VOL. 05 NO. 9,
AUGUST, 2024
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

A Review on Image Spam Detection Techniques using Some Machine Learning Algorithms

Ibrahim Idris
Department of Computer Science, Ibrahim Badamasi Babangida University Lapai Niger State, Nigeria.

Abstract
The most convenient and time efficient method of online communication is through email despite the emergence of alternative forms of online communication which include social networking, sending and receiving. The increase in online transactions via email has led to a significant increase in the global number of email spam which has relatively become a critical problem in the area of computing. There have been numerous techniques of machine learning for identifying unsolicited email spam. Despite the significant improvements made in the number of existing literatures reviewed, there exist no classification technique that has achieve 100% accuracy, each algorithm employs a limited number of features. As a result, determining the most appropriate technique is a critical task because their effectiveness needs to be weighed relative to their drawbacks. To improve on the existing spam detection techniques, we develop a multi-modal architecture capable of detecting image spam, train the model and then, provide an experimental result which indicates that the resultant model performs efficiently.

Keywords: Image Spam, Detection Techniques, Machine Learning, Algorithms, Review

Introduction
The increasing rate of email spam is continuous and alarming, It has created a major problem for service providers, jeopardizing user confidentiality and causing resource loss. Because they cause enormous losses for organizations, ranging from the loading of mail server and waste of bandwidth to the client profitability because of the period of time spent while detecting and dealing with the email spammers. Email spam contents are used for a variety of cybercrime and to circumvent the security measures not only to increase device functionality and loss of storage facility. This violence has the potential to be employed for abusing client data integrity and to steal a percentage of sensitive information including financial information and passwords. The

high volume of spam mail that circulates among networked computers has a negative effect on the memory space of the email server, bandwidth management, the system processing power and application time. Spam email is becoming more of a threat on an annual basis, accounting for more than 77 percent of all email traffic globally (Fonseca *et al.,* 2016).

There have been numerous machine learning techniques for identifying unwanted spam. Despite the relative improvements achieve in the number of existing literature reviewed, there is no classification technique that has achieved 100 percent accuracy (Chopra *et al.,* 2015). For classification, each algorithm makes use of a limited set of features and properties. Because of the important role of splitting ham (non-spam) from spam emails, two approaches have emerged: Knowledge engineering, which uses rules to separate spam from ham email and Content based engineering, which is a machine learning technique that takes decision based on the information similarities and uses heuristic approach to extract ham from spam email through learning from email traffic and then, training the rest corresponding email for the procedure to remain even with no further training (Fonseca *et al.,* 2016).

The majorities of current spam filters are only capable of identifying text based spam or image spam. In this research, a multi-modal architecture is developed that can identify textual spam, image spam, and mixed spam.

THE APPROACH TO SPAM DETECTIONS

Foqaha and Mohammed (2016) show that machine learning technique has gotten a lot of attention, and there are many other approaches that can be used in e-mail filtering. Among these are Nave Bayes, support vector machines, Artificial Neural network, K-nearest neighbor, Rough sets, and the artificial immune system. At the moment, there are two general approaches to spam filtering or email filtering which are:

i. **Content Based Filtering:** Content-based Filtering is a Machine Learning technique that makes decisions based on information similarities. This technique is commonly employed in recommender systems, which are algorithms that advertise or recommend goods to consumers based on information obtained about the user. These methods are based on feature extraction and image content analysis. These filters are used to evaluate and analyze the image's content and processes. The machine learning technique has gotten a lot of attention, and there are many other approaches that can be used in e-mail filtering. Among these are Nave Bayes, support vector machines, Artificial Neural network, K-nearest neighbor, Rough sets, and the artificial immune system (Foqaha & Mohammed, 2016). At the moment, there are two general approaches to spam filtering or email filtering. Two examples are knowledge engineering and content-based filtering in e-mail classification. To classify emails as spam or ham, the knowledge engineering technique must include a set of rules such as the user, the filter (the business software that delivers a certain service).

ii. **Knowledge Engineering Approach:** Masoumeh and Seeja (2015) indicate that knowledge engineering and machine learning are two general approaches used in e-mail filtering. The machine learning approach is more efficient than the knowledge

engineering approach since it does not require the specification of any rules. Instead, a collection of training samples is used, with these samples consisting of post e-mail content. The categorization rules are subsequently learned from these e-mail messages using a specific algorithm. Blacklisting, detecting bulk emails, greylisting, and scanning message headings are all part of the knowledge engineering strategy (Masoumeh & Seeja, 2015). They include blacklisting, detecting bulk entails, scanning message, greylisting.

MACHINE LEARNING METHODS

Machine learning algorithms take data elements and identify them as Yes or No. It's always a dual label situation. The machine is first trained by the human, who labels it Yes or No. Classification is the process of encoding and embedding data and deciding its labels as Yes or No on a consistent basis using machine learning approaches. Machine learning is a data analytics technique that helps computers to learn through expertise the same manner living things do. Instead than relying on a model based on a predetermined equation, machine learning techniques use computer techniques to extract information complex pattern recognition. As the number of samples available for learning increases, the performance of the algorithms changes. The machine learning technique does not require the establishment of any explicit rules," requiring instead a set of pre-classified samples. The categorization criteria are then taught using any suitable approach from this data (Borde*et al*., 2017).

Categories of Machine Learning

i. **Supervised Machine Learning:** By applying what they've learnt in the past to fresh data, algorithms can anticipate future events using labeled instances (x,y). To predict future output values, the learning approach creates an inferred function based on a study of a known training dataset. After a sufficient amount of training, the system may provide targets for any new input. Furthermore, the classification model may contrast its output to the desired output and find errors, allowing the model to be modified as needed. Supervisory machine learning algorithms encompass regression analysis, neural networks, and Support Vector Machines (SVM).It creates a statistical equation of a dataset that has both the desired inputs and outputs (Divya*et al.*, 2018). The data is referred to as training data. It solves classification and regression problems with labeled datasets, resulting in predictive modeling. Linear regression, support algorithms, logistic regression, naive Bayes, and K-Nearest Neighbor are some examples.

ii. **Unsupervised Machine Learning:** Without Supervision, Machine Learning Algorithms are utilized when the training data is neither classed nor labeled (y). Unsupervised learning investigates how a system might infer a function from unlabeled input to describe a hidden structure (that is, an inherent structure). The system does not establish the proper output; instead, it analyzes the input and employs datasets to infer hidden patterns and relationships from large dataset. These algorithms find structure in a set of data that only consists of inputs, such as data

point grouping or clustering. As a result, unlabeled or unprocessed test data is used to train the algorithms. It is used to solve clustering (grouping) and anomaly detection difficulties. Unsupervised learning is used in the k-means clustering and association rule, making it descriptive modeling. Unsupervised Machine Learning Algorithms include: Anomaly detection utilizing the Density Function and Gaussian (normal) Distribution using K-means Clustering.

MACHINE LEARNING ALGORITHMS

A. **Clustering Technique:** Clustering works by combining related pattern classes into a single group. Clustering is a type of method that breaks case studies into groups and compares them. This technique has attracted the interest of scientists and academics, and it has been employed in a variety of sectors. These are unsupervised learning approaches that were used to an email spam dataset with true labels, (Dipika & Kanchan, 2017). Given the availability of adequate representations, a variety of clustering approaches can categorize email spam datasets into spam or ham clusters as demonstrated in their work on email spam clustering.

B. **Support Vector Machine:** Support Vector Machine (CSVM) area type of controlled learning algorithm that has been shown to outperform other learning algorithms. SVM refers to a group of algorithms that were developed to solve classification and regression problems. SVM has used application to provide answers to quadratic programming issues with inequality defects and sequential equality by dividing distinct classes using hyper plane. It makes the best possible use of the border (Dipika & Kanchan, 2017). Although the SVM is slower than other classification methods, it has a higher accuracy due to it's ability to use a multifunctional model border that is not linear or sequential.

C. **Decision Tree:** A Decision Tree (DT) is a type of classifier that uses a tree structure and follows a similar pattern to a decision tree. According to Dipika and Kanchan (2017), the decision tree classifier is a unique method that contributes to categorizing information. A decision tree node can be a leaf node that conveys the definition of the initial purpose (class) or a decision node that indicates that a particular verification should be performed with one branch and areas of work as a subset of the bigger tree that reflects any test outcome. Thus, decision tree learning has been effectively used for filtering spam email.

D. **Random Forest (RF):** This is a popular instance of an ensemble learning technique that is suitable for classification of data in to classes (Dipika & Kanchan, 2017). For the first time, random forest was proposed by Torabi*et al.*(2015). The technique makes specialized predictions using a tree structure. At the stage of training, some decision trees are created by the writer of the program. These decision trees are then applied for the task of predicting the group; this is done by considering the chosen groups of each tree and the category. These decision trees are then used for the purpose of predicting the group, this is done by taking into consideration the selected groups in each tree and the group with the highest number of votes is taken as an output. Random

forest approach is gaining more prominence these days and has been applied in a number of field and literature to solve the analogous problem according to Torabi*et al.*(2015).

Table 1: Summary for the classification Algorithms

Pub. Year	Techniques	Advantage(s)	Limitation(s)
2016	Decision Tree	Has a very short training time.	Not flexible for adjustment.
2016	Random Forest	Provides higher performance and maintain lower classification error.	Longer training period.
2015	Support Vector Machine	Efficient mechanism during the data lost It has capacity to model un-sequential and un-straight forward borderlines which are not (multidimensional).	Slow classification.
2016	Clustering Technique	Capable of processing encrypted message content.	Cannot locate sensitive comparators.

Performance Metrics

Various performance indicators, such as Accuracy, Recall, F1-score and Precision, were employed to assess the usefulness of the suggested strategy. Furthermore, The Confusion Matrix, often known as the Error Matrix, is a popular tool for analyzing performance algorithm results. The confusion matrix is used as the primary evaluation metric for spam detection.

Confusion Matrix (CM)

A Confusion matrix table is a table which is employed frequently to describe performance effectiveness of a classification model (classifiers) on a given test data which is also known as true data. It is a brief and logical categorization of task and prediction outcomes.

The confusion matrix is depicted in the Table 2, along with the anticipated column and actual class row names.

Table 2: A Confusion Matrix.

	Class1 Predicted	**Class 2 Predicted**
Class1 Actual	TP	FN
Class2 Actual	FP	FN

The following components can be deduced from the definition of Confusion Matrix (CM):

- True Positive (TP): The percentage of spam classified correctly.
- The number of correctly categorized legitimate emails in the True Negative (TN) category (Ham).

- The number of genuine emails that are incorrectly labeled as false positives (Ham) (FP).
- The number of spam messages that have been misclassified as false negatives (FN).

Each of the metrics described is applied to data so as to evaluate the performance measure and compare the various classification reports in order to determine the best classifier model efficiency of the algorithm in the supervised machine learning method being used. It is a feature of every good model to be able to successfully generalize in order to test data which is significantly different from the given training dataset. A model built for training data by learning which scenarios fill the best may not perform well on test data. The confusion matrix is also calculated using the following parameters:

Accuracy Rate

The relationship between classification rate and accuracy is:

$$\text{Accuracy Rate (Acc)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Accuracy Rates: This metric indicates the efficiency of the binary classification test works. What the percentage is of correctly predicted events actually occur. Accuracy alone isn't a good measure because it doesn't tell you how effectively the model detects positives and negatives independently. It assumes that the costs of both types of errors are equal. Depending on the difficulty, 99 percent accuracy might be outstanding, decent, middling, poor, or even terrible.

Recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Rates of recall:- The recall is calculated ratio of the classified positive spam to the total positive spams. It describes how effective a test is at detecting positive spam. To put it another way, positive outcomes are predicted to be positive. The class was accurately identified if the recall was high (small number of FN). A good model should have a high recall rate. Sensitivity or TP Rate is other terms for recall. The recall (Rs) metric indicates how many spam messages the filter successfully prevented (i.e. the degree of its effectiveness of blocking actual spam correctly)

Precision

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

The total number of properly categorized positive spam is divided with the total expected positive number of spam to determine the precision rates.. A high precision suggests that a test data email that has been flagged as spam is, in fact, spam (small number of FP Rate) the percentage of positive projections are correct. A high precision level is excellent. Precision (Ps)

measures the proportion of messages labeled as spam by the filter that were, in fact, spam.

F1-Measure

Because there are two measurements (precision and recall), having a better measurement that encompasses both is advantageous. The F1-measure is calculated using Harmonic Mean rather than Arithmetic Mean since extreme values are penalized more severely. The smaller precision or recall value will always be closer to the F1-Measure. It is highly beneficial to have a high f1_measures.

$$\mathrm{F1-Measure} = \frac{2 * \mathrm{Recall} * \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}}$$

Table 2: Summary of the classification Techniques with different performance metrics.

Year	Dataset	Techniques	Accuracy	Precision	Re- call	F- Measure
2015	Dredze	SVM	90%	-	-	-
2018	Spam base	Random Forest	94.2 %	94.2%	94.2 %	94%
		Decision Tree Bayes	88.2 %	88.5%	88.5 %	88.5%
2015	Spam base	SVM	79.5 0%	79.02 %	68.6 9%	-
		Clustering	76.2 4%	70.59 %	72.0 5%	-
2018	Spam base	Random	92.41%	92.40%	92.4%	-
		SVM	0.85	-	0.91	-
2016	Dredze	SVM	0.97	0.97	0.68	-
		Decision Tree	0.92	0.92	0.92	0.92
2017	Spam base	Clustering	0.89	-	0.90	-

CONCLUSION

Sending and receiving emails has remained the most convenient and time efficient method of online communication. The increase in online transactions via email has led to a significant increase in the global number of email spam which has relatively become a critical problem in the area of computing. There have been numerous machines learning technique for identifying unwanted email spam. Despite the significant improvements made in the number of existing literatures reviewed, there is no classification technique that has achieve 100% accuracy, each algorithm employs a limited number of features. As a result, determining the most appropriate technique is a critical task because their effectiveness needs to be weighed relative to their drawbacks. In order to improve spam detection rate, some machine learning algorithms were reviewed and were used to implement a spam detection system capable of handling text of email formats. The existing literature reviewed has indicated that a significant progress has been made in email spam detection; however, more

research effort needed to improve on the performance and effectiveness of the spam email detection. More effort is also required to work more on the availability of well labeled dataset in order to enable a successful email spam classification. In this study, a variety of real-world text spam datasets have been utilized to develop powerful classifiers based on machine learning algorithms.

REFERENCES

Abdullahi, M., Bashir, S. A., Abdulmalik, D. M., &Abisoye, O. A. (2021). A Review on Machine Learning Techniques for Image Based Spam Email Detection. *2020 IEE 2nd International Conference on Cyber Space (CYBER NIGERIA),*2(25), 120-125. doi:10.1109/CYBERNIGERIA51635.2021.9428826.

Abdulhamid, S. M., Osho, O.,Ismaila, I., &Alhassan, J. K. (2018). Comparative Analysis of Classification Algorithms for Email Sparn Detection. *International Journal on Natural Language Computing (IJNLC),*9(6), 111-123. doi:10.5815/ijcnis.2018.01.07

Alexy,Y. &Shyamanta, T. (2016). Approaches to E-Mail Spam Detection and Classification Using SVM and Feature Extraction. *International Journalon Natural Language Computing (IJNLC),* 2(4), 119-126.Retrieved from https://www.ijconline.edu/42689123.

Al-Shboul, A. B., Hakh, H. H., Faris, I. A., &Alsawalqah, H. (2016).Voting-based Classification for E-mail Spam Detection. *Journal of ICT Research and Applications*, 10(1), 29-42. doi:10.5614/itbj.ict.res.appl.2016.10.1.3

Bhowmick, A,. & Hazarika,S. M.(2016). Machine Learning for E-mail Spam Filtering Review, Techniques and Trends. *International Journal of computer and Network Secuiry*, 4(2), 91–97.doi:10.48550/arXiv.1606.01042

Chopra, Nisha, D., and Gaikwad, K. P. (2015). Image and text spam mail filtering. *International Journal of Computing Technology and Electronic Engineering (IJCTEE),* 5(3), 17-29.doi:10.1007/572-1109-10-9785-9_2

Choudhary, N., & Jain, A.K. (2017). Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique. *international Journal* of *Advanced Informatics for Computing Research,*712, 18 - 30.doi:10.1007/978-981-10-5780-9_2

Divya, K. S., Hargavi, P. B.,&Yothi, P. J. (2018). Machine Learning Algorithms in Big Data Analytics. *International Journal of Computer Sciences and Engineering*, 6(1), 63-70.doi:10.26438/ijcse/v6il.6370

Fonseca, D.M., Fazzion, O. H., Cunha, E., Guedes, P. D., Meira, W., & Chaves M. (2016). Measuring Characterizing, and Avoiding Spam Traffic Costs, *IEEE international Journal of Internet Computing*, 20(4), 1-1. doi:10.1109/mic.2016.53