IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.569

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

Development and Analysis of Spam Mail Identification Model

Jayant Batra[1], Kirti Bhatia[2], Rohini Sharma[3], Shalini Bhadola[4]

P.G. Student, Sat Kabir Institute of Technology and Management, Bahadurgarh, Haryana, India[1]

Assistant Professor, Sat Kabir Institute of Technology and Management, Bahadurgarh, Haryana, India[2]

Assistant Professor, Government College for Women, Rohtak, India[3]

Assistant Professor, Sat Kabir Institute of Technology and Management, Bahadurgarh, Haryana, India[4]

ABSTRACT: Information sharing within the divisions of the organizations is a necessary part of today's computerized business oriented realm. Email is an indispensable and valuable tool of fast and low-cost communication. With the help of the e-mails people can be easily connected to one another. But the facility of e-mail also includes some problems like spam mails (or unwanted bulk email). These spams have become a challenge for business organizations and IT industry. This speedy development of spams, unnecessary fills a user inbox, swamping some worthy emails, exhausting memory and bandwidth and it takes enough time in deleting them.In this study, we looked at the problem of spam email, which is a major concern in the web world. The escalating volume of undesirable e-mails (spam mails) has necessitated the development of a consistent anti-spam filter. Machine learning (ML) processes are now being used to effectively screen spam e-mail on a continuous basis.In this work, we looked at specificpopular ML techniques and their applicability to the problem of spam mail categorization. Based on the number of Spam Assassin users, we have produced descriptions of the processes and the deviation in their execution. This dissertation suggests a content-centered email spam filtering scheme with a group of fresh algorithms and few prevailing algorithms. We have extracted the features of the e-mails. We have preprocessed the e-mails and created the vocabulary list. We have used support vector machine (SVM) ML classifier to train the datasets of the e-mails. Then we have classified the e-mails into spam and non-spam mails. All the tests have been conducted on the famous freely accessible corpuses.

KEYWORDS: Spam & Ham, Machine Learning, support vector machine

I. INTRODUCTION

Now a day, the E-mail is a very well-known and used form of communication. The astonishingly rapid reception of this communication channel can be described by the large number of existing users, assessed to be as near to the billions of peoples, and it has been rising day by day [1]. This system of communication has the modest benefit of being practically prompt, spontaneous in use, and practically very cheap. The most common protocol used by the e-mail system is SMTP [2].The current email service is prone to abuse due to its relative simplicity. The methods were designed with the expectation that email users would not take advantage of the ability to transmit messages to one another. Over time, the misappropriate usages and overuse of the messaging facility has taken numerous forms. Forged emails, unsolicited emails (spam), unscrupulous activities and identification embezzlement via "phishing" emails are all examples of common misuse. Virus and worm attachments, as well as email denial-of-service assaults, are examples of abuse. All of these kinds have one common factor: they take advantage of the email system's absence of sender and recipient controls and authentication. Email does not require prior approval, and anyone can transmit a message to anyone.

In our work, we have focused on the problem of spam mails which are received in bulk by every user on a daily basis. The spam mails are the undesirable mails delivered by any unsolicited person also termed as spammer [3] with the purpose of producing wealth. Users spend the majority of their time [4] deleting spam mails from their inboxes. Many replicas of the similar message are delivered on multiple occasions, which not only costs money to an organization [5], but also frustrate the recipient. Spam mails intrude on users' inboxes, and also generate a considerable volume of unnecessary data, reducing network capacity and consumption. This study proposes a Spam Mail Detection (SMD) system for classifying email data into spam andham emails. The three major stages of spam screening are the email address, subject, and substance of the message [6].

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

The topic of the email and the body of the email are common elements in all emails. Filtering the content of a spam email can help classify it. The identification of spam mail is predicated on the notion that the spam mail's matter is not same as that of genuine or ham mail. For instance, words associated with commercial advertisements, service endorsements, dating-related content, and so on. The procedure of detecting spam emails can be alienated into 2 classes: a technique based on knowledge engineering and ML [7]. Knowledge engineering is a network-based methodology to email classification that takes into account IP (internet protocol) addresses, network addresses, and a set of predetermined rules. Although the method has yielded encouraging results, it is time demanding. For some users, the work of maintaining and changing rules is inconvenient. . Machine learning, on the other hand, does not require any rules and is more productive than expert systems [8].The email is classified by the classification algorithm based on its content and other criteria. The procedure of feature extraction and selection is critical for the majority of classification issues. In the classification process, features are quite important. For feature extraction, a CFS [9] approach is adopted in this paper. For effective classification results, the CFS technique extracts the best characteristics from a pool of features. A unique mix bagged procedure is incorporated in the recommended spam mail detection (SMD) system to address the shortcomings of the conventional model. The basic procedure of email filtering is depicted in Fig1.

II. RELATED WORK

Knowledge engineering (KE) and machine learning are the most prevalent spam mail filtering methods. In KE based methods, we specified a collection of rules to identify spam emails as well as normal e-mails. This collection of rules can be specified by the consumer, or by the software dealer which bestows a specific rule cantered spam straining program. In order to be successful, these rules should be continuously renewed and preserved, which is always not possible or it canbe inconvenient for many persons. After many successful Machine learning (ML) approach, it was concluded that Machine learning methods are better than knowledge engineering methods; as these do not need to specify any instructions [10].Rather, in machine learning methods, there are a group of training models which are a group of pre specified e-mail posts. Next, an explicit algorithm has been utilized to acquire the classification rules from these e-mail posts. In recent times, various Machine learning methods (e.g. artificial immune system, J48 classifier, SVM, Naïve Bayes and Neural Networks) have been extensively reviewed and can be used for e-mail filtering. This work emphasized on the analysis filtering of spam mail and protecting normal e-mails.A filtering method is needed for the categorization of e-mail into ham or spam. The authors [11] have proposed a spam email filtration method through distinctive features selection process to categorize the emails into normal and spam. After the pre-processing of the dataset (English and Malay email) descriptions, they have used TF-IDF and rough set theoretical technique. Next, they have applied machine learning approach for the categorization and obtain good performance.

Authors in [12] have recommended a new ML centred technique for the categorization of email data. The implementation of the algorithmic contains KNN and Naïve Bayes algorithms andpresenting applicable outcomes in case of application of algorithms on pre-processed dataset. Additionally, authors in [13] have planned an ontology centred email filtering approach. The used dataset has been categorized via J48 decision tree centered method. In pursuance of testing the outcomes received after the categorization, a RDF linguistic centered ontology was produced by Jena.

Authors in [14] have used FFN network focused technique to recognize the spam emails and to precise the outcomes. They have used Krill Herd algorithm to train the dataset which is uniformly allocated into two splits for the training and testing reasons.Authors in [15] have projected an inclusive analysis of latest growths in the uses of ML processes for Spam filtering. They have emphasized on both textual as well as image centered methods. Rather than take into account Spam filtering as a customary classification issue, they have highlighted the significance of reflecting particular features of the problem, especially perception meaning for the designing of different filters. They have analysed the problem faced in revising a classifier bestowing to the bag-of-words illustration and a key transformation among ancient naive Bayes models. Authors in [16] have provided a detailed overview of spam mail identification techniques.

III. SPAM MAIL FILTRATION PROCESS

In the classification process, some intermediate steps are involved which are done in the pre-processing and post-processing. When an email file enters into the spam filter, some steps like Tokenization, Lemmatization and Stop word removal are used in extracting most informative features from bunch of email files. These steps are falling into the pre-processing. Further, representation step is engaged in finding a structure between the email files and associated feature obtained from pre-processing. This structure basically represents the relationship between features and email files that are used to train the classifier mask image and then find boundary of target region. This step is

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

done in post processing (Figure 2).The task of selecting the appropriate class label for a given input is known as classification. Each input is processed separately from all other inputs in basic classification tasks, and the set of labels is defined in advance. The following are some instances of categorization tasks:Identifying whether or not an email is spam. Choosing a news article's topic from a predetermined list of options such as "sports," "technology," and "politics." determining whether the term bank refers to a river bank, a financial institution, the act of tilting to the side, or the process of depositing something in a financial institution. An email contains two parts, a header and body.

IV. CONFIGURATION OF AN EMAIL SPAM CLASSIFIER

The intricacy presented by spammers has made it problematic to discern between Ham (genuine emails) and Spam, which has sparked interest in study into Spam Classification (spontaneous emails). It is a war fought by both sides. When a new filtering scheme is discovered, spammers introduce different attacks to counter them. Figure 3 gives a structure of E-mail classifier.

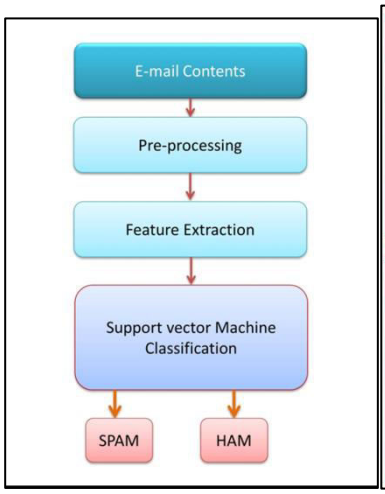

Fig. 1: Fundamental Approach of Spam Mail Filtration

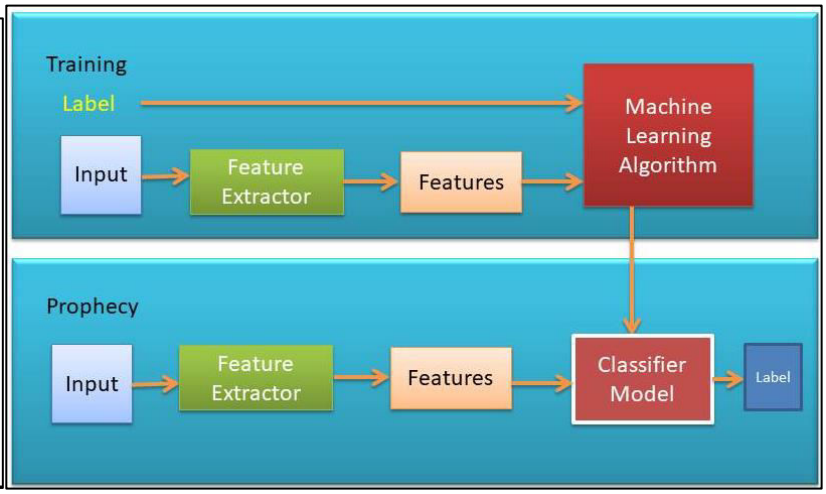

Fig 2: Relationship between features and email files to train the classifier

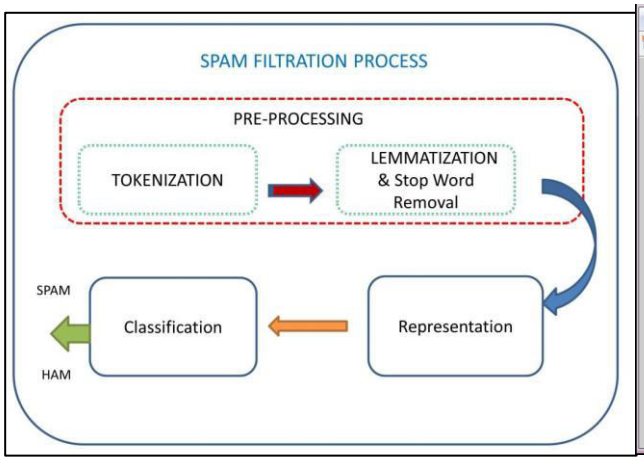

Fig 3: E-mail Classifier

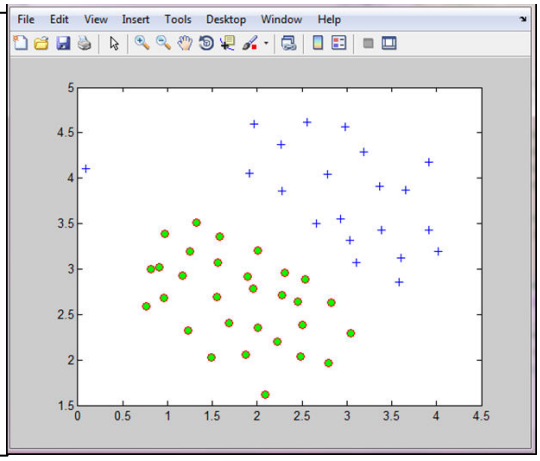

Fig 4: Trained Data of 2D Example1

V. PROPOSED WORK AND RESEARCH METHODOLOGY

We have used SpamAssassin corpus for our research experiments. This corpus contains some older and fresh Spams generated by sources other than spam traps. For this study, 2300 Spam mail files were designated from the full corpus. In addition, this compilation has certain simple and complex Genuine (Ham) files that are integrated to produce 2300 Ham email files. In this corpus, easy ham emails might be recognized and categorized but complex ham emails are

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

difficult to be measured. Complex ham contains certain attacks like use of HTML, uncommon HTML mark-up, colored script, axiom on the characteristics, done by spammers that make it very critical to distinguish emails.

The set of data is separated into two parts: a training and test set, each containing 702 and 260 emails, divided evenly between spam and ham mails. Text cleansing is the initial stage in any text mining challenge, where we eliminate terms from the document that may not add to the information we wish to retrieve. Emails may involve a large amount of undesired characters, such as punctuation marks, stop words, numbers, and so on, which may make it difficult to spot spam.

a) *Elimination of stop words* – Stop words such "and," "the," "of," and "other" are frequent in all English phrases and have no value in determining whether an email is spam or authentic, thus they have been detached from the emails.

b) *Lemmatization* – It is the procedure of combining a word's various inflected forms so that they can be investigated as a single item. The words "include," "includes," and "included" are all characterized as "include." In contrast to stemming, lemmatization preserves the context of the sentence.Non-words such as punctuation marks and special characters must still be removed from the postal documents. It can be done in a variety of ways. We will remove such words after constructing a dictionary, which is a very convenient way because you only have to remove such words once when you have a dictionary.

VI. CREATING WORD DICTIONARY

The topic of the email appears on the first line, while the email body appears on the third line. To detect spam emails, we will just use text analytics on the matter. To begin, we must compile a lexicon of words and their occurrence. A training collection of 700 emails is used for this purpose. Once the dictionary has been built, to eliminate the non-words we discussed in step 1, we can add a few lines of code to the above function.

VII. FEATURE EXTRACTION PROCESS

After the vocabulary is complete, we can retrieve a 3000 dimension word count vector (our feature here) for all emails in the training group. Every word count vector represents the occurrence of 3000 words in the training file. Actually, you've probably guessed that the majority of them will be zero. Let's look at an example. Assume we have a vocabulary of 500 words. Every word count vector in the training, the frequency of 500 dictionary words is recorded in this file.

VIII. TRAINING THE CLASSIFIERS

Support Vector Machines (SVM) classifier has been trained. SVMs are supervised binary classifiers that come in handy when there are a lot of features to consider. SVM is used to extract a subset of training data called support vectors from the rest of the data (boundary of separating hyper-plane). The SVM model's decision function, which predicts the test data's class using support vectors and a kernel trick, is based on support vectors. We can assess the models' performance on a test-set after the classifiers have been trained. With the trained SVM model, for each mail in the test set, we retrieve the word count vector and predict its category (ham or spam).A SVM's main idea is to create a nonlinear kernel function to transform data from the input space to a potentially high feature space, then generalize the optimum hyper-plane with thebiggest difference between the 2 categories. A linear SVM classifier is provided a T-element training setset (xi,yi):xiRD,yi-1,1,i=1,...,T.:

$$F(x) = \sum_i a_i y^i x \cdot x^i + b = x \cdot \sum_i a_i y^i x^i + b = x \cdot w + b \quad (1)$$

The multipliers ai have non-zero values in only a small subset of the training set, known as the support set, and its constituents, the support vectors. Because they build a hypothesis by means of a subset of the data comprising the most revealing outlines, SVMs are good contenders for vigorous or discerning sampling methodologies that look for these trends. A decent heuristic algorithm will request labels for the patterns that will form support vectors if the data is originally unlabeled. Active selection would be especially useful in instances when the data labelling process is costly or the dataset is vast and unlabeled.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

IX. ALGORITHM

1) Initialization
Seed SVM classifier with a certain instances of all classes (spam and ham)
Train a preliminary SVM filters
2) Learning
- Classify xi
- Query the true label of xi
- If the filter's prediction is wrong, retrain SVM filters based on SVi + xi
3) Concluding
Recurrence until xn is over

This strategy should produce a model that is identical to or comparable to what would have been achieved if all of the data had been used to train. The rationale for this is that the SVMs approach will save the key class boundary information shown so far as support vectors in a very compressed manner, that will help deduce the new idea suitably in the next iteration.

X. WORKING OF SUPPORT VECTOR MACHINES

We train a SVM first, and then cross-check the classifier, as we would with any supervised learning model. Use the machine that has been trained to classify (predict) fresh data. Furthermore, we can utilize several SVM kernel functions to get adequate predicted accuracy, but we should modify the constraints of the kernel functions.

XI. LINEAR BOUNDARY SEPARATION (ASSUME DATA SET ONE)

We started with the 2Dimension sample dataset, which is parted by a linear border. Figure 4 shows how the training data was plotted. The locations of good examples (shown by a +) and negative examples (indicated by a –) have created a clear distinction (signified by o). This is the difference between these two SVM decision boundaries. We can utilize dissimilar values of parameter C with SVM. When the C value is high, it suggests that SVM is attempting to categories all cases with high precision. C is equivalent to the logistic regression regularization parameter ().When C=1, the SVM places the decision border at the intersection of the two datasets, misclassifying the far left data point (Figure 5). If C=100, the SVM classifies, all examples with high precision, however, the decision boundary that does not seems to be real perfect for the data.

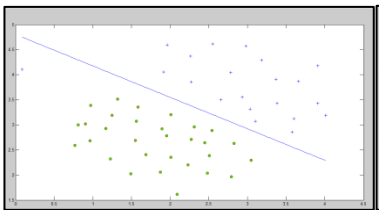

Fig 5: Decision Boundary of SVM with C=1

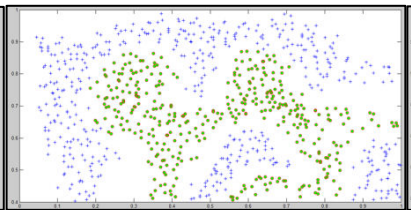

Fig 6: Trained data set two

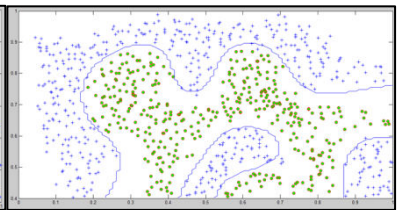

Fig 7: SVM and Gaussian Kernel based Decision Boundary for data set two

XII. SVM USING GAUSSIAN KERNELS

In this experimentation, we have used SVMs to accomplish non-linear classification. Here the used dataset is non-linear separable. We should implement a Gaussian kernel, to set non-linear decision boundaries using the SVM. Let us consider the Gaussian kernel as a similarity function that computes the distance amid a couple of instances, (x(i); x(j)). Moreover, the Gaussian kernel has been parameterized with a bandwidth value σ, which is used to decide the decreasing speed of resemblance metric (to 0), because the instances are not very near.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

XIII. LINEAR BOUNDARY SEPARATION (DATA SET TWO)

Now the SVM classifier load and plot data set two and outcomes are shown in figure 6. This time there is no linear separation boundary between +ive and –ive examples of data set two. Though, if we implement the Gaussian kernel with the SVM, we can have a non-linear decision boundary which can enact sensibly fine for the dataset two. With the help of Gaussian kernel function, we have a clear decision boundary as revealed in figure 7. The decision boundary is capable of separating maximum of the +ive and -ive examples properly and tracks the contours of the dataset properly.

XIV. CLASSIFICATION OF SPAM MAILS

Most of the present E-mail services provide a system or filter which can separate spam mails from Ham mails. Here we have used SVM classifier to make our own spam filter. We have trained a classifier which classifies Ham (y =0) and spam (y=1) mails. We have converted each mail into a feature vector $x \in R^n$ .We have used Dataset which is subset of public corpus SpamAssassin. Since we have built a content based filter, we will use only the body part of the e-mail for feature extraction and classification.

(a) *Preprocessing E-mail*

Let we have a sample e-mail as shown in figure 8. The preceding scenario includes a URL, an email address (in last), numbers, and monetary values. Whereas various emails will include identical categories of data (like numbers, URLs, or email addresses) and the particular objects (like the precise URL or dollar amount) in virtually every email will be unique. As a result, one common way for processing emails is to standardize them "these values, so that all URLs, integers, and other variables are handled the same. For example, we might use the unique string httpaddr to substitute all URLs in the email "to indicate the presence of a URL.This allows the spam classifier to classify messages based on whether or not any URLs were present, instead of a precise URL was present. Because spammers frequently randomize URLs, the chances of seeing the same URL in a fresh member of spam are quite low. The following are the steps for normalizing and processing e-mails.

i. **Lower-casing:** The whole email has been transformed into lower case, so that capitalization can be overlooked (e.g., SmAlL is considered similar to small).
ii. **Shedding HTML:** All HTML tags have been detached from the emails. Most of the emails have formatting of HTML; we have removed all the HTML tags, so that only matter is present in the body.
iii. **Standardizing URLs:** All URLs have been substituted with the text \httpaddr".
iv. **Standardizing Email Addresses:** All email addresses have been substituted with the text \emailaddr".
v. **Normalizing Numbers:** All numbers have been substituted with the text \number".
vi. **Normalizing Dollars**: All ($) signs have been substituted by the text \dollar".
vii. **Word Stemming**: The stemmed form of words is used. Like, \discount", \discounts", \discounted" and \discounting" are all substituted with \discount".

viii. Elimination of non-words: Non-words and punctuation are separated. Tabs, newlines, and spaces have all been cut down to a single space character. After pre-processing we have resultant e-mail contents (figure 9).

(b) *Vocabulary List*

For each email, after the task of preprocessing the emails, we have a catalog of words. Furthermore, we will select the words which we would like to use in our classifier and which we would want to cancel out. To make the vocabulary list, we have used most frequently used words. An example of vocabulary list is as follows (figure 10). Our vocabulary list was designated through selection of all words in the spam corpus that appear at least 100 times, consequential in a list of 1900 words. Generally, Commonly, a vocabulary list of 10,000 to50,000 words has been employed. We can now map each word in the precompiled emails into a list of word indices that carries the index of the word in the vocabulary list by usingvocabulary list. The word "anyone" was first normalized to "anyon" in the sample email, and then mapped to index 86 in the vocabulary list (highlighted).

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

(c) *Features Extraction from Emails*

Next we have implemented the feature extraction which changes each email into a vector in Rn. The feature $xi \in \{0|1\}$; for an email links to whether the email contains the i-th word from the dictionary. That is, if the i-th word appears in the email, xi = 1; if the i-th word does not appear in the email, xi = 0.

Fig 8: Sample E-mail

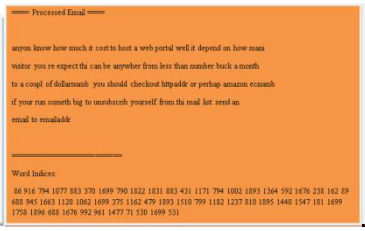

Fig 9: Pre-processed E-mail

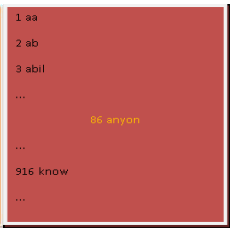

Fig 10: Vocabulary list

XV. TRAINING SVM FOR SPAM CLASSIFICATION

We have loaded a preprocess dataset which is used to train the SVM classifier. Each e-mail has been processed, feature vector was created and features were extracted. After the training finishes, the classifier has a training precision of almost 99.8 percent and a test correctness of nearly 98.5 percent, as can be seen. We can look at the parameters to determine which phrases the spam classifier thinks are the most analytical of spam to have a better understanding of how it works. The classifier's parameters with the highest positive values are displayed, along with the words that correlate to them. As a result, an email including phrases like "guarantee," "delete," "dollar," and "price" is likely to be categorized as spam. See figure 11. So finally our result conclude that processed email is a spam email (figure 12).

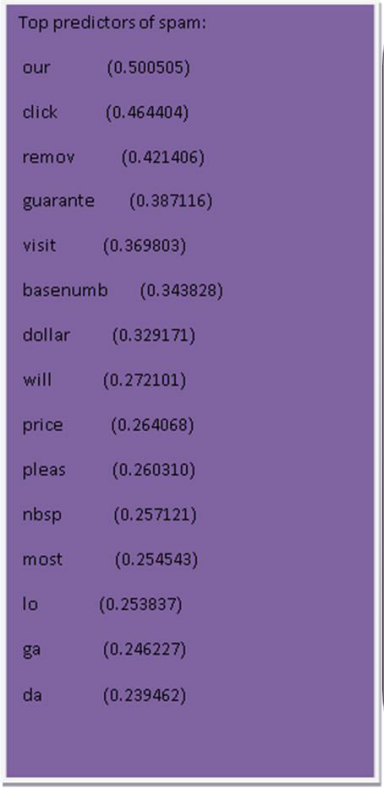

Fig 11: Spam Prediction

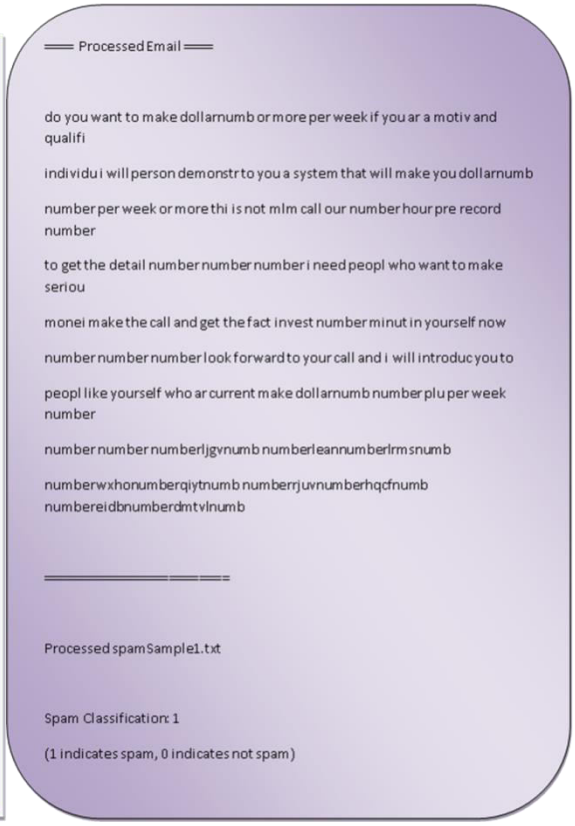

Fig 12:Spam Mail Identification

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 8, August 2021 ||

| DOI:10.15680/IJIRSET.2021.1008114 |

XVI. CONCLUSION

In this work, we have reviewed certain well known ML techniques and their suitability to the issue of spam e-mail categorization.In this research, body part of the email is to be considered for the experiment because it has widely accepted for the content based machine learning methods. In the pre-processing, several feature selection and feature search methods have been described for selecting most informative features. To identify most informative features, comparative studies have been performed separately on feature selection and feature search methods and discussed in the subsequent chapters. We have preprocessed the e-mail and constructed the vocabulary list. We have matched the word indices with this list. After than we performed feature selection and concluded the feature vector. After this we have Trained the SVM classifier. We have created a set of words. We check a word belong to spam or Ham. Our results give an accuracy of 98 %.

REFERENCES

[1] Email mailboxes to increase to 1.2 billion worldwide by 2005. Technical Report DC #W25335 7.
[2] https://tools.ietf.org/html/rfc821.
[3] C. Yang, R. Harkreader and G. Gu, “Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers", In: Recent Advances in Intrusion Detection, Springer Berlin/Heidelberg, pp.318-337, 2011.
[4] S. Kumar, and S. Arumugam, “A Probabilistic Neural Network Based Classification of Spam Mails Using Particle Swarm Optimization Feature Selection”, Middle-East Journal of Scientific Research, Vol.23, number 5, 874-879, 2015.
[5] N. P.DíAz, D.R.OrdáS, F.Riverola, and J.R. MéNdez, “SDAI: An integral evaluation methodology for content-based spam filtering models, Expert Systems with Applications”, Vol.39, number 16, pp.12487-12500, 2012.
[6] A. K Sharma, S. K Prajapat and M.Aslam, “A Comparative Study Between Naive Bayes and Neural Network (MLP) Classifier for Spam Email Detection:, In: IJCA Proceedings on National Seminar on Recent Advances in Wireless Networks and Communications, Foundation of Computer Science (FCS), pp.12-16, 2014.
[7] W. Ma, D. Tran and D. Sharma, “A novel spam email detection system based on negative selection”, In: Proc. of Fourth International Conference on Computer Sciences and Convergence Information Technology, ICCIT'09, Seoul, Korea, pp.987-992, 2009.
[8] T.S. Guzella, and W.M. Caminhas, “A review of machine learning approaches to spam filtering, Expert Systems with Applications”, Vol.36, number 7, pp.10206-10222, 2009.
[9] M. Mohamad, and A. Selamat, “An evaluation on the efficiency of hybrid feature selection in spam email classification”, In: Proc. of 2015 International Conference on Computer, Communications, and Control Technology (I4CT), Kuching, Sarawak, Malaysia, pp.227-231, 2015.
[10] T.S. Guzella, and W.M. Caminhas, “A review of machine learning approaches to Spam filtering”, Expert System, 2009.
[11] M. Mohamad, and A. Selamat, “An evaluation on the efficiency of hybrid feature selection in spam email classification”, In: Proc. of 2015 International Conference on Computer, Communications, and Control Technology (I4CT), Kuching, Sarawak, Malaysia, pp.227-231, 2015.
[12] A. Harisinghaney, A. Dixit, S. Gupta, and A. Arora, “Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm”, In: Proc. of 2014 International Conference on Optimization, Reliabilty, and Information Technology (ICROIT), Faridabad, Haryana, pp.153-155, 2014.
[13] S. Youn, and D. McLeod, “Efficient spam email filtering using adaptive ontology”, In: Proc. of Fourth International Conference on Information Technology, Las Vegas, NV, USA, pp.249-254, 2007.
[14] H. Faris, and I. Aljarah, “Optimizing feedforward neural networks using Krill Herd algorithm for e-mail spam detection”, In: Proc. of IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, pp.1-5, 2015.
[15] T. Guzella, and W.Caminhas, “A review of machine learning approaches to spam filtering”, Exp System Application, vol. 36, number 7, 10206–10222, 2000.
[16] Jayant Batra, Kirti Bhatia, Rohini Sharma, ShaliniBhadola, “An Overview on Machine Learning Based Spam Mail Identification Approaches”,International Journal of Innovative Research in Computer and Communication Engineering, vol. 9, number 7, pp. 8987-8994, July 2021.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.569

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details