International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

A Machine Learning Based Email Spam Classification Framework Model: Related Challenges and Issues

Deepika Mallampati, Nagaratna P Hegde

Abstract*: *Spam emails, also known as non-self, are unsolicited commercial emails or fraudulent emails sent to a particular individual or company, or to a group of individuals. Machine learning algorithms in the area of spam filtering is commonly used. There has been a lot of effort to render spam filtering more efficient in classifying e-mails as either ham (valid messages) or spam (unwanted messages) through the ML classifiers. We may recognize the distinguishing features of the material of documents. Much important work has been carried out in the area of spam filtering which cannot be adapted to various conditions and problems which are limited to certain domains. Our analysis contrasts the positives methods as well as some shortcomings of current ML methods and open spam filters study challenges. We suggest some of the new ongoing approaches towards deep leaning as potential tactics that can tackle the challenge of spam emails efficiently.

Index Terms: SVM, Machine learning, Deep neural network neural networks, spam, ham

I. INTRODUCTION

In recent days internet has become an important part of daily life and email seems to have become a potent tool for exchanging information. Together with the advancement of technology and e-mail, spam has grown dramatically in recent years. Spam can arrive from anywhere in the world where Internet access is accessible. The amount of spam messages keeps increasing exponentially, following the advancement of Antispam programmers and technology. That organization needs to evaluate the resources available to find out how best to combat spam in its community and tackle the widespread problem. Spam can be described as "unrequested" bulk emails [1], which are typically commercial, as per the direct and most common description. Spam e-mails are unacceptable notifications sent either directly or indirectly, systematically, without user communication or interaction [2]. Spam communications are called spam because they are not useful to the receiver.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
Deepika Mallampati*, Research Scholar, Osmania University, and Assistant Professor, Department of CSE, Neil Gogte Institute of Technology, Hyderabad, Telangana, India.
Dr. Nagaratna P Hegde, Professor, Department of CSE, Vasavi College of Engineering, Hyderabad, Telangana, India.

Many email users frequently remove unwanted emails (spam), even when an increasing amount of such emails takes up computer storage space and absorbs network bandwidth. The probability of spam mail increases globally, accounting for more than 77% of global email traffic [3]. Furthermore, the creation and enhancement of software classifiers that can discriminate between genuine e-mail and spam presents a continuing problem. Several reported experiments tested spam detectors utilizing Naive Bayesian methods and broad binary feature sets that assess the presence of spam keywords and Naive Bayesian strategies are used in other commercial applications, too. Spammers accept such efforts to block their communications and establish strategies to circumvent certain filters, but these confrontational techniques are habits that human users can often quickly identify. The purpose of this research was to establish an alternative strategy using a neural network (NN) classification system that uses a corpus of e-mails sent by several users. The collection of applications used in this research is one of the key enhancements because the feature set utilizes descriptive properties of terms and messages identical to those used by a human user to recognize spam, and because the algorithm to select the best feature set is focused on a range of forward applications. Another goal in this work was the improvement of the spam detection by using artificial neural networks, which is almost 95% accuracy; nobody achieved more than 89% accuracy with ANN.

Fig. 1. A word cloud of common words in spam e-mail.

The approach to machine learning seems to be more successful than the approach to software engineering because no rules are needed [4]. Alternatively, a variety of testing examples are a collection of e-mail messages pre-classified. A certain algorithm is then employed to know from these e-mail messages the classification rules. Machine learning techniques have been studied extensively and many algorithms can be used to process e-mails.

In [5] the authors have created several spam filters that block such spam emails from accessing a user's box, so there is not enough work to concentrate on text improvements. Naive Bayes is currently one of its most popular ways for the classification of spam due to its flexibility and performance. Naïve Bayes is also very precise; nevertheless, emails cannot be properly categorized. Therefore, we have developed a new algorithm to improve the accuracy of the Naive Bayes Spam Filter, enabling us to identify text changes and label the email as spam or ham properly.

The Python program blends semantic, keyword and machine-learning algorithms to improve Naive Bayes ' precision by more than 95% relative to the Spam assassin. Besides, we found a link between the duration of the email and the spam ranking, suggesting that Bayesian harming, a contentious topic, is a common phenomenon that spammers use. The major focus on the work and its influence towards spam filtering discussed below

a. They provided an analysis of the key features of e-mail spam, trends and improvements. We have thus outlined significant research directions that should be concentrated to more extent.

b. We addressed spam filter design and ML techniques for the Gmail, Yahoo mail and Outlook message spam filtering method. The various components of the email spam system have been explored beautifully.

c. We published an in-depth study of various techniques for email spam filtering and a thorough overview of spam filtering literature during the time (2006-2018).

d. We also introduced researchers to use effective algorithms that are possible with ML techniques for spam filtering purpose.

e. We also clearly indicated our observations of some open challenging issues with spam filters and proposed proactive steps to improve machine learning approaches to reduce future development of new methods of spam that could easily avoid filters.

II. RELATED WORKS

A brief survey is provided to investigate the weaknesses in the use of message filtering and retrieval technology to logically and potentially postulate email spam identification to allow the implementation of an effective spam filtering technique. To increase the readability of the text and to boost readers ' comprehension, the layout of this paper is shown in Fig. 2 below:

Spam filters are in two broad categories:

1) Non-Machine-based learning

2) Machine-based learning

Non-machine-based learning: Many early anti-spam methods belong to this category; for example use blacklisted spammers, white lists of secure sources, or a collection of man-made keywords, such as "make wealthy." Nonetheless, such static lists may often be used by spammers, such as modifying or spoofing the address or domain of the sender.

Spammers have learned to avoid / miss terms purposefully or to avoid the spam filters. Such methods require regular manual changes, and the risk of screening out an honest messages is high because spam is more extreme than filtering. Inaccurate anti-spam technologies may be liable for losing more than five million working hours a year for users to verify, based on estimates from the British Computer Society, that legal massages have not been wrongly quarantined.

Heuristics Benefits: The Heuristic mail filters are deemed simple, highly accurate and effective against the regulatory speech.

Limits: Heuristic mailed filters do not have intelligent learning capability (not suited to new SPAM features; they allow manager intervention of 2 modifications, so that rule changes or guideline sets need to be periodically updated, as well as high rates of false-positive when the sensitivity rises.

Benefits: Based on the high resistance of collision hash functions, signature mail filters produce low levels of false positives.

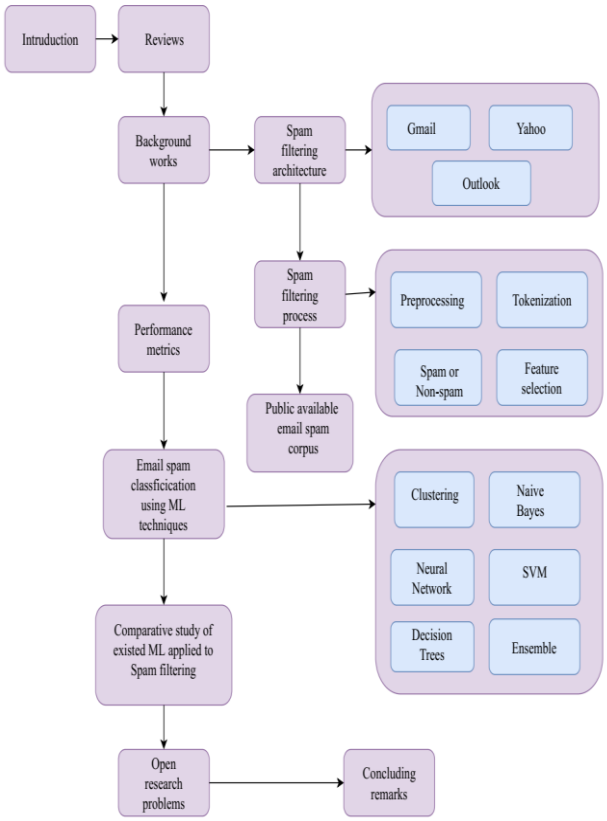

Fig. 2. Overview of the paper that described flow process

Limitations: The signature mail filters do not have intelligent learning functionality (they cannot recognize hazards in the case of new SPAM emails), allow the user to refresh the SPAM hash list or the list should be periodically fetched from a delivery server, and if updated, the filtering system does not identify the pre-known SPAM emails. Changes to pre-known SPAM emails create a different hash than the filtering system existing. Then, the updated SPAM email passes through the filtering system.

Benefits of blacklists: Blacklist filtering/whitelisting are deemed simple, quick and easy to enforce.

Limitations: One of the big drawback to blacklisting/ whitelisted filtering is that the sender's email address can be quickly spoofed.

Traffic Analysis Benefits: Mail filters are considered relatively complicated for traffic analysis. The method, though, recommends improved and fast mail filtering compared to actual e-mail content analysis as they review SMTP logs only.

Limitations: Mail analysis spam filters do not have sophisticated learning skills (they don't respond to new SPAM features). At this stage, it is not feasible for filters to determine which email traffic characterizing attributes are most appropriate for a certain email traffic source.

Machine-based learning: Machine learning focused on conventional technologies, this approaches dynamically interpret the output of the messages obtained and construct more reliable models. We can, therefore, be more successful and constantly revised to cope with the strategies of spammers.

Many machine learning approaches, including spam filtering, have been recently used.

Bayesian Benefits: Intelligent modeling (ML) and improved filtering focused on content analysis are used for Bayesian mail filters. It allow e-mail users to modify the filters to the form of SPAM that users will get. The Bayesian mail filters are thus found to be very reliable.

Limitations: The tokens are composed of single words in Bayesian filters. Combined terms can, therefore, leakage from identification. The filters cannot often evaluate consecutive terms that represent can spammer phrases, such as the expression "special offer," in the email content. Alternatively, every term is independently evaluated. Within SPAM emails, other phrases and mixed terms are often found. If such terms are not known, the filtering function for identifying these SPAM e-mails is restricted. Nevertheless, there are already other algorithms that help to evaluate permutations of single words, successive words, and words within a gap from each other.

Gray Listing Benefits: SPAM e-mails are fast, simple and streamlined to identify using a gray listing. This uses the standard mail protocol format. E-mail blocking is a function of the mail servers (MTAs). Therefore, the implementation of external hardware or software is not required. It also offers a safe way of controlling SPAM e-mails by not first accepting (rejecting) messages from spamming sites. The system has proved effective in preventing spamming server zombie computers.

Limitations: The gray list mail denial cannot be treated as a full anti-SPAM remedy. Despite its effectiveness, e-mails that require a prompt response can cause great inconvenience. For instance the websites requesting user feedback by email to complete their domain registrations. The technique can prevent users from registering for some time. Another major drawback may arise if the waiting time provided to the returned email is not met by the source email server. In this scenario, the mail is sent after the period expires. The mail will therefore not hit the receiver. As a consequence, the mail server receiver can regard and block the source mail server as a spammer.

III. BACKGROUND

Architecture for filtering of Spam:

It is intended to reduce the number of phishing emails to a low. Also, it is the process of retrieval for messages in order to reorder them according to certain criteria. Mail filters are typically used for handling incoming emails, screening spam, identifying and deleting emails carrying some harmful cryptograms like viruses, Trojans and malware.

Gmail filter spam: A test will be conducted using the sensitivity threshold score decided by the spam filter of each user. And it is therefore known to be genuine or spam text. Gmail often utilizes OCR to secure image spam consumers from Gmail spam. Furthermore, machine training algorithms designed to integrate and identify suitable datasets from Google that permit Gmail to increase its spam detection. The changing essence of spam is based on factors like the credibility of the network, connections with associated formatting message headers. This kind of filter primarily relies on "password" settings, which are continuously updated with contemporary software, updated detection mechanisms of spam and reviews from Gmail users on potential spammers.

Yahoo filter spam: It seems to be the world's formerly online webmail service with more than 320 million subscribers. The e-mail service has its own spam algorithms which are used for spam detection. Yahoo's specific strategies for spam detection include URL scanning, email material, and user spam reports. By comparison to Gmail, Yahoo scans domain emails and not IP addresses. Thus it is the integration of spam filtering methods that can provide mechanisms to prevent a valid user from erring for a spammer.

Outlook email spam filter: It is a bunch of applications that belong to Microsoft, including a webmail service from Outlook. The webmail software from Outlook allows users to send and receive emails through their web browser. People will add cloud storage resources to their accounts so they can choose data either from the local device or else OneDrive account correspondingly from Google Drive, Box and Dropbox while sending an email with file attachments. For a fact, the Outlook Webmail program often helps consumers to encrypt their e-mails and deny recipients. If the message sent to the mail is encrypted at the Outlook.com server, the authentication user can decode and read the message. Such a security measure means that the document as shown in Fig.3 can only be interpreted by the intended recipient.

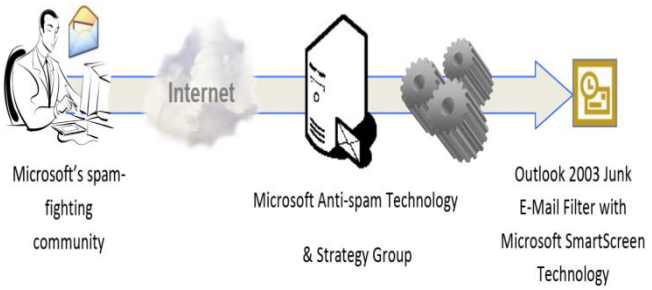

Fig.3. Flow of Microsoft smart spam E-Mail Filter process

Email spam filtering process:

Fig. 4 describes the usual use of a spam filter from a single user's perspective. Incoming messages are handled on a one-by-one basis and are marked as ham (a colloquial term commonly used for non-spam) or spam. Ham is forwarded to the daily read inbox of the customer. Spam is sent to an irregularly read quarantine server that can be scanned to try and find ham messages that the device misclassifies.

Pre-processing: This is the first step when a mail is sent. This is the first point. This phase requires tokenization.

Tokenization: This is a mechanism by which the terms are separated from the email address. It also divides a word into its significant fragments.

Feature selection: The pre-processing phase is preceded by the feature selection phase. It is a reduction in spatial distribution estimation that essentially displays interesting email text fragments as a feature type matrix.

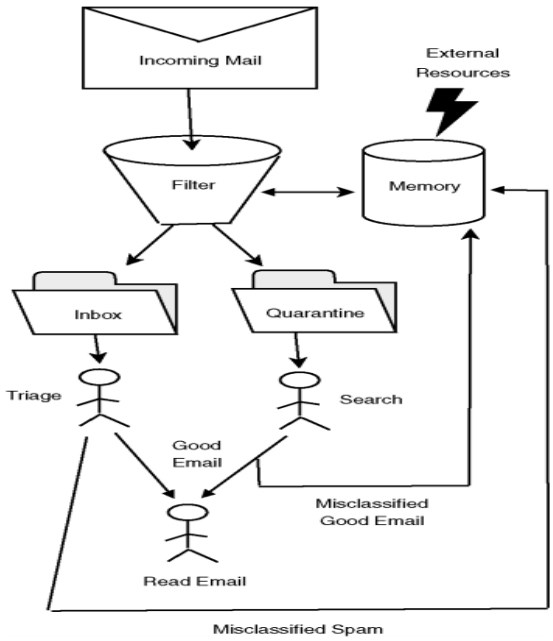

Fig. 4 Spam filter usage

IV. METHODS AND METHDOLOGY

Several scholars have already made efforts to combat spammers. In [6], the authors define a multi-layer spam detection system. However it also extended to anti-spam initiatives on both the server and the client side. In [7] the authors suggest a multi-stage phishing system. The product of the classification in the primarily class is given to a second classifier. If the second and first stage levels fit, the outcome is called the appropriate performance. Furthermore, if the findings vary, a different type is used to categorize the e-mail. The third type performance is assumed to be the right e-mail form. The choice of three classifiers are therefore used to achieve an email rating. Furthermore, if an e-mail of trust above the threshold can be listed in our suggested model, the lower level is not invoked. So in this carried work will accommodate more than three classifying levels in order to condense as much as possible the false positive rate. In [8] the writers mentioned a hierarchical spam detection system. The first layer is a document description. But in our situation, we have taken other spam email habits such as sender's blacklisting, sender integrity etc. into account. Finally, the last layer includes negative selection-based filtering which is more efficient against fresh spam e-mails. In [9], the authors propose a method that focuses primarily on image spam. They have two types of construction the primary belongs to non-image type spam, and secondary type is image spam.

Classification algorithms are named to classify a given instance into a variety of discrete categories. Such algorithms are used to create a model or set of rules that translate a specified input into a series of discrete output values. Many classification algorithms may take input in any way, whether discreet or continuous, although all the inputs are also discreet in some classification algorithms. The result is always a discrete value. Types of classification algorithms include decision trees and Bayes nets. In order to apply algorithms for classification on our weather example, we need to translate the output attribute into groups. This is usually done by discretization, which splits a continuous variable into groups.

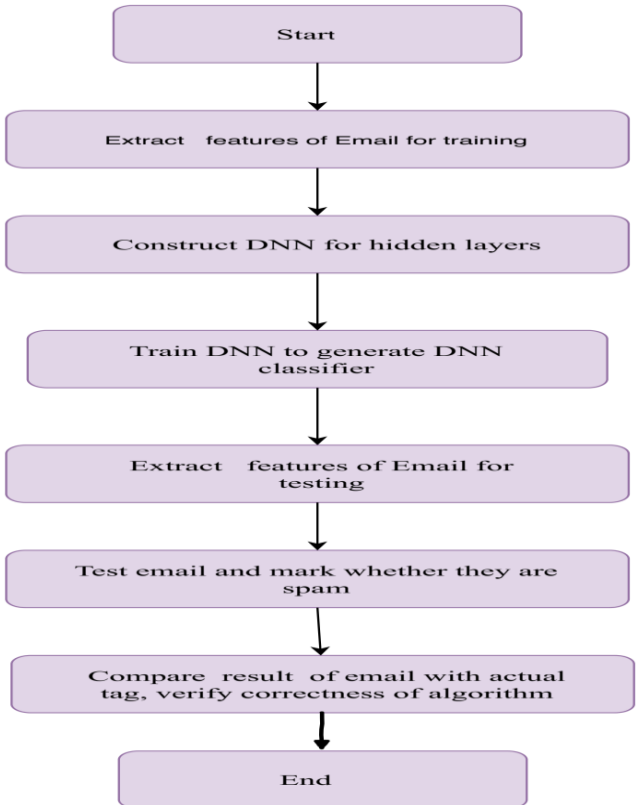

Fig. 5 Methodology of DNN classifier

Deep Neural Network (DNN): A binary distinction that determines whether all the e-mail is spam or ham. The pre-trained data is split into training data (80%) and test data (20%) before they have been submitted to the Deep Neural Network (DNN). The training data is entered in the DNN classification and the system is trained to detect spam & ham SMS messages. The trained DNN classifier has been used in the test data to model the SMS. DNN operates a binary analysis to decide if the SMS is spam or ham as shown in Fig.5.

Naive Bayes algorithm: It generates a probabilistic model by studying the conditions of each input attribute with a potential value obtained from the output attribute. This model is used to predict an output value when a collection of inputs is given. This is achieved by applying the Bayes rule on the condition that the output value can be seen when the attribute values in the particular instance are viewed together. We first describe the Bayes law before explaining the algorithm.

The rule of Bayes notes that when P (A) is specified as the likelihood of observation A, because B occurs. P (A) is called post-trust probability, and before probabilities are called P(B), P(A) and P(B). The theorem of Bayes defines a connection between the posterior likelihood and the previous probability. This allows one to consider the possibility of A given B when the actual probabilities of A and B are identified, and the likelihood of B provided A is also understood. The naive Bayes algorithm uses a number of training examples to identify the Bayesian method in a new example.

For e.g., the Bayes rule is implemented to find the likelihood of experiencing each output class, provided the input attributes, and the instance is associated with the class with the highest probability.

The values of probability used are derived from the numbers of attributes in the training set.

Support vector Machine (SVM): Its job is to classify the objects in just one class out of two. It will do this operation by looking for the maximum hyperplane margin that classifies

Table 1 Summary spam filtering using Machine Learning techniques

Authors	Compared with others	Merits with Performance metrics	Algorithm/Technology used	Demerits	Dataset used
Authors in, [10]	Compared with other Spam classification methods implemented in Hybridized ACO and SVM with KNN, NB and SVM	Maintaining optimal solutions to such nonlinear problems. Accuracy, precision and recall	Uses the feature selection using Ant Colony Optimization and SVM combine	It has very poor performance	Spambase dataset
Authors in, [11]	Compared with kNN with spearman and kNN with Euclidean	Maintains improvement in accuracy with higher F-measure.	Used KNN technique for classification with associated spearman correlation.	It has poor performance	Spambase dataset
Authors in, [12]	Compared with Bayesian classification, kNN, ANNs, SVMs, Artificial Immune System and Rough sets.	It is measured in terms of spam recall, precision and accuracy.	It introduces a Bayesian classification, k-NN, ANNs, SVMs, Artificial Immune System and Rough set.	Lots of the contemporary spam classification methods were not studied.	SpamAssassin
Authors in, [13]	Compared with ABFPA, BPSO, SFLA for feature selection.	It accomplishes the global optimization, and lower computational cost.	It uses the efficient adaptive binary flower pollination algorithm (ABFPA)	Does not used any evaluation metrics for the proposed work	Dataset not mentioned
Authors in, [14]	Compared with NN, MLP and Perceptron.	Capable of filtering efficiently spams	Usage of Artificial neural network with back propagation	Takes longer time for the training phase	Randomly collected emails
Authors in, [15]	Firefly, NB, NN and PSO algorithm.	Spambase dataset the accuracy of the proposed algorithm performs 26.7% better than Neural Network and 2.4% better than PSO algorithm. Sensitivity, specificity and accuracy	Best suited approach that deals with classification of the spam email using firefly and Naïve Bayes classifier.	It has poor performance	CSDMC2010 dataset
Authors in, [16]	No comparison done	Not stated	It uses an efficient genetic algorithm	Performance not related with other method	Words in data dictionary
Authors in, [17]	Compared with NSA, PSO, SVM, NB and DFS-SVM	Accuracy	It uses an Negative selection and PSO	Accuracy is only used for assessing it performance	Ling dataset

Authors in, [18]	No comparison done	Classification accuracy.	Designed GA with Heuristic Fitness Function	No Accuracy metric of proposed work is related.	Created dataset with 2248 emails
Authors in, [19]	Compared with PSO, SOM, kNN and SVM	Area under curve AUC	Implemented work with PSO, ANN and SVM	Area under curve is only used for evaluation	Spambase dataset
Authors in, [20]	Compared with MLP, C4.5, NB	Prediction Accuracy, FP	Implemented with WEKA	MLP consumes more time compared with NB and DT	Dataset not mentioned
Authors in, [21]	No comparison	Accuracy	Implemented with kernel function in R	High FP	UCI ML dataset

Table 2 Summary of publicly available email spam corpus

S.No.	Name of the Dataset	No. of messages		Spam Rate
		Spam	Non-Spam	
1.	Spam archive	15090	0	100%
2	Spambase	1813	2788	39%
3	Lingspam	481	2412	17%
4	PU1	481	618	44%
5	Spamassassin	1897	4150	31%
6	Gen spam	1205	428	18%
7	Trec 2005	52790	39399	57%
8	Biggio	8549	0	100

the textual function category between two classrooms (0 or 1), the maximum distance from both classes between the hyperplane and the closest object.

Decision Tree (DT): This approach is based on the hierarchical decomposition of the training data field during which the text (attribute) attributes of the tree nodes and the distinctions between them are defined by the weight of the item in tests, which essentially label the leaves as class names. The inclusion or absence of one or more terms (features) is used to separate the results.

Random Forest (RF): This method of sorting encompasses several decision-making systems. And the input characteristics checked by each of the trees in the forest to identify a new item.

J48 Classifier: It is an ID3 branch with several features in J48 include incomplete properties, decision-making chains, constant meaning sets for attributes, law mathematical formalism etc. J48 is a free software Java application of the C4.5 algorithm in the WEKA data mining mechanism.

Table .1 shows that the different spam filtering using Machine Learning techniques with their findings and merits for datasets.

V. RESULTS AND DISCUSSION

Usage of open dataset for email spam corpus:

The dataset in a database shows an important aspect for determining the output spam filter. Although numerous traditional data sets are used usually for text classification, it is recently spam filtering experts are trying to make accessible to the public the corpus they are used to evaluate the effectiveness of their proposed filter.

Table 2 provides an extensive list of corpuses made publicly available in the various techniques discussed in this article. Human corpus has unusually defining attributes suggested by the relevant information used in studies to test the efficiency of the spam filter.

Evaluation Metrics:

Traditional information retrieval methods were used to test the feasibility of the suggested solution. The main metrics used in the assessment of search methods were accuracy and recall.

Precision, is defined as the process of finding out the exactness and also coined to be predictive value that is more positive outcome, Precision *p* is can be put in the mathematical form as:

$$\text{Precision P} = \frac{TP}{TF + FP} \qquad (1)$$

Accuracy is defined as the process of finding out more number of messages to be positive for filter of spam emails.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

Recall, also known as responsiveness, may be defined as a ratio of total number, is the measure of completeness.

It is the average of the number of records retrieved to the total number of records in the database.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

The paper also provides a list of issues, difficulties and brief analysis of various challenges and problems presented by different authors, utilizing particularly different methods and techniques of defense.

• No suitable unified collection of coherent techniques which could encompass all huge Internet-based e-mail spam groups.
• The existence, through the use of currently distributed computing tools, of a flexible classification of spam mails.
• The absence of a complex pattern dependent spam filtering techniques of mails,
• The absence of machine effective classification process of spam as well its management schemes, which profit with a mixture of software utilities.
• Developing more effective spam filters for images. Some spam filters can only categorize text spam messages. Most experienced spammers, however, send spam emails as text inserted in an image (stego image) to avoid the identification by filters.
• The need for flexible, robust and optimized filters to be built by spam filtration with ontology and the semantic web.
• Lack of plugins that can update the screen space dynamically. Many existing spam filters cannot attach or erase characteristics on an incremental basis without completely re-creating the software to match up with current developments in email spam filtering.

Table 3 Summary of comparative results of classifiers

Evaluation criteria	Naïve Bayes	J48	DNN
Training time	0.12	0.22	2.66
Correctly classified instances of spam	1420	1370	1520
Accuracy	97.2	96.3	99.5

It is clearly shown from Table .3 that the summarization of three ML techniques and its evaluation parameters for comparative results of classifiers.

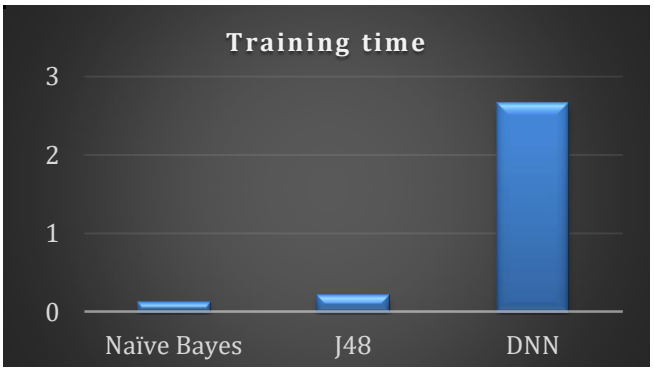

Fig. 6. Comparison of training times of different ML techniques

Fig. 6 shows the train training time of DNN technique is more as compared to other two techniques.

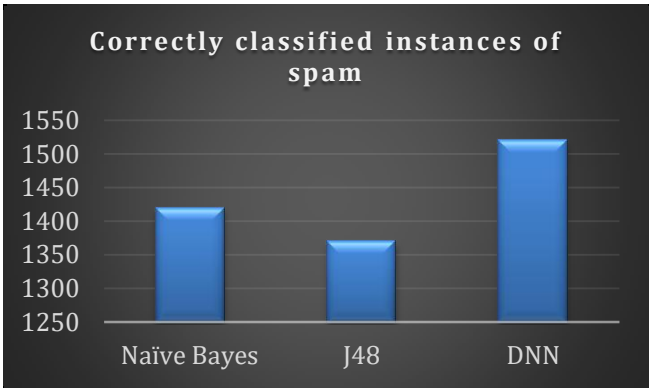

Fig. 7. Comparison of classified instances correctly of different ML techniques.

Fig. 7 shows the classified instances correctly for DNN technique is more as compared to other two techniques.

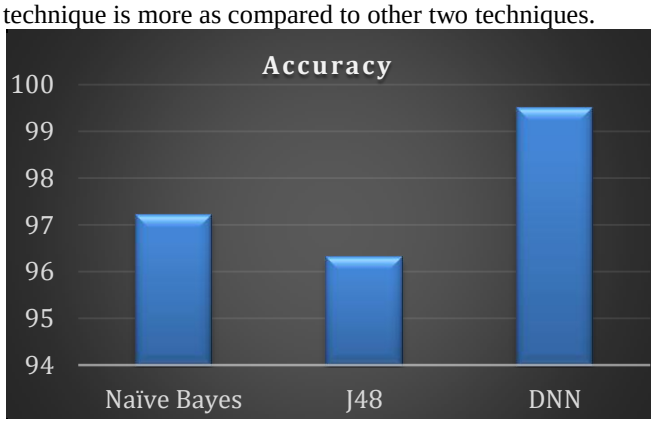

Fig. 8. Comparison of accuracy of different ML techniques

Fig.8 shows the accuracy for DNN technique is more as compared to other two techniques.

VI. CONCLUSION

In this paper, we looked at efficient methods towards spam filtering of emails using machine learning approaches. Also, the study of these improved methods was carried out to classify texts as spam or ham. Some academics have been exploring attempts to address the spam issue by using machine learning classifications. Since discussing the problems of transparent spam filtering, more research is needed to increase the efficiency of spam filters. This will allow the production of spam filters an important field of research for researchers and professionals looking for machine learning approaches for powerful spam filters.

REFERENCES

1. I. Androutsopoulos, J. Koutsias, K. Chandrinos and C. D. Spyropoulos, "An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages," Computation and Language, pp. 160-167, 2000.
2. G. V. Cormack, "Email Spam Filtering: A Systematic Review," Foundations and Trends® in Information Retrieval, vol. 1, no. 4, pp. 335-455, 2006.
3. M. Siponen and C. Stucke, "Effective Anti-Spam Strategies in Companies: An International Study," Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), 2006.
4. Guzella, T. S. and Caminhas, W. M."A review of machine learning approaches to Spam filtering." Expert Syst. Appl., 2009.

5. Linda Huang, Julia Jia, Emma Ingram, Wuxu Peng, “Enhancing the Naive Bayes Spam Filter through Intelligent Text Modification Detection”, 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications.
6. Jianying Zhou, Wee-Yung Chin, Rodrigo Roman, and Javier Lopez, (2007) "An Effective MultiLayered Defense Framework against Spam", Information Security Technical Report 01/2007.
7. Rafiqul Islam, Jemal Abawajy,"A multi-tier phishing detection and filtering approach (2013)",Journal of Network and Computer Applications, Volume 36, Issue 1, January, pp. 324–335.
8. Xiao Mang Li, Ung Mo Kim, (2012)"A hierarchical framework for content-based image spam filtering", 8th International Conference on Information Science and Digital Content Technology (ICIDT), Jeju, June, pp. 149-155.
9. Z. Wang, W. Josephson, Q. Lv, M. Charikar and K. Li.(2007) Filtering Image Spam with near Duplicate Detection, in Proceedings of the 4th Conference on Email and Anti-Spam CEAS.
10. R. Karthika, P. Visalakshi, A hybrid ACO based feature selection method for email spam classification, WSEAS Trans. Comput. 14 (2015) 171–177.
11. A. Sharma, A. Suryawansi, A novel method for detecting spam email using KNN classification with spearman correlation as distance measure, Int. J. Comput. Appl. 136 (6) (2016) 28–34.
12. W.A. Awad, S.M. Elseuofi, Machine learning methods for spam E-mail classification, Int. J. Comput. Sci. Inf. Technol. 3 (1) (2011) 173–184.
13. S.P. Rajamohana, K. Umamaheswari, B. Abirami, Adaptive binary flower pollination algorithm for feature selection in review spam detection, in: IEEE International Conference on Innovations in Green Energy and Healthcare Technologies, 2017, pp. 1–4.
14. I.J. Alkaht, B. Al-Khatib, Filtering SPAM Using Several Stages Neural Networks, Int. Rev. Comp. Softw. 11 (2016) 2.
15. K.R. Dhanaraj, V. Palaniswami, Firefly and Bayes classifier for email spam classification in a distributed environment, Aust. J. Basic Appl. Sci. 8 (17) (2014) 118–130.
16. [16] M. Choudhary, V.S. Dhaka, Automatic E-mails classification using genetic algorithm, in: Special Conference Issue: National Conference on Cloud Computing and Big Data, 2013, pp. 42–49.
17. C. Palanisamy, T. Kumaresan, S.E. Varalakshmi, Combined techniques for detecting email spam using negative selection and particle swarm optimization, Int. J. Adv. Res. Trends Eng. Technol. 3 (2016). ISSN: 2394-3777.
18. J.N. Shrivastava, M.H. Bindu, E-mail classification using genetic algorithm with heuristic fitness function, Int. J. Comput. Trends Technol. 4 (8) (2013) 2956–2961.
19. M. Zavvar, M. Rezaei, S. Garavand, Email spam detection using combination of particle swarm optimization and artificial neural network and support vector machine Int.
J Mod Educ. Comput.Sci. (2016) 68-74.
20. Deepika Mallampati, “An Efficient Spam Filtering using Supervised Machine Learning Techniques” in IJSRCSE, Vol.6, Issue.2, pp.33-37, April (2018).
21. Deepika Mallampati, K.Chandra Shekar and K.Ravikanth “Supervised Machine Learning Classifier for Email Spam Filtering”, © Springer Nature Singapore Pte Ltd. 2019 and Engineering, https://doi.org/10.1007/978-981-13-7082-341.