Indonesian Journal of Electrical Engineering and Computer Science
Vol. 14, No. 3, June 2019, pp. 1508~1517
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v14.i3.pp1508-1517

Youtube spam detection framework using naïve bayes and logistic regression

Nur'Ain Maulat Samsudin[1], Cik Feresa binti Mohd Foozy[2], Nabilah Alias[3], Palaniappan Shamala[4], Nur Fadzilah Othman[5], Wan Isni Sofiah Wan Din[6]
[1,2,3,4]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia
[5]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[6]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang (UMP), Malaysia

Article Info

Article history:

Received Dec 12, 2018
Revised Feb 13, 2019
Accepted Feb 27, 2019

Keywords:

Classification
Detection
Machine learning
Spam

ABSTRACT

YouTube has become a popular social media among the users. Due to YouTube popularity, it became a platform for spammer to distribute spam through the comments on YouTube. This has become a concern because spam can lead to phishing attack which the target can be any user that click any malicious link. Spam has its own features that can be analyzed and detected by classification. Hence, enhancement features are proposed to detect YouTube spam. In order to conduct the experiments, a YouTube Spam detection framework that consists of five (5) phases such as data collection, pre-processing, features selection and extraction, classification and detection were developed. This paper, proposed the YouTube detection framework, examined and validate each of the phases by using two types of data mining tool. The features are constructed from analysis by using data collected from YouTube Spam dataset by using Naïve Bayes and Logistic Regression and tested in two different data mining tools which is Weka and Rapid Miner. From the analysis, thirteen (13) features that had been tested on Weka and RapidMiner shows high accuracy, hence is being used throughout the experiment in this research. Result of Naïve Bayes and Logistic Regression run in Weka is slightly higher than RapidMiner. In addition, result of Naïve Bayes is higher than Logistic Regression with 87.21% and 85.29% respectively in Weka. While in RapidMiner there is slightly different of accuracy between Naïve Bayes and Logistic Regression 80.41% and 80.88%. But, precision of Naïve Bayes is higher than Logistic Regression.

Corresponding Author:

Cik Feresa Mohd Foozy,
Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia (UTHM),
Parit Raja, Batu Pahat, 86400 Johor, Malaysia.
Email: feresa@uthm.edu.my

1. INTRODUCTION

YouTube is one of the famous and well-known social media. YouTube is functioning as for the user to upload or share any relevant videos. Any Internet user from all over the world can watch the video online. From the video in YouTube, users not only can share their videos, but also can comment on the videos. Comments that came from the users sometimes not only to praise the good video or criticize videos they dislike but also post an unwanted or unsolicited and unrelated electronic message that is sent in bulk to a group of recipient which also known as spam [1].

Spam causes many problems, including wasting the user's time, memory and use up network bandwidths. Organizations and users could face financial loss due to the threat of spam [2]. Some of the

spammers use the comment part on YouTube for advertising issues, while others are responsible for distributing computer viruses and there are some spam messages intended to steal the user financial identities [3]. The most concerned threats of spam are when involving malicious spam that will lead to phishing websites once the users click the link [3] and the distribution of malware [4]. Gandra [4] states that 100 to 1 ratio of spam is on YouTube as shown in Table 1. This shows the seriousness of spam attack on YouTube.

Table 1. Report on Popular Social Platform for Spam [4]

Description	Data
Social media apps that are spammy	5%
Spammy social media apps that are brand-owned	20%(that is 1% overall)
Average number of social profiles contacted by a spamming account	23
Number of new spam accounts created	5 out of every new accounts
Most Popular social platform for spammers	**Facebook & You Tube**
Percentage of spam that contain URL	15%
Overall number of social media messages	1 out of every 200

2. LITERATURE REVIEW

Spam attack had been word widely distributed. Not only in social media such as Facebook, Twitter, YouTube, blogs, but also in SMS and e-mails. For email spam is defined as unwanted emails sent by different users daily by [5]. According to Tran *et al.* [6], email spam brings the meaning of unsolicited bulk emails received by users. While unsolicited commercial email or junk email is the definition of spam defined by Stone [7]. For Short Message Service (SMS) spam is known as unsolicited or unwanted message received on a mobile phone [8]. In web spamming, spam brought the meaning of an intended activity to mislead search engine to ranking some page higher than they reserved [9], [10]. Spam comments had been identified as a comments which consists of commercial content that is unrelated to the discussion with unwanted content or requests [1]. In addition, video spam also been defined by Yusof and Sadoon [1] as unrelated, unwanted content compared to its video's title.

2.1. Existing YouTube Spam Detection Framework

For YouTube spam detection framework used by Yusof and Sadoon [1], Alberto *et al.*, [11], Chowdury *et al.*, [13], and Kiran [12] is discussed. In a study conducted by Yusof and Sadoon [1], a framework used consists of five (5) phases which is data collection, pre-processing, feature construction, spam detection and evaluation. The framework used by Alberto *et al.*, [11] has three (3) phases such as processed data, pre-processing and classification. While framework in research conducted by Kiran [12] has three (3) phases consists of data collection, feature selection and classification. Besides that, framework used in research by Chowdury *et al.*, [13] consists of data collection, select attribute and classifications. Table 2 shows the comparison of YouTube spam detection frameworks.

Table 2. Comparison of YouTube Spam Framework

Author	Title	Framework				
		Data	Pre-processing	Features	Classification	Evaluation
[1]	Detecting Video Spammers in YouTube Social Media	✓	✓	✓	✓	✓
[11]	TubeSpam: Comment Spam Filtering on YouTube	✓	✓		✓	
[12]	Detecting spammers in YouTube: A study to find spam content in a video platform.	✓		✓	✓	
[13]	A Data Mining Based Spam Detection System for YouTube	✓		✓	✓	

From the Table 2, it shows most of the researchers applied data collection, feature selection and classification in their YouTube Spam detection framework. Therefore, for the framework used in this research resembled from the framework used by Yusof and Sadoon [1], there are data collection, pre-processing, features selection, classification and detection.

2.2. Data Collection

There are two (2) types of data being collected by existing researcher. The two types of dataset are YouTube comments and YouTube video. Alberto *et al.*, [11] used YouTube comments dataset downloaded

from UCI Machine Learning Repository [28]. The dataset contain a total of 1005 of spam comments and 951 of ham comments. On the other hand, Kiran [12] used YouTube Video data are gained from Crawling Algorithm with a total of 473 spam video and 119 ham video. Besides that, another YouTube video dataset is used by Yusof and Sadoon [1] which is extracted from web pages. The dataset consist of a total of 30621 spam and ham videos. Meanwhile, Chowdury et al., [13] obtain the dataset from TubeKit with 685 of spam videos and 1115 of ham videos.

Table 3 shows the type of dataset and total number of datasets had been used in the existing research. Due to the availability of YouTube comments is from UCI Machine Learning Repository [28] and had been used by Alberto *et al.,* [11], the same YouTube spam comments dataset is used in this research.

Table 3. Dataset Collection

Author	Dataset Type	Total Number of Dataset
[11]	YouTube Comment	1005 Spam, 951 Ham
[12]	YouTube Video	473 Spam, 119 Ham
[1]	YouTube Video	30621 Spam and Ham
[13]	YouTube Video	685 Spam, 1115 Ham

2.3. Feature Selection

There are many types of features can be selected to be used as a parameter in a research. For video spam in YouTube study conducted by Yusof and Sadoon [1], Alberto *et al.,* [11], and Kiran [12] different type of features is being selected. Study by Yusof and Sadoon [1], the feature being selected is the Edge Rank Algorithm. This algorithm is implemented because the algorithm is the same algorithm being implemented by Facebook in detecting spam. Next research conducted by Kiran [12] used three (3) features which are video based, user based and social network features in order to identify spam users. Features used by researcher Chowdury *et al.,* [13] is the number of users, number of comments, number of distinct users, number of rating counts, and number of different categories.

On the other hand, features used by Alberto *et al.,* [11] in detecting spam in YouTube spam comments is the most occurrence spam keywords. YouTube spam comments are still a new research area, therefore features of YouTube spam are rarely identified. So, feature of comments from web blog, Twitter, and SMS is being studied and can be implemented. Features extracted from comments in the research conducted by AlSaleh and AlArifi [14] such as post-comments similarity, the interval between post and comments, number of words in the comments, a number of sentences in the comments, comment length, phone information, email information, Uniform Resource Locator (URL) link, black word list, stop word ratio and word duplication ratio. Next is the features used by Uysal *et al.,* [15], is includes message length, number of terms, uppercase character ratio, non-alphanumeric character ratio, alphanumeric character ratio and the presence of URL in the comments. Perveen [16] conducted a study with the features of negative word count, negative word counts, URL, positive word count, positive word ratio.

Table 4 shows most researchers used heuristic, keyword and Uniform Resource Locator (URL) link. Three of the researchers used Heuristic and the presence of URL links as features while only Alberto *et al*., [11] used keyword features in the research. Therefore, the combination of heuristic, keyword and presence of URL links is chosen as features selection and extracted from the datasets.

Table 4. Comparison of Comments Features

Author	Title	Features		
		Heuristic	Keyword	URL Links
[9]	TubeSpam: Comment Spam Filtering on YouTube		✓	
[14]	Combating Comment Spam with Machine Learning Approaches	✓		✓
[15]	The Impact of Feature Extraction and Selection on SMS Spam Filtering	✓		✓
[16]	Sentiment Based Twitter Spam Detection	✓		✓

2.4. Classification

Classification is a crucial process in detection research. In this phase, classifier is being chosen to run features selected by researchers to identify the result. In the research conducted by Yusof and Sadoon [1], Chowdury *et al.,* [13], Alberto *et al.,* [11], and Kiran [12], classification techniques is used. Yusof and Sadoon [1] uses nine (9) classifier in the research. All of the classifier is being classed into three (3) classes

which is Decision Tree (DT) that consists of Functional Tree (FT), J48, Random Forest (RF). Class Function has LibLINEAR (LL), LibSVM (LSVM), Logistic (LR), Multilayer Perceptron (MLP) and Class Bayesian which consist of Bayes Network (BN), Naïve Bayes (NB). A study conducted by Alberto *et al.,* [11] were using six (6) classifier to find which classifier give better performance in detecting YouTube spam comments. Classifier user consists of K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), and Logistic Regression (LR). In order to detect malicious user in YouTube, Kiran [12] used Support Vector Machine (SVM) as a classifier to detect the malicious user. Other than that, Naïve Bayes, Clustering and Decision Tree are the classier used by Chowdury *et al.,* [13] in their research.

From the comparison Table 5, it shows that three out of four researchers chose Naïve Bayes in detecting spam. Then, two of the researchers choose Logistic Regression, Random Forest, and SVM techniques.

Table 5. Comparison of Classifier Technique

Classifier / Author	[7]	[9]	[10]	[11]
FT	✓			
J48	✓			
RF	✓	✓		
LL	✓			
LSVM	✓			
LR	✓	✓		
MP	✓			
BN	✓			
NB	✓	✓		✓
KNN		✓		
DT		✓		✓
SVM		✓	✓	
C				✓

2.5. Detection

This section is the result of detection performance of the classifier that had been experimented by Yusof and Sadoon [1], Alberto et al., [11], Kiran [12], and Chowdury *et al.,* [13]. For Alberto *et al.,* [11] out of every classifier had been used the most reliable accuracy and better ranking position is DT, Naïve Bayes, SVM, Random Forest and Logistic Regression with 99% of confidence level. While K-Nearest Neighbor performance is the worst.

Apart from that, Kiran [12] who is using SVM as a classifier in detecting spam user in YouTube measure by using True Positive rate, True Negative rate, False Positive rate, False Negative rate, Accuracy and F-Measure. The result obtained is True Positive Rate is 46.9%, which indicates that SVM detect 46.9% spammers. For legitimate user, 99.1% is identified. Besides that, the accuracy result is 90%.

Yusof and Sadoon [1] using split percentage and cross validation in representing the accuracy result. The split percentage is by 70:30, 80:20, and 90:10.The result obtained from the research shows that Naïve Bayes computes the highest accuracy in all three split percentages with 98% in average. While Multilayer Perceptron techniques show lowest accuracy with 90.67% in average.

Result from the Chowdury *et al.,* [13] study is for 40% of the population, result of Naïve Bayes is 99.75%, Decision Tree 98.66% and clustering with 98.98%. But, at 85% of total population Naïve Bayes has predicted accuracy of 80.20%, decision tree has 82.11% and clustering has 65.79%. Hence, the conclusion from the result is stated that Naïve Bayes and Decision Tree give better performance when the test case is higher in number. The summary performance is shown in the Table 6.

Table 6. Performance Result Summary

Author	Title	Accuracy Result
[1]	Detecting Video Spammers in YouTube Social Media	NB – 98%, LR – 95.67 %
[11]	TubeSpam: Comment Spam Filtering on YouTube	DT, Naïve Bayes, SVM, Random Forest and Logistic Regression = 99% of confidence level
[12]	Detecting spammers in YouTube: A study to find spam content in a video platform.	SVM – 90%
[13]	A Data Mining Based Spam Detection System for YouTube	NB- 80.20% DT – 82.11% C – 65.79

Table 6 shows the result of each classifier used in the existing research. Based on the results obtained, it shows that Naïve Bayes produce high accuracy in detecting spam followed by Logistic Regression Techniques. Therefore, both Naïve Bayes and Logistic Regression were chosen to validate the techniques by using features selected in this research.

3. RESEARCH METHODOLOGY

This section will be explained flow for this research that will include dataset used and framework proposed for this research so that the experiments can be organized carried out. Figure 1 illustrates an overview of this research methodology. This research methodology consists of five (5) phases that need to be followed. The phases starting with data collection, pre-processing, feature selection and extraction, classification and ended with detection.

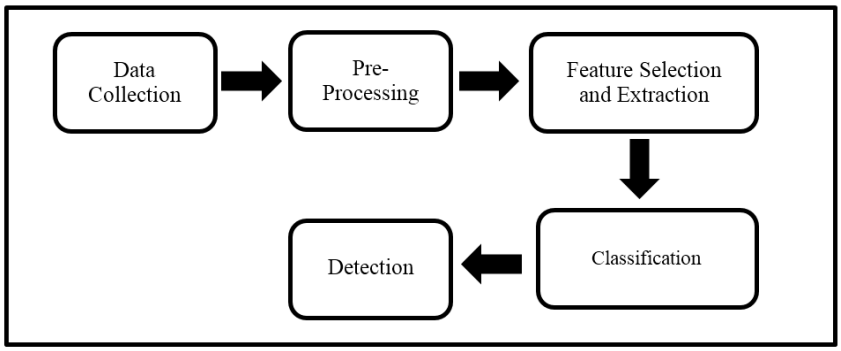

Figure 1. Overview of research methodology

3.1. Datasets Collection

The datasets that will be used in this paper will be YouTube comment from YouTube Spam Collection Data Set from Machine Learning Repository [26]. In the datasets, there are five (5) YouTube videos comments had been collected with the total of 1956 comments. Out of 1005 of the comments is spam while the rest are ham (legitimate) comments.

Table 7 shows the datasets. This dataset is also used by Alberto *et al.*, [11] while conducting their research. Apart from that, these datasets is a publicly available dataset on the Internet.

Table 7. Number of Spam and Ham in the Dataset [26]

Name	Number of Spam	Number of Ham
Psy	175	175
Katy Perry	175	175
Eminem	245	203
LMFAO	236	202
Shakira	174	196
Total	1005	951

3.2. Pre-Processing

Pre-processing is one of the important steps in machine learning techniques. This process will clean the dataset in order to identify the suitable features for this detection framework. The flow chart at Figure 2, shows the process how the pre-processing experiments were done by using machine learning tool.

There are two processes taken in pre-processing steps for this research. The process is tokenization and stemming. Figure 2 illustrates the pre-processing phase.

1) Tokenization is the process of splitting the comments by a space (-), and punctuation symbols (!,?) and used by Patwari [17], Gomes *et al.*, [18], Verma [19], Yang and Qian [20]. Tokenization made the comments is read word by word. It eases the next stemming process.
2) Stemming process is the process of changing the word into its root word example “Subscribe” will be “Subscrib” by Alberto *et al.*, [11], Gomes *et al.*, [18], Verma [19] and Yang and Qian [20].

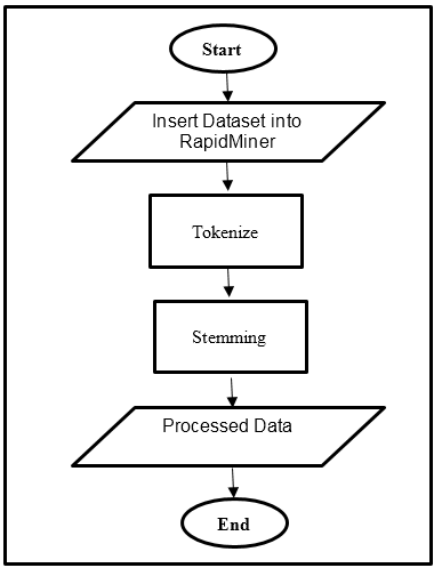

Figure 2. Process of pre-processing

3.3. Features Selection and Extraction

Features Selection and Extraction is the third phase for this detection framework. The process in this phase are shown as below in Figure 3. There are Data that has been cleaned, identify the suitable features based on the YouTube comment, split the features into three set in order to identify the best features set and finally is test the features by using classification techniques such as Naïve Bayes, Logistic Regression, Naïve Bayes and Logistic Regression.

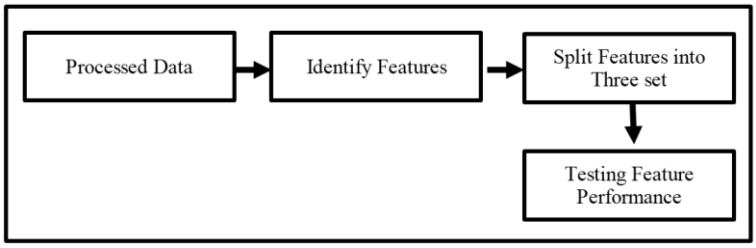

Figure 3. Features selection and extraction

For this research, three (3) experiments had been conducted to identify most suitable features to be used in this research. Therefore, the selected features are as follows:

a. Presence of Links
   Research on detecting spam by AlSaleh and AlArifi [14], Perveen [16], Uysal *et al.*, [15] also uses the presence of links as one of the features. The presence of links or URL is commonly detected as a spam message or comments. This feature we noted in boolean expression where a value of 1 as presence and value of 0 as an absence.
b. Length of Comments
   The length of comments on this research is calculated after pre-processing. The value of this feature is numerical. This features also being used in the research by AlSaleh and AlArifi [14], Hijawi *et al.*, [21], and Cristina [22].
c. Spam Keyword
   This feature also denoted as Boolean expression where if there are the spam keyword in the comment, the value 1 will be denoted while value 0 if there are no spam keyword. These features had been used by Alberto *et al.*, [11] and Mercer *et al.*, [23].

   Table 8 shows the spam keyword being used in this research.

Table 8. Spam Keyword Used in this Research

Spam Keyword
Click
Visit
Subscrib
Spam
Money
Check
Pleas
Com
www
http
Channel

3.4. Classification

A test is conducted between Naïve Bayes and Logistic Regression with KNN and SVM. The result shows that Naïve Bayes and Logistic Regression produce higher accuracy than KNN and SVM. Therefore, Naïve Bayes and Logistic Regression were chosen as classifier in this research. Figure 4, shows the process of classification techniques used in the experiments.

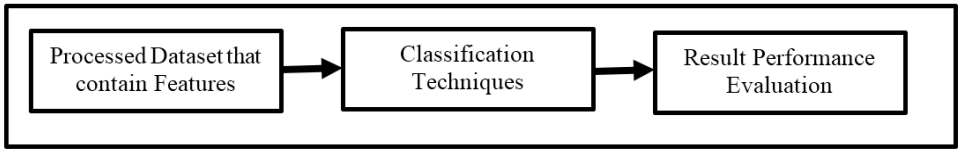

Figure 4. Classification

3.4.1 Naïve Bayes

Naive Bayes classifiers are linear classifiers that are known as a simple but very efficient classifier. The probabilistic model of Naive Bayes classifiers is based on Bayes' theorem. The 'naïve' actually comes from the assumption that the features in a dataset are mutually independent respectively [21]. This is proven by the use of Naïve Bayes by Yusof and Sadoon [1], Stone [7], Chowdury *et al.*, [13], Raschka [24].

3.4.2 Logistic Regression

Logistic regression classifier is very popular and widely used classification technique [25]. This is simple, easy to implement, and provide good performance on a wide variety of problems such as predicting spam. Logistic Regression also best in predicting discrete probability where the output of the probability either yes or no or win or lose. Logistic Regression is simple to execute, and give great execution on a wide assortment of issues [26].

3.4.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) classifier is considered as an example-based classifier, means the training documents are used for comparison, rather than an explicit category representation, such as the category profiles used by other classifier. As such, there is no real training phase. When a new document needs to be categorized, the *k* most similar documents also known as neighbors are found and if a large enough proportion of the neighbor have been assigned to a certain category, the new document is also assigned to this category, otherwise not. In order to decide whether a message or comments is a spam or ham, it is referred to the class of the messages that are closest to it. The comparison between the vectors is a real time process [27].

3.4.4 Support Vector Machine (SVM)

A support vector machine (SVM) is defined as a set of related supervised learning methods that is used for classification. In a simple word, if it is given a set of training examples, each marked as belonging to one of two categories. The SVM training algorithm builds a model that predicts whether a new example falls into which specific category. SVM constructs a hyper plane or a set of hyper planes in a high dimensional space, which can be used for classification [25].

4. RESULT AND DISCUSSIONS

This section will be discussing on results obtain from the experiments that had been conducted.

4.1. Features Selection Experiments

There are three (3) experiments had been conducted to find the most suitable features to be used in this research. The features are being tested with Naïve Bayes and Logistic Regression in two (2) data mining tools of Weka and RapidMiner.

Table 9. Result of Features Selection Set

Features	Weka		RapidMiner	
Set	Naïve Bayes	Logistic Regression	Naïve Bayes	Logistic Regression
7	74.26%	74.26%	72.85%	72.75%
8	75.54%	75.42%	71.24%	72.70%
13	87.21%	85.42%	80.41%	80.88%

From Table 9, the first experiment which was using seven (7) features (presence of links, keyword such as bad word, click, visit, subscrib, spam and money) the result is not satisfied which was below than 80% of Accuracy in both data mining tools.

The Feature Set of 8 experiments is a heuristic feature is added which is the comment length. Based on the Table 9, the result started to be increasing but it still below 80% of Accuracy. Then, the features set of 13 experiment, are used another keyword such as check, pleas, com, http, channel and www.

Based on the experimental above, it shows that the more features used in this detection framework, the accuracy result is increase which give more than 80% result. To validate the result, two machine learning tool such as Weka and RapidMiner are been used in this experiments and it is proven the result is in similar range which is above 80% accuracy.

4.2. Classifier Experiments

Classification is a process to classify the YouTube comment into Ham or Spam message. The classification techniques used in this experiments are Naïve Bayes and Logistic Regression are being tested with KNN and SVM classifier. The set features used in this experiments is 13. Two machine learning of Weka and RapidMiner has been used to test the result. According to the result, it is shows that Naïve Bayes and Logistic Regression classifier perform better in detecting YouTube spam comments. Table 10 shows the results.

Table 10**.** Classifier Result

Techniques	Weka	RapidMiner
Naïve Bayes	87.21 %	80.41 %
Logistic Regression	85.42 %	80.88%
KNN	84.53 %	66.41 %
SVM	85.04 %	73.68 %

From Table 10, it shows that Naïve Bayes and Logistic Regression computes higher accuracy in detecting YouTube spam comments. Therefore, Naïve Bayes and Logistic Regression classifier will be chosen to test YouTube spam comments in term of Accuracy and Precision.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Where

i. TP is True Positive
ii. TN is True Negative
iii. FP is False Positive
iv. FN is False Negative

4.3. Detection Experiments

In this experiment, the performance of two (2) classifiers was compared by using two (2) different data mining tools. In order to make the comparison fair, this research uses the same YouTube Spam comments dataset to extract the features. Two (2) evaluation measures such Accuracy and Precession are used to evaluate the developed spam detection models. The evaluation is computed by the two data mining tools (Weka, and RapidMiner). There are 13 features set has been used in this experiments. Naïve Bayes and Logistic regression shows acceptable result in classify the YouTube comment into Ham and Spam class. Table 11 summarizes the results.

Table 11. Results in Weka and RapidMiner for Naïve Bayes and Logistic Regression

Techniques	Weka		RapidMiner	
	Accuracy (%)	Precision (%)	Accuracy (%)	Precision (%)
Naïve Bayes	87.21	87.2	80.41	75.27
Logistic Regression	85.42	85.7	80.88	74.13

From the Table 11, the Naïve Bayes and Logistic Regression result, it shows Naïve Bayes performed better detecting YouTube spam comments in Weka than Logistic Regression. While in RapidMiner, there is a slightly different of Accuracy and Precision result. The result of Logistic Regression computes slightly higher accuracy than Naïve Bayes. But, the precision of Naïve Bayes is higher than Logistic Regression. Despite the differences of results computed by both data mining tools, it is validated that both of the classifier performed good performance due to the accuracy that both had achieved more than 80% in both data mining tools after being tested.

5. CONCLUSION

In this research, the development of a spam comment detection framework by using machine learning techniques has been done. It is important to improve security since the Internet nowadays that indication the security issues [29]. There are many studies aimed to reduce attack and to protect user privacy but yet lacking in applying the techniques for social media [30]. This paper also wants to contribute by examining the suitable features based on the real comment from social media site for developing spam comment detection framework.

There are several phases involved in the development of this framework such as Data Collection, Pre-processing, Features Selection and Extraction, Classification and Detection. Each of these phases has been validated through experiments by using machine learning techniques.

The Data Collection is downloaded from UCI Machine Learning and the Pre-processing will clean the dataset before the experiments are performed. Based on the result, it shows that the feature selection contribute a good, accurate result and some classification techniques is not suitable with the features set. However it is important to develop a framework in detecting spam comment in order to develop a good Spam detection tool in future.

There are three (3) types of features which are presence of links, length of comments and most occurrence spam keyword that appeared in the processed datasets. The feature then is being tested with Naïve Bayes and Logistic Regression. Both accuracy and precision result were computed by using two data mining tools which are Weka and RapidMiner for data validation. The experimental result showed Naïve Bayes and Logistic Regression are good classifiers in detecting YouTube spam comments. For future works a spam detection tool may be added and tested with other classifier because it is significant to have a tool to detect spam comment in order to avoid the user to click the malicious link.

ACKNOWLEDGEMENTS

The authors express appreciation to the University Tun Hussein Onn Malaysia (UTHM). This research is supported by GPPS Grant vot number H061 and Tier 1 Grant vot number H237.

REFERENCES

[1] Y. Yusof and O. H. Sadoon, "Detecting Video Spammers In Youtube Social Media," no. 082, pp. 228–234, 2017.
[2] U. K. Sah and N. Parmar, "An approach for Malicious Spam Detection In Email with comparison of different classifiers," *IRJET*, vol 4, i.8, pp. 2238–2242, 2017.

[3] I. Daugher and R. Antoun, "Ham- Spam Filtering Using Different PCA Scenarios," *IEEE Int. Conf. Comput. Sci. Eng. IEEE Int. Conf. Embed. Ubiquitous Comput. Int. Symp. Distrib. Comput. Appl. to Business*, Eng. Sci., pp. 542–545, 2016.
[4] S. Gandra, "Implementation Of Prototype To Detect Spam In YouTube Using The Application TubeKit And Naïve Bayes Algorithm," 2014.
[5] M. Esmaeili, et al., "An Anti-Spam System using Naive Bayes Method and Feature Selection Methods," *International Journal of Computer Applications*, vol. 165, no. 4, pp. 1–5, 2017.
[6] K. Tran et al., "Towards a Feature Rich Model for Predicting Spam Emails containing Malicious Attachments and URLs," in *Proceedings of the 11-th Australasian Data Mining Conference*, 2013, pp. 161–171.
[7] T. Stone, "Parameterization of Na ¨ ıve Bayes for Spam Filtering," 2003.
[8] M. Shafie et al., "A Review on Mobile SMS Spam Filtering Techniques," vol. 5, 2017.
[9] H. Garcia-molina, "Web Spam Taxonomy," pp. 1–9.
[10] J. Zhang and G. Gu, "Neighbor Watcher : A Content-Agnostic Comment Spam Inference System," no. 2.
[11] T. C. Alberto, J. V. Lochter, and T. A. Almeida, "TubeSpam: Comment spam filtering on YouTube," Proc. - 2015 *IEEE 14th Int. Conf. Mach. Learn. Appl. ICMLA 2015*, no. 2012, pp. 138–143, 2016.
[12] P. S. Kiran, "Detecting spammers in YouTube : A study to find spam content in a video platform.," *IOSR Journal of Engineering (IOSRJEN)*, vol. 05, no. 07, pp. 26–30, 2015.
[13] R. Chowdury, N. M. Adnan, G. A. N. Mahmud, and R. M. Rahman, "A Data Mining Based Spam Detection System for YouTube," pp. 373–378, 2013.
[14] M. Alsaleh and A. Alarifi, "Combating Comment Spam with Machine Learning Approaches," 2015.
[15] A. K. Uysal, S. Gunal, S. Ergin, and E. S. Gunal, "The Impact of Feature Extraction and Selection on SMS Spam Filtering," *Elektronika Ir Elektrotechnika*, pp. 67–72, 2013.
[16] N. Perveen, "Sentiment Based Twitter Spam Detection," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 7, no. 7, pp. 568–573, 2016.
[17] A. Patwari, "Identifying Undesireble Behaviour in Social Media : Towards Automated Fact-Checking and YouTube Meta-Data Spam Detection," *Purdue University*, 2017.
[18] S. R. Gomes, S. G. Saroar, M. A. Telot, B. N. Khan, A. Chakrabarty, and M. Mostakim, "A Comparative Approach to Email Classification Using Naive Bayes Classifier and Hidden Markov Model," in *Proceedings of the 2017 4th International Conference on Advances in Electrical Engineering (ICAEE)*, 2017, pp. 28–30.
[19] T. Verma, "Tokenization and Filtering Process in RapidMiner," *International Journal of Applieed Information Systems*, vol. 7, no. 2, pp. 16–18, 2014.
[20] T. Yang and K. Qian, "Spam Filtering using Association Rules and NaIve Bayes Classifier," pp. 638–642, 2015.
[21] W. Hijawi, H. Faris, J. Alqatawna, A. M. Al-zoubi, and I. Aljarah, "Improving Email Spam Detection Using Content Based Feature Engineering Approach," 2016.
[22] R. Cristina, "Identification of Spam Comments using Natural Language Processing Techniques," pp. 29–35, 2014.
[23] R. E. Mercer, R. Shams, and R. E. Mercer, "Classifying Spam Emails Using Text and Readability Features Classifying Spam Emails using Text and Readability Features," no. December, 2013.
[24] S. Raschka, "Introduction and Theory," pp. 1–20, 2014.
[25] J. Badresiya, Ashok; Vohra, Saifee; Teraiya, "Performance Analysis of Supervised Techniques for Review Spam Detection," *Int. J. Adv. Netw. Appl.*, pp. 21–24, 2014.
[26] C. Visani and N. Jadeja, "A Study on Different Machine Learning Techniques for Spam Review Detection," no. August, 2017.
[27] K. Zainal, N. F. Sulaiman, and M. Z. Jali, "An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka*," Int. J. Comput. Sci. Inf. Secur.*, vol. 13, no. 3, pp. 66–74, 2015.
[28] Lichman, M., "UCI Machine Learning Repository", [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
[29] Salleh, S. N. M., Din, R., Zakaria, N. H., & Mustapha, A., "A Review on Structured Scheme Representation on Data Security Application,". *Indonesian Journal of Electrical Engineering and Computer Science*, 11(2), pp. 733-739, 2018.
[30] Umapathy, K., & Khare, N., "An Efficient & Secure Content Contribution and Retrieval content in Online Social Networks using Level-level Security Optimization & Content Visualization Algorithm," "*Indonesian Journal of Electrical Engineering and Computer Science*, 10(2), pp. 807-816, 2018.