N. Elavarasan and K. Mani, "A Survey on Feature Extraction Techniques," pp. 52–55, 2015.
26. G. Da, S. Martino, M. Mohtarami, P. Nakov, and J. Glass, "Team QCRI-MIT at SemEval-2019 Task 4: Propaganda Analysis Meets Hyperpartisan News Detection ˜ 3 ," 2019.
27. N. E. C. L. America and P. Nj, "Large-Scale Machine Learning with Stochastic Gradient Descent," 2010.
28. A. A. Heidari, H. Faris, S. Mirjalili, I. Aljarah, and M. Mafarja, Ant Lion Optimizer : Theory , Literature Review , and Application in Multi-layer Perceptron Neural Networks. Springer International Publishing, 2020.
29. S. Ghosh, S. Biswas, and D. Sarkar, "ORIGINAL ARTICLE A novel Neuro-fuzzy classification technique for data mining," Egypt. Informatics J., vol. 15, no. 3, pp. 129–147, 2014.
30. K. Shu, S. Wang, and H. Liu, "Exploiting Tri-Relationship for Fake News Detection," 2016.
31. I. Retrieval, Introduction to Information Retrieval. 2008.
32. S. Y. Yuliani, M. F. Bin Abdollah, S. Sahib, and Y. S. Wijaya, "A framework for hoax news detection and analyzer used rule-based methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 10, pp. 402–408, 2019.

AUTHORS PROFILE

S. Y. Yuliani is Currently a Ph.D. student at Faculty of Information and Communication Technology, Universiti Teknik Malaysia Melaka (UTeM) and a member of Information Security, Digital Forensic and Computer Networking (INSFORNET) research group. Mainly his research interests include of Computer Security, Cyber Security, Information Systems Security Audit, Machine Learning, Text Processing, Hoax Detection. She is currently a senior lecturer at the Faculty of Engineering, Widyatama University, Indonesia. (sy.yuliani@widyatama.ac.id)

Shahrin Shahib is currently an professor at the University Teknikal Malaysia Melaka, Malaysia. He research interests include Networking, Computer Systems, Security, Network Administration and Design. Currently, he also contributes to Cybersecurity Malaysia as a Member of Malaysia Common Criteria (MyCC) Scheme Management Board. He also become a member of various conference committee and journal reviewer. (shahrinsahib@utem.edu.my)

Mohd Faizal Abdollah is currently an associate professor at the University Teknikal Malaysia Melaka, 2009. His research interests include malware analysis, network security, cyber terrorism, Intrusion Detection system and network management. Currently, he is active working in the research of malware at Nework mobile and IoT platform and cyber attack. He also become a member of various conference committee and journal reviewer. (faizalabdollah@utem.edu.my)

Fariska Zakhralativa Ruskanda currently is a PhD student in School of Electrical Engineering and Informatics in Institut Teknologi Bandung (ITB), Indonesia. Her research interests include artificial intelligence, computer graphics, digital image processing, bioinformatics, machine learning, and natural language research of Natural Language Processing especially in text processing, text classification, hoax detection, image classification, sentiment analysis (aspect-based), and Arabic text proccesing. She received a Bachelor in Informatics and also Master in Informatics from Institut Teknologi Bandung (ITB). She is currently a junior lecturer at School of Electrical Engineering and Informatics, Institut Teknologi Bandung.(fariska.zr@informatika.org)