International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 6, September-October 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

A Deep Analysis on Prevailing Spam Mail Filteration Machine Learning Approaches

Anu[1], Ms. Preeti[2]

[1]M Tech Student, [2]Assistant Professor,
[1,2]Meri College, Sampla, Haryana, India

ABSTRACT

In this work, we have reviewed the issue of spam mail which is a big problem in the area of Internet. The growing size of uncalled mass e-mail (or spam) has produced the requirement of a dependable anti-spam filter. Now a days the Machine learning (ML) proedures are being employed to spontaneously filter the spam e-mail in an effective manner. In this work, we have reviewed some of the prevalent ML approaches (such as Rough sets, Bayesian classification, SVMs, k-NN, ANNs and Artificial immune system) and of their use fullness in the issue of spam Email taxonomy. We have provided the depictions of the procedures and the divergence of their enactment on the basis of the quantity of Spam Assassin.

KEYWORDS: *Spam e-mail, Machine learning Procedures and Filtering*

How to cite this paper: Anu | Ms. Preeti "A Deep Analysis on Prevailing Spam Mail Filteration Machine Learning Approaches" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-6, October 2020, pp.1160-1167, URL: www.ijtsrd.com/papers/ijtsrd33261.pdf

IJTSRD33261

1. INTRODUCTION

With the increase in poularity of the Internet, we have also faced new types of threats such as virus attack and unwanted profit-making mass quantity of e-mails popularly known as spam. These types of e-mails consume time, communication bandwidth and storage space. In spite of so many firewalls, this issue has been growing day by day. According to some prevalanet survey, less than 50 % of all emails are spam which cost the internet users a huge price per year. In this scenario, an automated e-mail filteration mechanism can be the most suitable technique for defying spam at any time. Therefore there is a tuff fight among spammers and anti-spammers. In previous times, the spam was filtered through detction and blockage of address which send spam e-mails or through filtaration of messages having specific keywords or advertisements. In order to mislead filteration techniques, the spammers have strated to employ numerous tricks to disable the filtering techniques such as use of different sender IP addresses or to affix / suffix haphazard symbols to the beginning /end of the subject line of the message [1].

The most popular methods of spam mail filteration are knowledge engineering (KE) and machine learning (ML). In KE based methods, we specifie a collection of rules to identify spam emails as well as normal e-mails. This collection of rules can be specified by the user, or by the software vendor which bestows a specific rule centerd spam-filtering software. In order to be successful, these rules should be continuously renewed and preserved, which is always not possible or it can be 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 [2]. Rather, in machine learning methods, there are a set of training samples 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, support vector machines (SVM), Naïve Bayes and Neural Networks) have been extensively reviewed and can be used for e-mail filtering. This work emphasized on the analysis fn filtering of spam mail and protecting normal e-mails.

This paper has been planned as follows: section 1 contains introduction of the paper, section 2 synopsize the associated work considering several ML procedures, section 3 provides a common theoretic explanation on the ML approaches and section 4 present conclusion of the paper.

2. LITERATURE SURVEY

A filtering method is needed for the categorization of e-mail into ham or spam. The authors [3] have proposed a spam email filteration 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 theoritical technique. Next, they have applied machine learning approach for the categorization and obtain good performance.

Authors inn [4] have proposed a new ML based method for the categorization of email data [4]. The implementation of the algorithmic contains KNN and Naïve Bayes algorithms and presenting applicable outcomes in case of application of algorithms on pre-processed dataset.

Additionally, authors in [5] have planned an ontology centered 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 [6] 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.

AUTHOR	FINDINGS/ METHODOLOGY	RESEARCH GAP
Guzella T, Caminhas W. M.- 2000 [7]	In this paper, the authors 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, specially perception meaning for the designing of different filters.	They have analyzed the problem faced in revising a classifier bestowing to the bag-of-words illustration and a key transformation among ancient naive Bayes models. They have concluded the significant development in this field and pointed out different characteristics which are yet to be discovered, particularly in additional reasonable estimation situations.
Levent Özgür, Tunga Güngör, and Fikret Gürgen – 2004 [8]	The authors have proposed vigorous anti-spam cleaning approaches for different vernaculars in common and specifically for Turkish, founded on **ANN** and **Bayesian classifiers**. The procedures are flexible and have twin modules. The major module works on the morphology and the next module categorizes the mails through the roots. They have considered single layer and multi layer perceptron (MLP) (ANN tructures), and the inputs to the networks are controlled by binary and probabilily proototypes. They have used three approaches binary, probabilistic, and advance probabilistic for Bayesian classification prototypes.	The anti-spam filtering mechanism which has been defined in this work consists of two components: MM and LM. They have designed a turkish morphological invetigation method. A particular word in Turkish may relate to a idiom built up of numerous words so Turkish is known as agglutinative language. Due to this, the morphological examination of turkish language is more difficult than the investigation in other languages such as Hindi. In pursuance of determining the origin words that can be served as the aspects of the classification procedure, they have used the ideat of shared information. The aspect vector has been described as a set of acute words that can be employed in categorization. Initially, the candidate words (the words of the training data) has been recognised.
Enrico Blanzieri, Anton Bryl -2007, [9]	In this work, the authors have evaluated an example focused on spam filter supported by the **SVM**-NN classifier theory. They have combined the concepts of SVM and **kNN**. Initially, to mark a message the classifier discovers k nearest marked messages, and next a SVM prototype has been trained for these k samples and employed to mark the unidentified sample. The authors have compared SVM-NN with SVM and k-NN and presented the results.	To find the adjacent neighbors and the linear kernel for SVM, they have used Euclidean metric in their experiments. All messages were deliberated as an unordered list of strings disjointed by gaps. Occurrence of a particular token in a specific section of a message, specifically in the region of the header, was reflected as a binary aspect of that message. Next, the utmost recurrent characters in the training dataset were choosed and utilized. Therefore, every message has been characterized through a vector of the binary descriptions. The authors have assessed a learning centered spam filter. The proposed method has overtake SVM considerably over the minute dimensions of the feature domain.
Mawuena Glymin and Wojciech Ziarko- 2007 [10]	This work exhibits a fundamental summary of methods used for spam recognition through probability assessment table focused analytical data modelling. The main emphasis is on the presentationof the solution which joins modest methods with particular heuristics to make comprehensive rough estimates of spam and genuine e-mails via the VPRSM **rough set** technique. Experimentations were organized to	The spam recognition system working has been allocated into two stages: a training stage and a categorization stage. The rough set centered machine learning has been used in the training stage and it is used for the pool of pre-categorized e-mails. In the categorization stage, hierarchy of trained decision table was employed which forecast the choice class of receiving e-mails.

	investigate the uses of VPRSM for creating a smart worker for spam filteration. They have conducted investigation of possibility of utilizing the hierarchy of VPSRM probability tables for spam recognition filter.	For training purpose, they have collected e-mails from hotmail platform and tested the determination prototype.
Hsu Wei-Chih, Tsan-Ying Yu- 2009 [11]	The **SVM** is a dominant categorization method of data mining and has been effectively used in several physical world applications. The classification performance is effected by the selection of parameters of SVM specifically through training process. Though, selection parameter in SVM has been normally recognized by knowledge or simple grid search. They have suggested Taguchi technique for improvement of grid search and employed to enhance the SVM centered Spam Filteration prootootype. It is simple to realize the orthogonal arrays without repetition. A physical world mail dataset was chosed to show the efficiency and possibility of the process. The experimentation exhibits that the Taguchi technique may discover the operative prototype with high categorization precision and good sturdiness.	The authors have discovered the finest blend of two coefficients (C, γ) of SVM using Taguchi technique. While creating spam filteration prototype, the coefficients of SVM were considered as influence factors. After the selection of the coefficients of SVM, they verified the results of classification and equated with grid search. The SVM is a vigorous supervised learning model which is founded on the organized hazard lowering standard from arithmetical learning concept. SVM has amazing enactment for text classification process.
Ibrahim Eldesoky- 2009 [12]	**Artificial Immune System** (AIS) prototype has been stimulated from the normal immune system. The authors have proposed the AIS prototype which fulfill the spam filteration procedure. The contents of the e-mail messages has been employed for both training as well as testing procedures. The words which make the e-mail are weighted and employed in estimating the empathy amid the antibody and an antigen. To remunerate the cells that properly identify the spam e-mail, they have initiated a learning phase. Rather than using clonal assortment, negative assortment was employed for the training period and it caused improved testing enactment owing to the condensed quantity of sensors. As equated with other methods, the fraction of training group elements is very less than others and hence it confirms the improved enactment. The procedure was verified against the global Masses of spam and normal mails.	The detector or sensor is the key module of the spam immune scheme and these are known as antibodies (AB). The ABs are required to confirm the subjects of an e-mail posts. To transform an email message into an AB, initially they have computed the native weight of every word of the email. The native weights were computed in a standardized way through totaling of the existences of every word in relation to the number words in the email. The anticipated affinity assessment method is supported by native weights rather than totals of the analogous words of the two emails. Next, they have checked this affinity against a standard value to indicate the type of the email. Though, the artificial prototype has been trained for both spam and desired elements, but it discard the cells which erroneously identify new antigens through the testing progression.
Loredana F., Camelia L., Rodica P. -2010 [13]	This work represents a fresh spam recognition filter, which conglomerates numerous attributes which are prominently based on frequency of words, and usesof the **kNN** process to categorize the emails. The recommended architecture has been separated into two key modules: one for the collection of the data and for the extraction of the aspects of all messages and another for the Classification of the emails and interpretation of the results. Initially, an email has benn selected to assemble the data for analysis. They have designed spam traps to attract spammers. For feature extraction, they have anayzed the message on the basis of different features such as length of message, quantity of responses and frequency of words. An email and these mined features signify the sample of the data set. The authors have merged the tokenization technique for the contents of the message with a curbing	The authors have classified the messages with the kNN method which is supported on a group of characteristics mined from the properties and contents of the email. The trained set has been resampled to the utmost suitable size and certain class dissemination has been decided by numerous experimentations. The proposed mechanism achieves a continuous renewal of the dataset and the record of maximum commonly words that exist in the email. Moreover, they have provided a feedback choice concerning nonclassified messages and suggested to be accomplished at a fixed rate conditional to the availability of the resources.

	procedure.	
Tiago A. Almeida · Jurandy Almeida · Akebo Yamakami - 2011[14]	Here, the authors have considered the enactment of distinct term-selection methods using various separate prototypes of **Naive Bayes (NB) spam filters.** They have designed the experiments which were meticulously designed to confirm statistically precise results. Furthermore, they have performed an investigation which concern the magnitudes normally used to assess the excellence of spam filters. Lastly, they have also investigated the advatages of working the MCC as a standard of enactment.	Initially the authors have showed an enactment assessment of numerous term based approaches in decline of dimensionality for the spam filtering field by classifiers according to the Bayesian decision scheme. They have accomplished the assessment of the enactment attained by dissimilar Naïve Bayes spam filters used to categorize messages from six famous, actual, public and huge e-mail datasets. They have also offered the MCC as the estimation measurement factor which offeres an extra stable estimation of the forecast than TCR, particularly if the two groups are of dissimilar dimensions.
Yuanchun Zhu and Ying Tan. - 2011 [15]	Stimulated from the living **immune system**, the authors have recommended a local concentration (LC) inspired feature mining method for spam filteration. This method is capable of extracting the location realted information through messages by converting every region of a message to an analogous LC aspect. They have implmented two schemes of the LC method by a static length and a varying length sliding windows. A generic LC model has been designed to include the LC method into the complete practice of spam filteration. Two kinds of detector groups have been producedthrough term selection approaches and a clear penchant threshold. Afterward a sliding window has been accepted to allocate the message into separate regions. Once segmentation of the Message is complete, next they have calculated the strength of detectors and measure the aspect for every native region. Conclusively, they have combined all the aspects of local areas as a aspect vector of the message.	To assess the anticipated LC systeml, they have performed numerous experimentations on 5 standards masses through the cross confirmation process. It has been proven that the LC method works well with 3 term selection techniques. When compared with the universal concentration centerd method and the existing bag-of-words technique, the LC technique has better enactment in context of both precision and measurement. Moreover, they have verified that the LC technique is vigorous for messages of different lengths.
Noemí Pérez-Díaz, David Ruano-Ordás, José R. Méndez, Juan F. G., Florentino F. – 2012 [16]	The authors have reviewed and combined the preceding methods and new substitutes for smearing the **rough set (RS)** mechanism over the spam filtering space through describing the three distinct rule implementation techniques such as MFD), LNO and LTS. Keeping in mind the aim of properly evaluating the correctness of the anticipated procedures, they precisely solve and review important queries for suitable prototype approval such as corpus assortment, pre-handling and representative problems, and distinctive exact standard processes. The experimentation conducted through numerous implementation methods for choosing suitable decision rules produced by RS. Their anticipated procedures outdid other popular anti-spam filteration procedures like SVM, Adaboost and distinct sorts of Bayes classifiers.	This work provides a complete revision about the uses of RS as key classifier for business spam filteration. They have presented and investigated different schemes via RSs along with a realistic analysis about their applications, drawbacks and advantages. After analysis, they have concluded that most of the preceding works give an idea only using corpora and an insufficient pre-handling. Keeping this in mind, the authors have carried out a different investigation for a huge, fresh and non-managed corpus which was supplied by the team of Spam Assassin. As approved by the results received, the RS-centered methods are continually an appropriate substitute of Naïve Bayes classifier, SVM and Adaboost method. The proposed MFD heuristic reaches the finest precision when matched with LNO and LTS.
Yazdan Jamshidi- 2016 [17]	For spam filtering, this work presents the usages of Interval's Number **KNN** (INKNN). Later,this method was defined as a lattice data set expansion of KNN method. A populace of spam e-mails was displayed using an IN. Then proposed classifier was utilized to differntitate spam e-mails from hams. To examine the efficiency of INKNN, they have performed wide experimentation over public	This work bestows a NN classification procedure for spam filteration founded on probability explained interval number and lattice concept. A real word benefit of lattice concept is the capability of modelling both indeterminate info anddissimilar kinds of lattice-ordered data. The suggested method is able to deal with different kind of data. It is able to manage both locations and intermissions. There is a fast process

	SpamAssassin corpus. The outcomes presents that the INKNN is capable of achieving the advanced enactment as comapared to other previous advanced ML methods.	of learning in the suggested method and hence it can be used in different areas where the data is so enormous that the examination process becomes time devouring. The key benefit of using interaval number is the lodging of grainy information. The experiment outcomes validate the efficiency of our anticipated prototype.
Ali Shafigh Aski, Navid Khalilzadeh Sourati– 2016 [18]	This work defines 3 ML procedures to separate spam from hams with less error rates and extreme competence through a multilayer perceptron prototype. Numerous extensively employed methods are decision tree (DT) classifier, multilayer perceptron (MP) and NB classifier. These methods are useful for training datasets either in the type of spam or hams. At last, they have discussed the results of measured methods are inspected in context of the planned prototype.	The proposed method is based on guidelines for appropriate recording in context of the competence of instructions. The instructions were delivered in three types: a) information analysis of email header, b) tallying of keywords, and c) key content of the message.
Priti S., Uma B. – 2018 [19]	This work is meant to suggest a ML based fused bagging method by applying the two ML procedures: NB and J48 (DT)) for the spam email recognition. Here, the dataset has been shared into dissimilar groups and provided as input to all algorithms. They have conducted various experiments and the obtained results were compared in context of accuracy, memory, exactness, f-measurement, true and false negative/positive rate. One experiment was conducted using Naïve Bayes and another using J48 methods. One more experiment was conducted through hybrid bagged method. The complete precision of 87.5% attained by the hybrid bagged method of spam mail detection system.	On the basis of the idea of entropy, the J48 method is a multcast decision tree classifier. It forms decision trees of the training dataset. The DT is created through J48 and relies on the training data attribute values for the categorization of the fresh data value. It adhere to the idea of dividing the data into numerous groups, every aspect quality of data can be utilized to make a decision. The technique works iteratively till every data attribute has been managed and classified.
Abdul J. S., Asif K., Bharanidharan S., Sami A., Krishnan K., Mirjam J. and Friso D. – 2019 [20]	Spam mails are also popular as non-self/ unwanted moneymaking or malevolent mails. These mails are composed and forwarded to disturb either a person or a company or a unit of peoples. Other than promotion, the mails may have associations with phishing or malware hosting websites and can sneak private info. In this work, the authors have studied the efficacy of employing a NSA process for inconsistency recognition smeared to spam filteration. The proposed method has an extreme enactment and a low false recognition rate.	The proposed model is trained by forming a memory of the previous conduct of spam emails. It forbids the similar type of conduct for future incoming messages, because the model has been trained against a certain conduct. This procedure is known as Negative Selection (NS). The strategy of the NSA has been founded on the self and nonself discernment conduct of the mammalian learned immune system. The aim is to develop a prototype of irregularities, variations, or unaware data through production of patterns which do not relate to or equal to a prevailing body of accessible patterns. The NSA determines gap among regularity and irregularity by having information of self and non-self conduct.

TABLE: 01 LETRATURE REVIEW / GAP

3. CLASSIFICATION OF MACHINE LEARNING IN E-MAIL

The classification task of E-mails is generally allocated into numerous sub-tasks. Initially the collection and demonstration of Data are most problemsatic, next, feature selection and reduction of e-mail try to lessen the size for the residual stages of the task.

Naïve Bayes classifier technique: Its works on the dependent incidence and the likelihood of an incidence happening in the upcoming time which may be recognised from the preceding occuring of the analogous incidence. This method may be employed to categorize spam e-mails. Here the likelihoods of words perform the key rule. If any words take place frequently in spam but not in normal e-mail, then this inward e-mail is perhaps a spam. Figure 1 shows the working of classifier method.

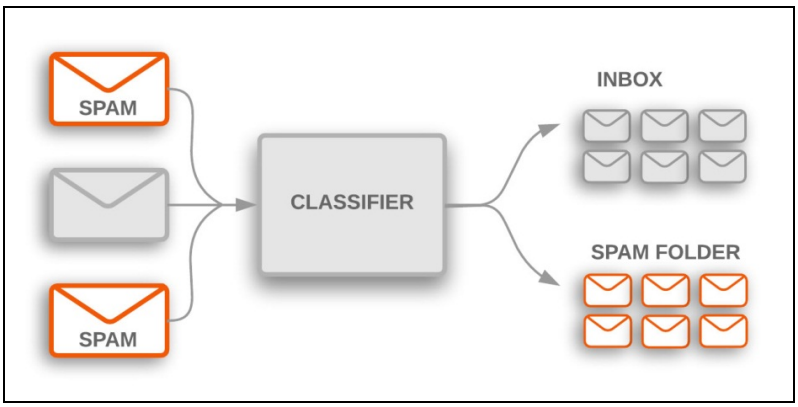

Fig 1: An instance of Naïve Bayes classifier

K-NN classifier technique: It is measured as an instance centered classifier. Here comparison is done over training documents instead of an specific class illustration, like the class profiles are employed by other classifiers. It means actually there is no training stage. For the classification of a fresh document, the k most identical documents (neighbours) are located and if a sufficient big share of them have been allocated to a definite class, the fesh document has also been allocated to this class. Moreover, using customary indexing approaches cause the discovery process of the nearest neighbours very fast. To dtermine the class of the message (spam or ham), message which are close to the class are searched quickly. Figure 2 shows the kNN based system classification [13].

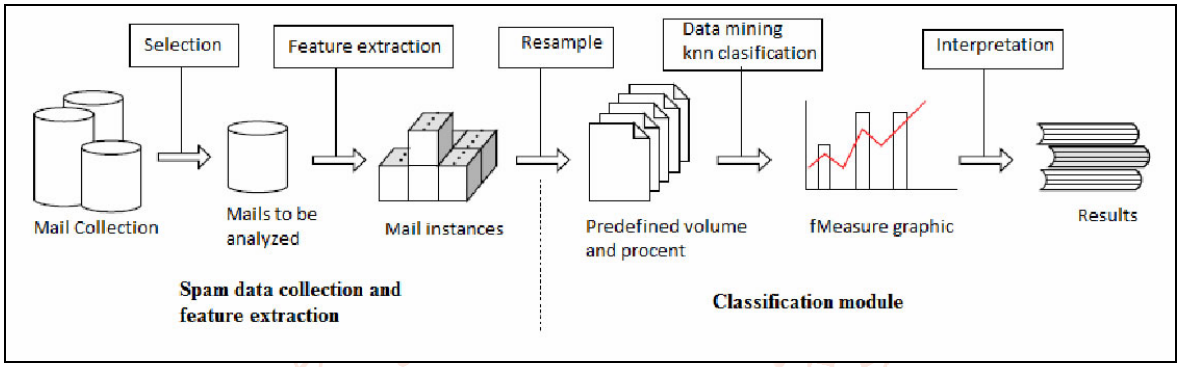

Fig 2: KNN based spam mail filteration

Artificial Neural Networks (ANN) classifier technique:
It is also recognized as Neural Network (NN) and it is a computation prototype inspired by biological NN. It includes an interrelated group of artificial neurons. An ANN is a flexible system which modifies its configuration as per the information flow within the artificial network throughout a learning stage. The ANN is founded on the law of learning by instances. Moreover, there are two standard types of the NNs, perceptron and the MP. Figure 3 shows the working of ANN for spam mail filteration[20].

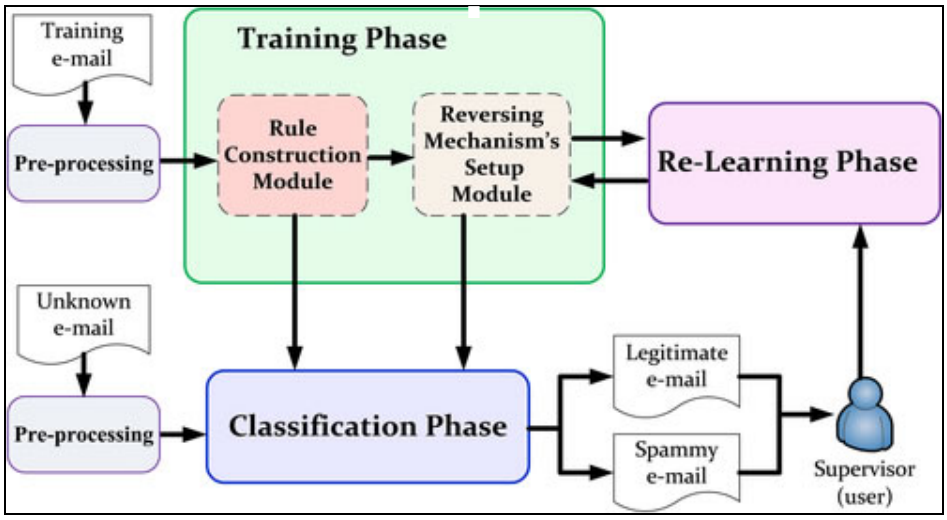

Fig 3: working of ANN for spam mail detection

SVM classifier technique: It is founded on the idea of determination planes that describe its limitations. A determination plane is splits among a group of entities having dissimilar groups participations. The SVM patterning procedure discovers an optimum hyper plane with the highest gap to split two groups. It needs resolving the optimization problem. Working of SVM is shown in figure 4 [21].

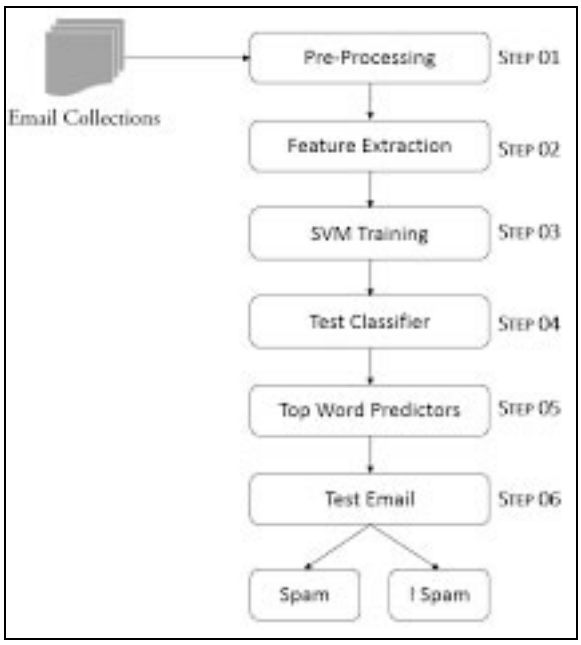

Fig 4: SVM classification of e-mail.

Artificial Immune System (AIS) classifier technique:

The normal immune system protects the person against dangerous illnesses and viruses. It is able to recognize and eliminate any different cell or particle. To accomplish its job, the immune organism has grown over complicated pattern identification and reply procedures follow several differential routes.

The fundamental task of natural immune organism is to differentiate among self and injurious non-self components.

The pathogens are the injurious non-self components of specific concern. Consequently the AIS can distinguish between self of legitimate email and a non-self of illegitimate email. Its working is shown in figure 5 [12].

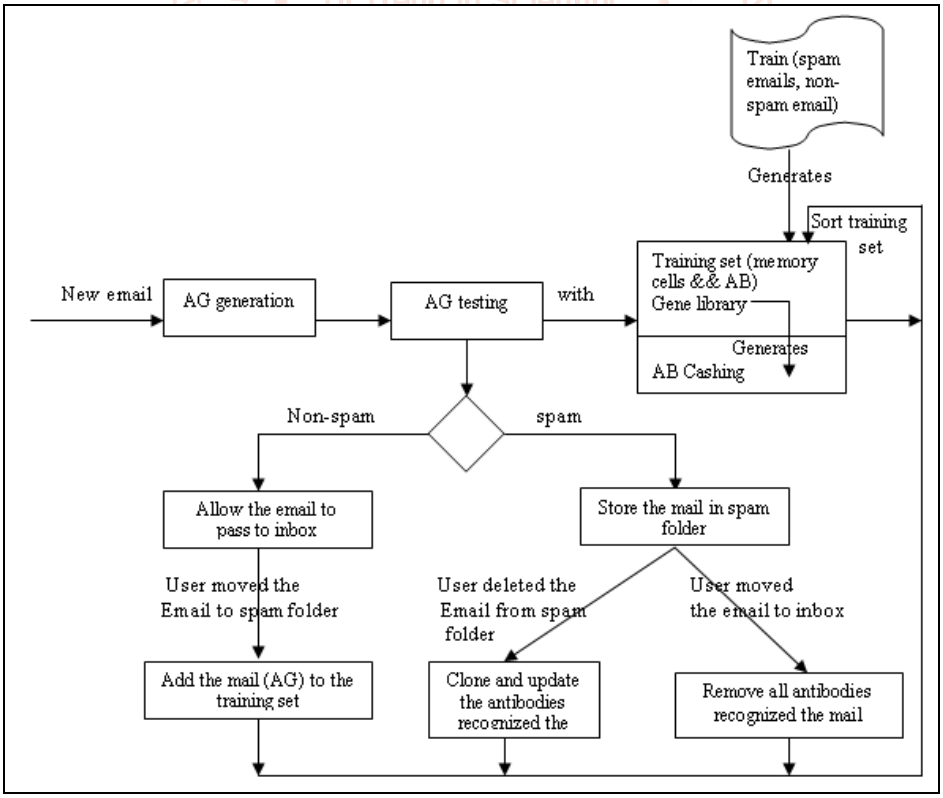

Fig 5: working of Artificial Immune System based spam filteration

Rough sets (RS) classifier technique:

these sets are capable of computing the diminutions of information schemes which may have certain characteristics that are unrelated to the goal idea (means: determination attribute), and certain surplus attributes. Diminution is required to produce modest valuable in formations from itself. It is a least subset of provision attributes in relevance to determination attributes. The RS structure works as follows: With the the received e-mails, initialy, we should choose the maximum suitable attributes to employ for classification. Then the input dataset is altered into a decision structure and splitted into the training and the

testing datasets. A classifier will be persuaded from the training dataset and smeared to the testing dataset to attain enactment assessment.

4. CONCLUSION

In this work, we have reviewed certain well konwn machine learning techniques and their suitability to the issue of spam e-mail categorization. We have presented the research gaps between the machine learning approaches towards spam e-mail filteration. We have presented brief descriptions of the popular machine learning methods like Naïve Bayes classifie, KNN, ANN, rough sets, and artificial immune systems. In context of precision, we conclude that the Naïve bayes and rough sets approaches are very satisfying in enactment amid the other approaches. There is a need of extra research to increase the enactment of the Naïve bayes and Artificial immune system either by amalgam system or by resolution of the feature dependence problem.

5. REFERENCES

[1] L. Cormack, Gordon. Smucker, Mark. Clarke, Charles "Efficient and effective spam filtering andre-ranking for large web datasets" Information Retrieval, Springer Netherlands. January 2011, pp. 1-24.

[2] Guzella, T. S. and Caminhas, W. M."A review of machine learning approaches to Spam filtering." Expert Syst. Appl., 2009.

[3] 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.

[4] 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, India, 2014.

[5] 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.

[6] H. Faris, and I. Aljarah, "Optimizing feed forward 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

[7] Guzella T, Caminhas W, (2000) A review of machine learning approaches to spam filtering. Exp Syst Appl, vol. 36, issue 7, pp.10206–10222, 2000.

[8] Levent Özgür, Tunga Güngör, and Fikret Gürgen, "Spam Mail Detection Using Artificial Neural Network and Bayesian Filter", in International Conference on Intelligent Data Engineering and Automated Learning, pp. 505-510, 2004.

[9] Enrico Blanzieri and Anton Bryl, "Instance-Based Spam Filtering Using SVM nearest Neighbor Classifier", American Association for Artificial Intelligence, www.aaai.org, pp. 441-442, 2007.

[10] Mawuena Glymin and Wojciech Ziarko, "Rough Set Approach to Spam Filter Learning", in RSEISP '07: Proceedings of the international conference on Rough Sets and Intelligent Systems Paradigms, pp. 350–359, june 2007.

[11] Hsu Wei-Chih and Tsan-Ying Yu, "E-mail Spam Filtering Using Support Vector Machines with Selection of Kernel Function Parameters", in Fourth International Conference on Innovative Computing, Information and Control, 765-767, 2009.

[12] M. H. Haggag and I. E. Fattoh, "Artificial Immune System for Spam Filtering", IJICIS, Vol.9, No.2, pp. 117-129, JULY 2009.

[13] Loredana Firte, Camelia Lemnaru and Rodica Potolea, "Spam Detection Filter using KNN Algorithm and Resampling", in International Conference on Intelligent Computer Communication and Processing (ICCP), 2010 IEEE, pp. 27-33.

[14] Tiago A. Almeida, Jurandy Almeida and Akebo Yamakami, "Spam filtering: how the dimensionality reduction affects the accuracy of Naive Bayes classifiers", J Internet Serv Appl,pp. 183-200, 2011.

[15] Yuanchun Zhu and Ying Tan, "A Local-Concentration-Based Feature Extraction Approach for Spam Filtering", IEEE Transactions on Information Forensics and Security, vol. 6, no. 2, pp.1-25,2011.

[16] Noemí P., David R., José R. M., Juan F. G. and Florentino F., "Rough sets for spam filtering: Selecting appropriate decision rules for boundary e-mail classification", Applied Soft Computing, pp. 3671–3682, 2012.

[17] Yazdan Jamshidi, "A nearest neighbour classifier based on probabilistically/possibilistically intervals' number for spam filtering", Int. J. Soft Computing and Networking, Vol. 1, No. 1, pp. 4-16, 2016.

[18] Ali Shafigh Aski and Navid Khalilzadeh Sourati, "Proposed efficient algorithm to filter spam using machine learning techniques", Pacific Science Review A: Natural Science and Engineering, pp. 145-149, 2016.

[19] Priti Sharma and Uma Bhardwaj, "Machine Learning based Spam E-Mail Detection", International Journal of Intelligent Engineering & Systems, Vol.11, No.3, 1-10, 2018.

[20] Jyh-Jian Sheu, Yin-Kai Chen, Ko-Tsung Chu, Jih-Hsin Tang and Wei-Pang Yang, "An intelligent three-phase spam filtering method based on decision tree data mining", Security and communication networks, vol. 9, issue 17, 401304026, 2016.

[21] Shradhanjali and Toran Verma, "E-Mail Spam Detection and Classification Using SVM and Feature Extraction", international Journal of Advance Research, Ideas and Innovations in Technology, vol 3, Issue 3, pp. 1491-1495.