IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

Machine Learning-Based Email Spam Filtering

Narendra Kumar Sahu and Wasim Khan
Government Women's Polytechnic College, Indore
Oriental University, Indore
naru_sahu@yahoo.com and wasukhan1982@gmail.com

Abstract: *Email spam, sometimes referred to as garbage email (unwanted email "more often than not of a business nature conveyed in mass"), is one of the real issues with today's Internet. It can cause financial harm to organizations and anger individual customers. Among all the methods created to stop spam, separation is an essential and well-known method. Two frequent uses for mail channels are the sorting of incoming emails and the removal of spam and computer malware. A less prevalent usage is evaluating employees' current email to ensure they are following applicable laws at particular companies. Additionally, clients can use a mail channel to arrange messages according to subject matter or other parameters, and then sort them into envelopes. The client has the option to introduce mail channels as standalone projects or as a part of their email program (email customer). Email clients have the option to create custom "manual" channels that automatically route mail based on selected criteria. In this work, we present a summary of the application of frequently used machine learning techniques to spam classification. Nowadays, the majority of email projects now include built-in spam separation functionality.*

Keywords: E-mail classification, Spam, Spam filtering, Machine learning, algorithms.

I. INTRODUCTION

Messages are become a common and necessary form of communication for the majority of Internet users. However, the worst aspect of email contact is spam, often known as sporadic business or mass emails. Spam is typically compared to paper junk mail. The difference is that spammers paid for the distribution of their materials, whereas junk mailers pay for extra transmission capacity, circular space, server assets, and lost efficiency. If spam continues to grow at its current rate, sooner rather than later the problem may become blatantly unmanageable.

A review indicates that more than 70% of business messages delivered in the modern day are spam [1]. Rising spam levels therefore give rise to a number of difficult issues, including stuffing letter drops for customers, drowning critical personal mail, wasting storage space and correspondence transmission speed, and denying customers the chance to remove all spam sends. While the content of spam emails is generally varied, they usually fall into one of the following categories: friend-making, business improvement, financial schemes, weight loss, explicit sexual content, etc.

- **Consistently changing –** Spam is continually changing as spam on new points develops. Likewise, spammers endeavor to make their messages as indistinct from honest to goodness email as could be expected under the circumstances and change the examples of spam to thwart the channels. [4]
- **False positives issue –** False positives are just inadmissible in this manner the prerequisites on the spam channel are extremely demanding.
- **OCR computational cost**– The OCR computational cost in content installed in pictures good with the gigantic measure of messages dealt with day by day by server-side channel. [4]

The utilization content darkening systems – Spammers are applying content clouding systems to pictures (see Fig. 2.), to make OCR frameworks ineffectual without trading off human comprehensibility. [5]

What is Spam?

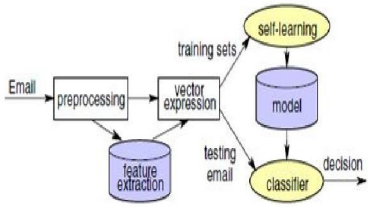
Spam is unsolicited, unwelcome emails from more interesting people delivered in large quantities to a large mailing list. These emails are usually of a business nature. Some argue that this definition should only apply in situations when the recipient of the email is not specifically selected; for example, it would not apply to emails that advertise job openings

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

or opportunities for research understudy. This definitional issue demonstrates how the definition depends on the collector and strengthens the argument for personalized spam filtering.

Fig. 1. An example of a spam mail.

1.2 Structure of an E-mail:



Not with standing the body message of an email, an email has another part called the header. The occupation of the header is to store data about the message and it contains many fields, for instance, following data about which a message has passed:

1.3 Spam Filtering:

Spam separating in Internet email can work at two levels, an individual client level or a venture level (see Figure 4). An individual client is ordinarily a man working at home and sending and getting email by means of an ISP. Such a client who wishes to distinguish and channel spam email introduces a spam separating framework on her individual PC. This framework will either interface specifically with their current mail client operator (MUA) (all the more for the most part known as the mail peruser) or all the more normally will go about as a MUA itself with full usefulness for forming and accepting email and for overseeing letter boxes.

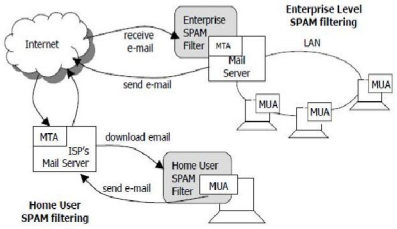


Fig. 2. Alternatives for spam filtering in Internet e mail.

Endeavor level spam sifting channels mail as it enters the inward system of a venture. The product is introduced on the mail server and connects with the mail exchange operator (MTA) characterizing messages as they are gotten.

Spam email, which is distinguished by the endeavor spam channel, will be arranged as a spam message for all clients on that system. Spam can be sifted at an individual level on a LAN too. An arranged client can channel spam locally as it is downloaded to their PC on the LAN by introducing a fitting framework.

By far most of current spam separating frameworks utilize govern based scoring strategies

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

An arrangement of guidelines is connected to a message and a score a masses in view of the tenets that are valid for the message.

Frameworks normally incorporate many standards and these guidelines should be refreshed frequently as spammers adjust substance and conduct to maintain a strategic distance from the channels. Frameworks additionally fuse list-based methods where messages from distinguished clients or areas can be consequently blocked or permitted through the channel.

On the off chance that the score for an email surpasses a limit, the email is delegated spam. Restricted learning capacities are starting to show up in frameworks, for example, Mozilla and the MacOS X Mail program yet these frameworks are still in their earliest stages. Credulous Bayes is by all accounts the procedure of decision for adding a learning ability to business spam sifting frameworks.

The design of spam separating is appeared in Fig. 5. Right off the bat, the model will gather singular client messages which are considered as both spam and authentic email. Subsequent to gathering the messages the underlying change process will start.

This model incorporates introductory change, the UI, highlight extraction and determination, email information grouping, and analyzer area.

Machine learning calculations are utilized finally to prepare and test whether the requested email is spam or honest to goodness.

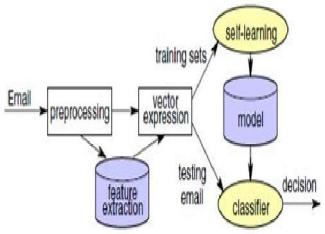


Fig.3. the process of spam filtering

II. SPAM TECHNIQUES

If an advertiser has a database with the names, addresses, and phone numbers of potential customers, they can pay to have that database cross-referenced with another database that has email addresses. At that time, the organization can send emails to those who haven't requested them, which may include those who have consciously chosen not to provide their email address. [6]

2.1. Image spam:

Picture spam is a jumbling strategy in which the content of the message is put away as a GIF or JPEG picture and shown in the email. This keeps content based spam channels from distinguishing and blocking spam messages. Picture spam was allegedly utilized as a part of the mid 2000s to publicize "pump and dump" stocks.[7] Frequently, picture spam contains counter-intuitive, PC produced content which essentially disturbs the peruser.In any case, new innovation in a few projects attempts to peruse the pictures by endeavoring to discover message in these pictures. They are not extremely precise, and some of the time sift through guiltless pictures of items like a crate that has words on it.

A more up to date method, in any case, is to utilize a vivified GIF picture that does not contain clear content in its underlying edge, or to bend the states of letters in the picture (as in CAPTCHA) to maintain a strategic distance from identification by OCR devices.

2.2. Blank spam:

Clear spam will be spam without a payload commercial. Frequently the message body is missing out and out, and the title. Still, it fits the meaning of spam on account of its temperament as mass and spontaneous email.

Clear spam might be begun in various ways, either deliberate or inadvertently:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

Blank spam can have been sent in a catalog reap assault, a type of word reference assault for social affair legitimate locations from an email specialist co-op. Since the objective in such an assault is to utilize the skips to separate invalid locations from the substantial ones, spammers may forgo most components of the header and the whole message body, and still finish their objectives.

Blank spam may likewise happen when a spammer overlooks or generally neglects to include the payload when he or she sets up the spam run.

Often clear spam headers seem truncated, recommending that PC glitches may have added to this issue—from ineffectively composed spam programming to breaking down transfer servers, or any issues that may truncate header lines from the message body.

Some spam may give off an impression of being clear when in reality it is most certainly not. A case of this is the VBS.Davinia. B email worm [8] which engenders through messages that have no headline and seems clear, when in reality it utilizes HTML code to download different records.

2.3. Backscatter spam:

Backscatter is a response against worms, viruses, and spam emails, in which email servers that accept spam and other emails forward bogus messages to an innocent group of people. This occurs because the email address of the victim is revealed in the envelope sender of the initial message. A significant portion of these emails have a created from: header that matches the sender's address. These communications qualify as spontaneous bulk email or spam since the recipients did not request them, they are remarkably similar to one another, and they are sent in large quantities. In this regard, frameworks that generate email backscatter may end up being noted on various DNSBLs and violating the terms of service of network access providers.

III. THE ALGORITHMS

This segment gives a short review of the hidden hypothesis and usage of the calculations we consider. We should examine the Naïve Bayesian classifier, Modified Naïve Bayesian classifier , Support Vector Machine ,the k-NN classifier, the Neural system classifier, the bolster vector machine classifier and Artificial Immune System classifier.

3.1 Naïve Bayes Classifier:

The Naive Bayes classifier is a straightforward factual calculation with a long history of giving shockingly precise outcomes. It has been utilized as a part of a few spam order studies [9, 10, 11, 12], and has progressed toward becoming to some degree a benchmark. It gets its name from being founded on Bayes' control of restrictive likelihood, consolidated with the "innocent" supposition that all contingent probabilities are free [13]. Gullible Bayes classifier looks at all of the case vectors from both classes. It computes the earlier class probabilities as the extent of all occasions that are spam (Pr[spam]), and not-spam (Pr[notspam]). At that point (expecting paired characteristics) it gauges four restrictive probabilities for each quality: Pr[true|spam], Pr[false|spam], Pr[true|notspam], and Pr[false|notspam]. These evaluations are ascertained in view of the extent of occasions of the coordinating class that have the coordinating an incentive for that property.

To arrange a case of obscure class, the "guileless" variant of Bayes' manage is utilized to gauge first the likelihood of the occasion having a place with the spam class, and after that the likelihood of it having a place with the not-spam class. At that point it standardizes the first to the aggregate of both to deliver a spam certainty score in the vicinity of 0.0 and 1.0. Take note of that the denominator of Bayes' manage can be overlooked on the grounds that it is counteracted in the standardization step. As far as usage, the numerator has a tendency to get very little as the quantity of traits develops, on the grounds that such a variety of modest probabilities are being duplicated with each other. This can turn into an issue for limited exactness drifting point numbers. The arrangement is to change over all probabilities to logs, and perform expansion rather than augmentation. Note likewise that restrictive probabilities of zero must be stayed away from; rather a "Laplace estimator" (a little likelihood) is utilized. Note that utilizing double characteristics in the occasion vectors makes this calculation both less difficult and more proficient. Likewise, given the predominance of inadequate example vectors in content grouping issues like this one, double credits offer the chance to execute exceptionally noteworthy execution advancements.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

3.2 Modified Naïve Bays Classifiers

Compared to the current Naïve Bayes (NB) or Supporting Vector Machine (SVM) classifier, the Modified Naïve Bayes (MNB) classifier generates more accurate results and ensures the requirements with a very low percentage of training. [19]

All words in a mail have independent nature of spam level according to laws of probability the probabilities of independent event should not be added to sum of probabilities, which results more than one. For example, consider the word "Bumper" is a hammy word and "Prize" is also a hammy word but when these words combine together "Bumper Prize" which is a spammy word. This example shows that the combination of words can also create a spam which cannot be calculated by ordinary classification methods. Proposed scheme introduces a method to combine the probabilities of many independent events and take it as a single probability of an email and utilize it to evaluate whether the given mail is spam or not. This scheme is implemented via a slightly different approach in NB classifier. Its training Enron dataset also contains difference in ratio of ham and spam mail in order to show that a recipient will receive ham mails more than the spam emails. It starts by counting the number of appearance of ham words in a test document (AH) and number of appearance of spam word in a test document (AS).[19]

3.3 Support Vector Machine

Bolster vector machines (SVMs) are relatively new techniques that have gained popularity very quickly due to their amazing results in a variety of machine learning problems and their solid theoretical foundations in quantifiable learning hypotheses[14].

Bolster vector machine (SVM) calculations separate the n-dimensional space portrayal of the information into two locales utilizing a hyperplane. This hyperplane dependably augments the edge between the two areas or classes. The edge is characterized by the longest separation between the cases of the two classes and is registered in view of the separate between the nearest occasions of both classes to the edge, which are called supporting vectors [15]. Rather than utilizing direct hyperplanes, numerous usages of these calculations utilize alleged piece capacities. These portion capacities prompt non-straight characterization surfaces, for example, polynomial, outspread or sigmoid surfaces [16]. Formal definition - More formally, a bolster vector machine develops a hyper plane or set of hyper planes in a high-or unbounded dimensional space, which can be utilized for order, relapse, or different assignments. Instinctively, a great partition is accomplished by the hyper plane that has the biggest separation to the closest preparing information purposes of any class (supposed utilitarian edge), since all in all the bigger the edge the lower the speculation blunders of the classifier.

3.4 Artificial Neural Networks

A simulated neural system (ANN), more often than not called neural system (NN), is a scientific model or computational model that is enlivened by the structure or potentially useful parts of organic neural systems. A neural system comprises of an interconnected gathering of manufactured neurons, and it forms data utilizing a connectionist way to deal with calculation. As a rule, an ANN is a versatile framework that progressions its structure in view of outer or inner data that courses through the system amid the learning stage. Present day neural systems are non - straight factual information displaying devices. They are generally used to model complex connections amongst sources of info and yields or to discover designs in information.

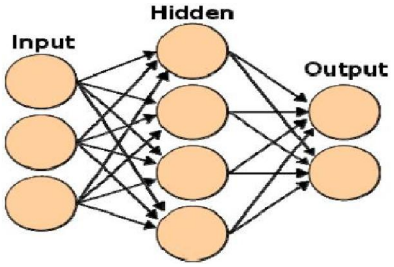


Fig.4. an artificial neural network

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

By definition, a "neural system" is an accumulation of interconnected hubs or neurons. See fig. 7. The best-known case of one is the human mind, the most unpredictable and modern neural system. On account of this cranial-based neural system, we can settle on exceptionally fast and solid choices in portions of a moment. [17]

ANN is an interconnected group of nodes , akin to the vast network of neurons in the human brain .

Spam exhibits a one of a kind test for conventional sifting innovations: both regarding the sheer number of messages (a huge number of messages day by day) and in the expansiveness of substance (from explicit to items and administrations, to back). Add to that the way that today's monetary texture relies on upon email correspondence – which is similarly expansive and abundant and whose topic logically covers with that of many spam messages – and you have a genuine test.

How it functions - Since a neural system depends on example acknowledgment, the hidden preface is that each message can be measured by an example. This is spoken to underneath in Fig. 8. Each plot on the diagram (otherwise called a "vector") speaks to an email message. In spite of the fact that this 2-D illustration is an over-disentanglement, it imagines the guideline utilized behind neural systems.

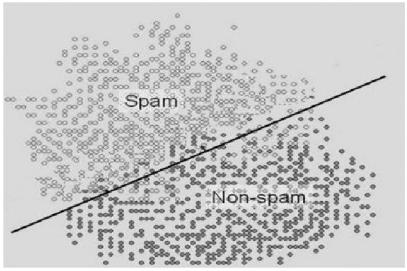


Fig.5. Distinctive patterns of good and spam messages cluster into relatively distinct groups

To distinguish these examples, the neural system should first be "prepared". This preparation includes a computational investigation of message substance utilizing huge delegate tests of both spam and non-spam messages. Basically the system will "learn" to perceive what we people mean by "spam" and "non-spam". To help in this procedure, we initially need a reasonable, brief meaning of "spam": Spam, n., email sent in mass where there is no immediate assention set up between the beneficiary and the sender to get email sales. U.B.E. (Spontaneous Bulk Email) is another acronym for spam that adequately typifies this definition. To make preparing sets of spam and non-spam messages, each email is painstakingly checked on as per this basic, yet prohibitive meaning of spam. In spite of the fact that the normal client regularly considers every undesirable email as "spam", messages that verge on "requested" (it was likely asked for sooner or later by the client) ought to be dismisses by and large. Cases of these might incorporate email sent from effectively conspicuous areas, for example, Amazon.com or Yahoo.com. A decent saying to take after here is: "if all else fails, toss it out". So also, non-spam email ought to be limited to individual email interchanges between people or gatherings, and keep away from any types of mass mailings, paying little mind to whether they were requested or not. Once these sets have been assembled and affirmed, the neural system is prepared for preparing.

The ANN framework now preprocesses each email in the separate preparing sets to decide precisely which of these important words are found in each spam email, and which of these words are found in the non-spam email. Next, the ANN is prepared to perceive certain blends or examples of fascinating or significant words to recognize spam, or on the off chance that it sees different mixes, to distinguish non-spam.

The counterfeit neural system utilizes an arrangement of complex numerical conditions to play out this sort of calculation. As some spam and non-spam messages will frequently "share" attributes, a reasonable qualification can't generally be made.

By the "non-spam" plots or vectors that wind up in the "spam" bunch and the other way around. In this "hazy area" lies the potential for false positives. After the preparation is finished, the ANN can now be utilized to check live-stream email. Each message is filtered to recognize important words, which are then prepared by the ANN.

In the event that the ANN again observes certain sorts of blends of word utilization demonstrating a likelihood of spam, it will report spam, alongside likelihood esteem. Taking after the case in Fig. 9, if the vector or plot registered for the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

message arrived over the isolating line, it would be considered "spam". Its likelihood or certainty level would rely on upon the relative separation far from the line.

To boost identifications while maintaining a strategic distance from false positives, a very much composed heuristics motor will suit diverse affectability edges, or levels of forcefulness, in recognizing spam. This means the cut-off or separating point amongst spam and non-spam can be balanced so that the probability of a false positive match will be significantly lessened. This can be found in Fig. 6 beneath.

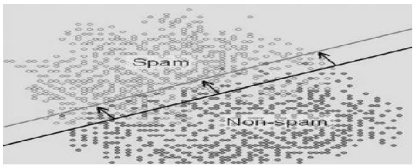


Fig.6. The sensitivity threshold can be adjusted to avoid the “grey” area.

At the end of the day, the further far from the focal isolating line amongst ham and spam email bunches, the lower the shot of false positive recognitions. Note in Fig. 9 that there are far less non-spam vectors or examples over the new cut-off or partitioning line.

3.5 K-closest neighbor classifier:

The k-closest neighbor (K-NN) classifier is considered an example-based classifier, meaning that preparation reports are used for analysis rather than an explicit classification representation, such as the classification profiles used by many classifiers. Taking everything into account, there isn't a true stage of preparation. When a new archive needs to be organized, the k closest reports (neighbors) are located. If a sufficient number of them have been assigned to a particular class, the new record is also assigned to that classification; otherwise, it is not. Furthermore, finding the closest neighbors can be revived utilizing customary ordering techniques. To choose whether a message is honest to goodness or not, we take a gander at the class of the messages that are nearest to it. The examination between the vectors is a constant procedure. This is the possibility of the k closest neighbor calculation:

Stage 1. Preparing:

Store the preparation messages.

Stage 2. Sifting:

Given a message x, decide its k closest Neighbors among the messages in the preparation set. On the off chance that there are more spams among these neighbors, characterize given message as spam. Generally characterize it as real mail.

We ought to note that the utilization of an ordering strategy keeping in mind the end goal to diminish the season of correlations incites a refresh of the example with a multifaceted nature O(m), where m is the specimen measure. As the majority of the preparation cases are put away in memory, this system is likewise alluded to as a memory-based classifier [24]. Another issue of the introduced calculation is that there is by all accounts no parameter that we could tune to decrease the quantity of false positives. This issue is effectively settled by changing the characterization manages to the accompanying l/k run the show:

In the event that l or more messages among the k closest neighbors of x are spam, order x as spam, generally arrange it as real mail.

The k closest neighbor run has discovered wide use when all is said in done arrangement errands. It is likewise one of only a handful few generally predictable characterization rules.

IV. CONCLUSION

Spam is turning into an intense issue to the Internet people group, undermining both the trustworthiness of the systems and the profitability of the clients. In this paper, we propose five machine learning strategies for hostile to spam separating. In this paper we talked about the issue of spam and gave an outline of learning based spam sifting systems. There is no regular meaning of what spam is, yet the greater part of the sources concur that the center element of the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

wonder is that spam messages are spontaneous. Spam causes various issues of both sparing and moral nature, which brings about specific in the endeavors of administrative definition and preclusion of spam.

The most mainstream and very much created way to deal with hostile to spam is learning based separating. The present cutting edge incorporates many channels in view of different grouping procedures connected to various parts of email messages.

Email sifting is the handling of email to arrange it as indicated by determined criteria. Regularly this alludes to the programmed handling of approaching messages, however the term likewise applies to the intercession of human insight notwithstanding hostile to spam strategies, and to active messages and in addition those being gotten. Email separating, programming inputs email. For its yield, it may go the message through unaltered for conveyance to the client's post box, divert the message for conveyance somewhere else, or even discard the message. Some mail channels can alter messages amid handling.

All in all, we can state that the field of hostile to spam assurance is at this point develop and very much created. At that point a question emerges, why our inboxes are still frequently loaded with spam? Reactivity of spammers assumes a part most likely, thus does the perplexing way of spam information. However, one more issue not to be thought little of here is that we more often than not don't secure against spam in all the accessible ways. At the end of the day, one point which ought to dependably be recalled by server heads and end clients is that the counter spam innovations ought to be composed and created, as well as conveyed and utilized.

REFERENCES

[1]. Aladdin Knowledge Systems, Anti-spam white paper, www.csisoft.com/security/ aladdin/esafe_antispam_whitepap er.pdf Retrieved December 28, 2011.

[2]. F. Smadja, H. Tumblin, 2002,"Automatic spam detection as a text classification task", in: Proc. of Workshop on Operational Text Classification Systems, 2002.

[3]. A. Hassanien, H. Al-Qaheri, 2009,"Machine Learning in Spam Management", IEEE TRANS., VOL. X, NO. X, FEB.2009.

[4]. P. Cunningham, N. Nowlan, 2011,"A Case-Based Approach to Spam Filtering that Can Track Concept Drift",[Online]Available:https://www.cs.tcd.ie/publications/tech-reports/reports.03/TCD-CS-2003-16.pdf Retrieved December 28, 2011

[5]. F. Roli, G. Fumera, 2011 "The emerging role of visual pattern recognition in spam filtering: challenge and opportunity for IAPR researchers" http://www.iapr.org/members/newsletter/Newsletter07-02/index_files/Page465.htm Retrieved December 28, 2011

[6]. H. West, "Getting it Wrong: Corporate America Spams the Afterlife". Clueless Mailers. (January 19, 2008).

[7]. B. Parizo,2006,"Image spam paints a troubling picture". Search Security. (2006-07-26)

[8]. Symantec (2011) VBS.Davinia.B, [Online] Available: http://www.symantec.com/security_response/ writeup.jsp?do cid=2001-020713-3220-99 Retrieved December 28, 2011

[9]. Androutsopoulos, J. Koutsias, "An evaluation of naive bayesian anti-spam filtering". In Proceedings of the Workshop on Machine Learning in the New Information Age, 11th European Conference on Machine Learning (ECML 2000), pages 9–17, Barcelona, Spain, 2000.

[10]. Androutsopoulos, G. Paliouras, 2000, "Learning to filter spam E-mail: A comparison of a naïve bayesian and a memory-based approach". In Proceedingsof the Workshop on Machine Learning and Textual Information Access, 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2000), pages 1– 13, Lyon, France, 2000.

[11]. Hidalgo, 2002"Evaluating cost-sensitive unsolicited bulkemail categorization". In Proceedings of SAC-02, 17th ACM Symposium on Applied Computing, pages 615–620, Madrid, ES, 2002.

[12]. Schneider, 2003, "A comparison of event models for naivebayes anti-spam e-mail filtering". In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics, 2003.

[13]. Witten, E. Frank, 2000,"Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations". Morgan Kaufmann, 2000.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, March 2024

[14]. N. Cristianini, B. Schoelkopf, 2002, "Support vector machinesand kernel methods, the new generation of learning machines". Artificial Intelligence Magazine, 23(3):31–41, 2002

[15]. V. Vapnik, 1998, "The Nature of Statistical Learning Theory ,Springer; 2 edition (December 14, 1998)

[16]. S. Amari, S. Wu, 1999, "Improving support vector machine classifiers by modifying kernel functions". Neural Networks, 12, 783– 789. (1999).

[17]. C. Miller, 2011,"Neural Network-based Antispam Heuristics" ,Symantec Enterprise Security (2011), www.symantec.com Retrieved December 28, 2011.

[18]. C. Wu, 2009, "Behavior-based spam detection using a hybridmethod of rule-based techniques and neural networks" , Expert Syst., 2009

[19]. Anitha P. U., Guru Rao C. V., Babu Suresh, 2017" Email Spam Classification using Neighbor Probability based Naïve Bayes Algorithm" 7th International Conference on Communication Systems and Network Technologies,2017

[20]. Awad Mohammed, Foqaha Monir , 2016," Email Spam Classification Using Hybrid Approach of RBF Neural Network and Particle SWARM Optimization", International Journal of Network Security & Its Applications (IJNSA) Vol.8, 2016