Improve Spam Filtering of Electronic mail

[1]Rojali Priyadarsini, [2]Akhilesh Bansiya

[1]MTech Scholar, [2]Assistant Professor

[1]Department of Computer Science Engineering, Vedica Institute of Technology,Bhopal,India

[2]Department of Computer Science Engineering, Vedica Institute of Technology,Bhopal,India

[1]rojali9@gmail.com [2]akhilesh2483@gmail.com

Abstract:-

Nowadays, maybe the most affordable kind of report in the globe is email, and its straightforwardness makes it helpless against lots of perils. One of the most essential perils to email is spam; unwanted email, especially when headway association send a group mail. Spam email could similarly contain malware as substance or further executable record. Inconsistently they in like manner fuse hazardous association or associations with phishing sites. These cruel spams's trade off the insurance and security of gigantic proportion of delicate information. In this way, a structure that can normally make sense of how to sort vindictive spam in email is incredibly appealing. In this paper, we hope to make finding of harmful spam all through component assurance. We suggest models that use a novel informational collection for the strategy of feature assurance, a development for improving arrangement in later stage. Feature assurance is obvious to make preparing time and precision of malicious spam area. This paper likewise shows the evaluation of various classifiers use during the cycle.

Keywords: Electronic mail, SVM, Spam, Dataset, Naive Bayes

1. Electronic Email Spam Filtering using SVM

Electronic mail is a key turmoil happening over standard correspondence structures on account of its, brisk, accommodating, basic, and pragmatic, to use nature. An essential bottleneck in electronic exchanges is the monstrous scattering of unfortunate, unsafe emails known as spam emails. A key concern is the making of fitting channels that can enough catch those emails and get prevalent rate. Machine learning (ML) researchers have developed different procedures to deal with this issue. Inside the structure of machine learning, uphold vector machines (SVM) have organized a colossal part to the improvement of spam email isolating. Taking into account Support Vector Machine, unmistakable arrangement have been orchestrated through text classification moves close (TC). A fundamental issue while using SVM is the assurance of pieces as they straightforwardly impacts the section of emails in the quality space [5]. Here fig 1; explains the spam isolating using SVM.

1.1 Spam Information Recognition Features on the Public Networks:

Various papers contain completed in the field of spam acknowledgment on top of the casual associations. These examinations have raised one or various features for spam acknowledgment. Some article has created as of late used for a relational association and diverse contain reviewed different organizations. Besides, a couple has created on the spam customer accounts area and a couple was about spam portion post revelation in the casual associations. We

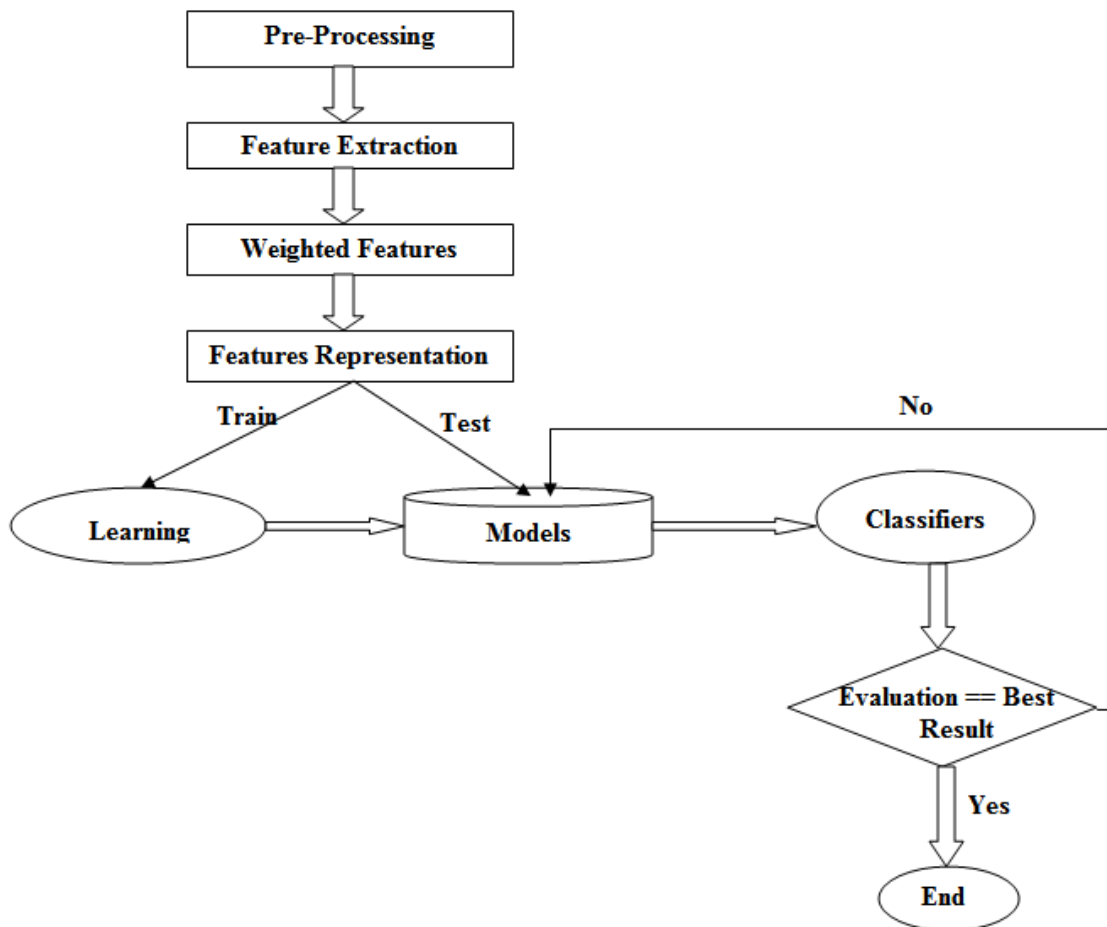will focus these cases unreservedly in different parts.

Fig 1: Spam Filter

In 2010, [5] disconnected the features of the Twitter spammers divided into two social affairs: content - based and graph based features and depict the notification completely about spammer disclosure. In content-based features division, it was referred to there are four features to perceive spam customer account.

Associations or Connections: If by far most of the send tweets beginning at a customer account encase the association, they will be seen as spammer.

Recurring Tweets: If customer accounts send rehashing tweets, it will be seen as a spammer.

Mentions and Replies: For the most part of the tweets send from a customer account contain the appropriate responses and notification, recognized as a spammer.

Trending topics: If customer accounts send unessential issue to moving subjects, it will be seen as a spammer.

URL ratio (R): The second next segment for spam area is occasion of URL in divider post. For the customer's appeal to spam site pages, spammers throw associates with guarantee divider post.

Ratio (R): Ratio of move friend requests numeral to the amount of customers that have developed the sales is orchestrated as a measure for spam revelation. Since spam is absolutely not a certified someone, so any one know it, in light of everything, and basically a limited quantity of the customer account recognize partner requests.

Selecting with searching by friends (F): The fourth one segment, is identified with this issue that, whether or not there are buyer accounts that have search the resolute record as their buddy or not. These properties are called F and portray to:

$$F = T_n/D_n \quad (2)$$

That T_n is the overall number of names between the friends' customer records and D_n is exceptional name number.[1]

Message similarity (S): The third one component is scattering of likeness among move messages by a customer.

2. Electronic E-mail Functions

As of late referred to in the beginning of this fragment, there has been a gigantic composition on text-mining. Moreover, there have been different works on classification and grouping messages to have been applied to email handling in create to lessen information over-trouble.

2.1 E-mails examination: The fundamental development in our email handling is to do an examination of the understudy's messages. The rule of this movement is to get a composed presentation that will be used to aggregate messages thusly to their semantics. For that, we prescribe to use the text mining techniques as a system for parsing understudy's messages.

2.2 E-mail Processing: Electronic mail can be viewed as an exceptional kind of chronicle as it is generally text close by some perceives information first class to it (e.g., to, from, cc, subject, associations, and so forth). In the couple of years earlier, through the beginning of text-mining, the evaluation of email started to procure an extended consideration of a growing number of experts.

2.3 E-mails Mining: The subsequent some segment of each textual report is parsed starting at now using the text-mining measure through some change in order to isolate the basic email body features. Most

likely, the text mining is applied toward textual information. Also since our treatment is done on the email, an individual kind of textual information, we name this action the pattern of Email mining.

2.4 E-mails Conversion: E-sends are indistinct basically. Thusly, the "Preparer Agent" changes over each email into a composed show. In this learning, we decide to address the email's HTML plan in a textual record that contains normally two areas: the first contains information of tending to, (for instance, the Subject, Recipients and Sender) and the resulting parts contain the collection of email. In the email change work, the "Preparer Agent" focuses just on the underlying section that will be parsed and tokenized all through the text-mining techniques to gain information about: Subject, Sender (From) and Recipient (To, CC, Bcc).

3. Literature Survey

As notice earlier, assemble a record of emails for assessment should be feasible for different purposes. One of the principal goals is spam acknowledgment. Spam is an issue concerning consent, not content. The Unsolicited Bulk Email ("UBE") message is an advert, pornography, a stunt, an asking letter or, more then likely a proposition of an open lunch, the content isn't associated - if the message was send unconstrained and in bulkiness, by then the message is spam. [4] A lot of studies have been disseminated sharing remarkable ways on the most ideal approach to fight spam such like the Rule Based Spam Filtering, Machine Learning methods, Content Hash Based Filtering, Support Vector Machines (SVM), Content-Based Filtering (CBF) and the Collaborative Filtering (CF) to give a few models. Among these systems, CBF has been the essentially wide used foe of spam course of action since it is uninhibitedly available with its business executions. [6] Current assessment revolve around improving substance classifier execution, by an improved preprocessing or redesign of learning figuring. Social events that solidify particular spam classifiers have moreover been organized. [7] However, comparably CF and CBF have drawback. CF faces issues, for instance, first-rater, and assurance. The basic issue is since of the complexity of request emails that have not been evaluated sooner than; the discretionary issue develops when customers rate only a couple of messages; and the prop up one issue depends upon what is shared [7]. One of the strong advantages of the CBF is to it diminishes bumble rates as legitimate email would not be barren whether or not the ISP from which it start, is

lying on a nonstop square once-over with it simply needs accidental alteration, which implies less issue for end-customer. [8] This sub zone depicts diverse exploration papers related with spam email classification.

3.1 Spam–non-spam email order:

We picked a couple of papers, base on reference, interrelated to spam affirmation or filtering. Those papers are: Blanzieri and Bryl, 2008 [10]; Zhuang et al., 2008 [9]; Mishne et al., 2005 [12]; Webb et al., 2006 [11]; Zhou et al., 2010 [14]; Sculley and Wachman, 2007 [13]; Xie et al., 2006 [15]; Bogawar et al. 2012 [17]; Katakis et al. 2007 [16]; Ozcaglar 2008 [18]. Not in the least like papers inspected the using of unprecedented calculations and besides apply the calculations in extraordinary spots between email senders close by gatherers.

Zhuang et al's. (2008) [9] article focused in on endeavoring toward find Botnets. Botnets are bundle at risk for scattering spam emails. Method is surveyed to recognize such wellspring of spam campaign that shares some wide features. Spammers how-ever attempt to change spam emails all through some normal goof or disarrays especially stylish isolated watchwords.

Blanzieri and Bryl (2008) [10] existing a specific clarification in 2008 to read data calculations for spam isolating. The papers talk about different perspectives identified with spam isolating, for instance, the suggestion toward change or modify email broadcast shows to include techniques to take out or reduce spam.

Webb et al's. (2006) [11] paper focused about web spam with how to apply email spam disclosure techniques to perceive spam webpage pages. The equivalent to the best approach to manage perceive spam in emails, site pages are analyze for explicit features that may arrange them as spam pages, for instance, using expressions stuffing, arbitrary well known words, etc. Mishne et al's. (2005) [12] paper addresses one more event of web or, without a doubt association spam research paper. Web diaries, public organizations, news or, more then likely even online business sites nowadays license customers to give their comments or analysis. Spammers use such ability to post spam messages during those posts. Thusly spam revelation procedures must be moreover used to permit customized acknowledgment of such posts.

Sculley and Wachman (2007) [13] talk about likewise calculations, for instance, VSM for email, web, and sites and web and association spam affirmation. The substance of the email or, no doubt the page is inspected by various ordinary language

preparing approach, for instance, NGram, Bags of words, etc. The impact of an exchange limit VSM is surveying using divergent setting regard made arrangements for such limit.

VSM execution and desire exactness is high while the assessment of this limit is high. Zhou et al. (2010) [14] organized a spam-based classification plan of three classifications. In adding to model spam and not spam classification, a third questionable classification is provide for additional versatility to the desire computation. Uncertain emails should be reexamined and assemble more information to be capable then to intellectual whether they are spam or most likely not. Makers use Sculley and Cormack (2008) [19] UCI Machine Learning vault or storage facility since their preliminary email dataset (machine learning safe or file).

Perez-Dıaz et al's. (2012) [12] paper 2012 survey apply unbalanced set on spam affirmation with unique rule execution intend to get the best organizing one. UCI Spam base is use in the investigational study (machine learning store or storage facility).

Xie et al's. (2006) [15] paper 2006 endeavor to sum up features that can see Botnets or spam delegate that are used to throw an epic number of spam emails. Makers see network

interrelated practices that can undoubtedly perceive such spam mediator.

3.2 Email Data Examination objective

In this segment, we will depict various papers identified with the appraisal of email messages for reason other than spam introduction.

Kiritchenko and Matwin (2001) [20] offered a paper on email request through solidify named and unlabeled data. Like to various papers, VSM is show to be the most grand classifier in game plans of desire or classification execution. Text game plan is used to sort emails into various envelopes based on predefined class. Makers endeavor to portray classes as fascinating close by monotonous classification. A basic overview of genuinely checked emails can be used for the future ordinary preparing and classification. VSM is show to profit by the co-preparing measure organized in this paper. Enron email record is used in different assessment papers in email request (Klimt and Yang, 2004) [21]. Shetty and Adibi's (2005) [22] paper use Enron email record in email course of action based on graph entropy model. The entropy endeavors to pick the renowned charming centers (that identify with emails) in the graph. Edges

identify with messages between various emails customers.

Yoo et al. (2009) [23] talk about balanced email prioritization in email correspondence and public networks or social events. Targets, for instance, bunch contacts and course of action (Using Newman clustering methodology) were survey in association through email messages close by casual networks.

Klimt and Yang (2004) [21] thought about relations in email messages, for instance, the relations among contacts and messages or likely series of messages. Series of messages consolidate different emails exchange between two or various individuals all through some email messages. Enron dataset is use in this learn like a lot of other related exploration papers here any spot it is considered as the major publically existing email messages dataset. For this particular paper, one even more little email dataset (CMU) is use. McCallum and Wang's (2007) [24] paper is furthermore in the area of casual networks close by email assessment through the goal of significant examination and course of action based on relations among people. Carmona-Cejudo et al's. (2011) [25] papers are identified with ceaseless email game plan and present GNUs mail open source used for email

envelope request. The application is made to parse emails from dissimilar email clients close by play out a couple of data mining assessment with WIKA data mining device. In email database request is moreover base lying on the hour of email messages (Bekkerman et al., 2004) [26]. The paper use Enron and SRI email datasets proposed for the case works out. A couple of new game plan procedure, for instance, MaxEnt was survey inside the paper. The key decision to shape in each email game plan papers is what features to pick. Features can be identified with email allocate, from or to addresses or likely can be interrelated to the content; words, plan of words, etc. Trademark language handling stunts, for instance, parsing similarly as stemming are then stressed to parse email contents close by discard any words that may not be associated for the classification strategy.

Flying animal's (2004) [27] papers analyze an approach to manage guess answer on emails based on mine data. Instance of answer desire can be identified with for example the most sensible individual to respond toward an email. Data Retrieval (IR) Latent Semantic Indexing (LSI) system can be use to parse and take out features from emails. Artificial Neural Networks (ANNs) are use and show to have

staggeringly extraordinary results in courses of action of desire exactness.

3.3 Ontology arrangement of email contents:

Cosmology's proposed for different purposes identified with the reusability of real factors, real factors sharing and examination and besides to parcel shared characteristics from contrast in the abnormal real factors domains. In the organized exploration subject of cosmology request or data extraction of Email contents, there have been distinctive assessment papers that endeavor to propose and begin thoughts all things considered found in Email contents. Such way of thinking could in like manner be used for email endorsement or, without a doubt spam ID. For instance, Taghva et al's. (2003) [28] paper masterminded email thoughts mining using Ecdysis Bayesian email classifier. Maker's focus email contents base on features together from the concentrate or arranged data and furthermore from DOE inclusionary or, without a doubt exclusionary records (Office of Civilian Radioactive Waste Management, 1992). Inclusionary thought contain: Organization, Department, and Message Topics and Email Agent. Exclusionary thoughts contain: Email Characteristics, and Count Characteristics, and Attachment Type

Characteristics. Those components join different interrelated attributes. Protege ontological contraption (http://protege.stanford.edu/) was used to work close by show the cosmology. Inside our case, MIME parser is use to parse from emails a huge amount of qualities of those depict in Taghva et al.[28] cosmology.

Yang and Callan (2008) [29] in 2008 offered moreover cosmology to isolate thoughts from a corpus of public notes (Mercury and Polar Bear datasets). NGram mining is use to see contender thought. Wordnet and surface text configuration contrasting are used with see associations close by the thoughts. Wordnet watchwords are use to oversee relationship of thought into arranged hierarchal affiliations.

Beseiso et al's. (2012) [30] paper masterminded a procedure for thoughts' extraction beginning from email systems. Makers talk about individual troubles of emails thoughts' extraction while generally speaking; customers' emails are territories explicit and particularly subordinate lying on the individual, their tendencies, and calling, etc. Makers complete NEPOMUK Message Ontology and describe email general thoughts similarly as space unequivocal thoughts. Makers use Enron and custom email datasets for evaluation.

Aloui and Neji's (2010) [31] paper broadened a system for modified email classification close by question answering. The philosophy masterminded three lots of emails base on their general subjects: Social, procedural, and scholarly limits. This paper extended an approach in the paper of Leˆ and Leˆ (2002) [32]. The 10 classifications contain: Requesting, Discussing, Thinking, Confirming, Clarifying, Referring, Complimenting, Greeting, Complaining, and Sharing.

Text bundling close by classification can be used for an expansive scope of uses. For instance, Altwaijry and Algarny's (2012) [33] paper use text classification system to describe network pay data similarly as traffic close by mastermind such data into peril (ruinous) or, no doubt non-risk data. A Naive Bayesian (NB) classifier is use. Such classifier is winding up being successful for classification in different zones. Makers use public KDD IDS dataset used for testing similarly as preparing.

One more critical application region for classification especially in data recuperation systems joins picture request (De and Sil, 2012) [34]. In this paper, makers use cushy rationale to administer fragile class names to the divergent pictures in the framed dataset. Such picture game plan can be use for web crawlers request close by all things considered pictures are associated with embedded text or, without a doubt text arranged about those photos.

Following are a part of the concentration inside the assessment of email study (Based on our examination of papers interrelated to investigate papers during data mining in emails' datasets):

1. Usually, email study can be requested under text classification inside its most activities. Calculations, for instance, KNN, VSM, Ripper, Winnow, Maximum Entropy (MaxEnt), ANN are instance of calculations use in email study.

2. A rule research subject in email classification is to arrange emails into spam or likely no spam emails. This can also use for the continuous figuring of spam emails.

Algorithms	Precision	Recalls	F-measure	Accuracy	FPR	Network's name	Decision Methods	Reference
DenStram	0.7272	1	0.8421	0.9711	0.31	Twitter	Clustering	[9]
StreamKm++	0.5591	1	0.7172	0.9393	0.65			
Combine	0.7939	1	0.8851	0.98	0.21			
Decision Tree	0.667	0.333	0.444	..	..	Twitter	Classification	[6]
Neural Network	1	0.417	0.588		..			
SVM	1	0.25	0.4	..	..			
Naïve Bayes	0.917	0.917	0.917	..	..			
Naïve Bayes	..	0.976	..	..	0.075	Twitter	Classification	[8]
Jrip	..	0.987	..	..	0.014			
J48	..	0.983	..	..	0.017			
Naïve Bayes	..	0.733	..	..	0.301	Combined		
Jrip	..	0.935	..	..	0.071			
J48	..	0.975	..	..	0.048			
Naïve Bayes	..	0.965	..	..	0.089	Facebook		
Jrip	..	0.912	..	..	0.09			
J48	..	0.898	..	..	0.081			
MCL			0.88			Facebook	Clustering	[11]
SVM	0.8732	0.893	0.883	0.922	0.063	Twitter	Classification	[13]
Suppliments M/c Learning	0.7396	0.7467	0.7431	0.7902	0.1067	Twitter	Classification	[10]
Naïve Bayes	0.9187	0.799	0.8546	0.8642	0.07	Twitter	Classification	[4]

Table 1: Spam classification methods

Some email spam classification research papers attempted to arrange emails spam based on the gender of the sender given a portion of the regular perspectives that may recognize emails from females or guys is in table 1. Email classification can be likewise used to consequently appoint emails to predefined organizers. Moderately spam alongside non spam emails, emails can be likewise grouping into: Interesting alongside dull emails.

Highlights are removing from the email content or probably body, title or, more than likely subject or, in all likelihood a portion of the other Meta data that can be extricate from the emails, for example, collector, sender, BCC, date of sending, number of beneficiaries, accepting, and so forth. This strategy separate component can be base on words, pack of words, and so forth. Email grouping too thought about bunching emails into unique subjects or, more than likely organizers.

The time data inside emails (for example at the point when: sent, gotten, and so on.) is utilized also in some examination papers to arrange emails. A few examination papers attempt to classify emails based on practically identical strings or subjects. Some email framework, for example, Gmail associate emails corresponded to one

another (for example by answer or, more than likely forward occasions) together.[2]

4. Problem Statement

Some email spam classification research papers endeavored to organize emails spam based on the sexual orientation of the sender given a bit of the ordinary points of view that may perceive emails from females or folks is in fig (b). Email classification can be moreover used to subsequently designate emails to predefined coordinators. Modestly spam close by non spam emails, emails can be similarly gathering into: Interesting close by dull emails.

Features are eliminating from the email content or presumably body, title or, without a doubt subject or, probably a segment of the other Meta data that can be remove from the emails, for instance, gatherer, sender, BCC, date of sending, number of recipients, tolerating, etc. This procedure separate segment can be base on words, pack of words, etc. Email gathering too pondered bundling emails into remarkable subjects or, without a doubt coordinators.

The time data inside emails (for instance right when: sent, gotten, etc.) is used additionally in some assessment papers to mastermind emails. A couple of assessment papers endeavor to order emails based on essentially indistinguishable strings or

subjects. Some email structure, for instance, Gmail partner emails related to each other (for instance by answer or, without a doubt forward events) together.[2]

Web spam which is a most huge issue through the current web search gadget; in this way it is principal for web crawlers to contain the capacity to perceive web spam among creeping. The game plan Models are considered by machine learning describe count. The one machine learning figuring is Naïve Bayesian Classifier which is excessively used in to part the spam similarly as non-spam sends. Gigantic Data separate structure which is likewise plot used for spam acknowledgment. Concentrate the tendency beginning at a message is a procedure for get the critical data. In Machine learning advancement can get from the preparation datasets moreover anticipate the tendency creation structure in this way they are widely use as a limited quantity of feeling demand through the extraordinarily exactness of framework.

5. Future Research Scope

This work proposes a model for improving affirmation of merciless spam in email. Our model reason use a novel dataset expected for the pattern of feature choice, and thereafter endorse the arrangement of picked features using three classifiers perceived in

spam recognizable proof: Support Vector Machine, Naïve Bayes, and Multilayer Perception. Feature decision is reached out to recover preparing time similarly as precision for the classifiers

6. Conclusions

To overview the outcomes of the speculation it will in general be expressed, that the arrangement of a Meta spam channel look good similarly as has its ground. Notwithstanding the way that the idea oversees existing spam channels similarly as email corpus, the over depict procedure can too be applied for extra channels as well. Examinations of Bayesian networks have given a fine base to the creation of a Meta spam channel.

References

[1]. Nasim eshraqi, Mehrdad Jalali and Mohammad Hossein Moattar 2015, Spam Detection In Social Networks: A Review, Second International Congress on Technology, Communication and Knowledge (ICTCK 2015) November, 11-12

[2]. Izzat Alsmadi, Ikdam Alhami, Clustering and classification of email contents, Journal of King Saud University – Computer and Information Sciences, 46-57.

[4]. Issam dagher, Rima Antoun, Ham- Spam Filtering Using DIFFERENT PCA SCENARIOS, 2016 IEEE International Conference on Computational Science and Engineering.

[5]. Ola Amayri, Nizar Bouguila, A study of spam filtering using support vector machines, Artif Intell Rev (2010) 34:73–108.

[4] Spam Haus http://www.spamhaus.org/consumer/definition/

[6] Garriss et al., 2006. Re: Reliable Email

[7] Lopes, C., et al, 2009. Symbiotic filtering for spam email detection

[8] Islam, Md. Rafiqul & Chowdhury, U. Morshed. 2005. SPAM FILTERING USING ML ALGORITHMS

[9] Zhuang, L., Dunagan, J., Simon, D.R., Wang, H.J., Tygar, J.D., 2008. Characterizing Botnets from Email Spam Records, LEET'08 Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats Article No. 2.

[10] Enrico Blanzieri, Anton Bryl, 2008. A survey of learning-based techniques of email spam filtering, Technical Report # DIT-06-056.

[11] Steve Webb, James Caverlee, Calton Pu, 2006. Introducing the Webb Spam Corpus: using Email spam to identify web spam automatically, CEAS.

[12] Mishne, G., Carmel, D., Lempel, R., 2005. Blocking blog spam with language model disagreement. In Proc. 1st AIRWeb, Chiba, Japan.

[13] Sculley, D., Gabriel M. Wachman, 2007. Relaxed online VSMs for spam filtering, SIGIR 2007 Proceedings.

[14] Bing Zhou, Yiyu Yao, Jigang Luo, 2010. A three-way decision approach to email spam filtering. Canadian Conference on AI, pp. 28–39.

[15] Mengjun Xie, Heng Yin, Haining Wang, 2006. An effective defense against email spam laundering, CCS'06, October 30–November 3,Alexandria, Virginia, USA.

[16] Ioannis Katakis, Grigorios Tsoumakas, Ioannis Vlahavas, 2007. Email Mining: Emerging Techniques for Email Management, Web Data Management Practices: Emerging Techniques and Technologies, IGI.

[17] Pranjal S. Bogawar, Kishor K. Bhoyar, 2012. Email mining: a review, IJCSI Int. J. Comput. Sci. Issues 9(1), No 1, January 2012.

[18] Cagri Ozcaglar, 2008. Classification of email messages into topics using latent dirichlet allocation, Master thesis, Rensselaer Polytechnic Institute Troy, New York.

[19] Sculley, D., Gordon V. Cormack, 2008. Filtering Email Spam in the Presence of Noisy User Feedback, CEAS.

[20] Svetlana Kiritchenko, Stan Matwin, 2001. Email classification with co training. In: CASCON '01: Proceedings of the 2001 conference of the Centre for Advanced Studies on Collaborative research.

[21] Klimt, Bryan, Yang, Yiming, 2004. The Enron corpus: a new dataset for email classification research. ECML, 217–226.

[22] Jitesh Shetty, Jafar Adibi, 2005. Discovering Important Nodes through Graph Entropy the Case of Enron Email Database, KDD'2005, Chicago, Illinois.

[23] Shinjae Yoo, Yiming Yang, Frank Lin, Il-Chul Moon, 2009. Mining Social Networks for Personalized Email Prioritization, KDD'09, June 28–July 1, Paris, France.

[24] McCallum, Andrew, Wang, Xuerui, 2007. Andre´ s Corrada-Emmanuel, Enron and academic email. J.Artif. Intell. Res. 30, 249–272

[25] Carmona-Cejudo, Jose´ M., Baena-Garcı´a, Manuel, Morales Bueno, Rafael, Gama, Joa˜ o, Bifet, Albert, 2011. Using GNUsmail to compare data stream mining methods for on-line email classification. J. Mach. Learn. Res. Proc. Track 17, 12–18.

[26] Ron Bekkerman, Andrew McCallum, Gary Huang, 2004. Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora.

[27] Christian Bird, 2004. Predicting Email Response using Mined Data, <http://www.cabird.com/papers/mlpaper.pdf> (last accessed 2014).

[28] Kazem Taghva, Julie Borsack, Jeffrey S. Coombs, Allen Condit, Steven Lumos, Thomas A. Nartker, 2003. Ontology-based Classification of Email, ITCC, IEEE Computer Society, pp. 194–198.

[29] Hui Yang and Jamie Callan. Ontology generation for large email collections. In Proceedings of the Eighth National Conference on Digital Government Research, 2008.

[30] Majdi Beseiso, Abdul Rahim Ahmad and Roslan Ismail," A New Architecture for Email Knowledge Extraction", International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.3, July 2012.

[31] Aloui, Awatef, Neji, Mahmoud, 2010. Automatic classification and response of E-mails. Int. J. Digital Soc. (IJDS) 1 (1).

[32] Thao Leˆ , Quynh Leˆ , 2002. 'The Nature of Learners' email communication. In: Proceedings of the International Conference on Computers in Education.

[33] Altwaijry, Hesham, Algarny, Saeed, 2012. Bayesian based intrusion detection system. J. King Saud Univ. Comput. Inf. Sci. 24, 1–6.

[34] De, Indrajit, Sil, Jaya, 2012. Entropy based fuzzy classification of images on quality assessment. J. King Saud Univ. Comput. Inf. Sci. 24, 165–173.