International Journal of Mechanical Engineering and Technology (IJMET)
Volume 9, Issue 1, January 2018, pp. 1037–1046 Article ID: IJMET_09_01_110
Available online at http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=9&IType=1
ISSN Print: 0976-6340 and ISSN Online: 0976-6359

© **IAEME** Publication **Scopus** Indexed

INVESTIGATION ON EMAIL TEXT MINING TECHNIQUES AND TOOLS

V.Sudheer Goud
Research Scholar, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, A.P, India

P. Premchand
Professor, Department of Computer Science and Engineering,
University College of Engineering, Osmania University, Hyderabad, T.S, India

ABSTRACT

Text Mining is an exciting area of research which uses techniques borrowed from data mining, machine learning, information retrieval, natural language understanding, case based reasoning, statistics, and knowledge management, to help the people to gain rapid insight into large quantities of semi structured or unstructured text. Email is one of the most popular forms of communication nowadays, mainly due to its efficiency, low cost, and compatibility of diverse types of information. In order to facilitate better usage of emails and explore business potentials in emailing, various text mining techniques have been applied on email so in this paper, discussing different types of text mining techniques for email data and tools. And designed a framework for email text mining and classification.

Keywords: TF-IDF, Email Text classification, Email Text Mining.

Cite this Article: V.Sudheer Goud and P. Premchand, Investigation on Email Text Mining Techniques and Tools, International Journal of Mechanical Engineering and Technology 9(1), 2018. pp. 1037–1046.
http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=9&IType=1

1. INTRODUCTION

Marti Hearst was solitary of the earliest researchers who conversed about Text Mining and obtainable a paper on it in 1999. According to him, Text Mining is the finding by computer of new, formerly unknown information, by repeatedly extract information from dissimilar written data sources. A key part is the connecting together of the extracted information jointly to form novel facts or new hypothesis to be explored additional by more predictable means of testing. Text Mining is dissimilar from what we're known with in web search. In search, the user is normally looking for impressive that is previously acknowledged and has been written by someone else. The difficulty is pushing aside all the substance that presently isn't pertinent to user requirements in order to discover the pertinent information. In Text Mining, the aim is to determine strange but helpful information from database or rather unstructured data. To the inexperienced, it may perhaps that Google and other Web search engines accomplish impressive similar, since they also through reams of email data in split second intervals. But, as experts note, search engines are only regain information, displaying lists of database that contain certain keywords. Text mining programs go additional, categorize information, making links among or else unconnected documents and given that visual maps to lead users down new pathways that they might not have been aware of.

RELATED WORK:

Spam emails are the biggest threat for today's internet. It causes financial crises and frustration among email customers. All approaches that have been developed to cater junk emails, filtering is one of the key approach. Spam emails, often referred to as junk emails or unsolicited emails are sent to those individuals who have never demanded for them. Main purpose of spam filters is to keep users' inbox free from spam emails. Many pitfalls are associated with this emailse.g. Consumes space in inbox, get mixed with important personal emails, use network bandwidth, and requires individual time and energy to sort through it [3]. Two significant approaches for classification were defined in paper [4]. First method is associated with the automated defined rules. The most common example of this system is rule based system. Rule based system is commonly used when classes are static they are easily separable on the basis of some common features. The second proposed system is on the basis of machine learning technique. In paper [5] criterion function is used for defining clusters of spam messages. In the above mentioned paper k-nearest neighbor algorithm is used to define criterion function. Main purpose of criterion function is to maximize the similarity between messages in clusters. Symbiotic Data Mining (SDM) [6] is a data mining approach that uses Content Based Filtering (CBF) and Collaborative Filtering (CF).In order to improve personalized filtering local filters are reused from diverse entities while privacy is maintained. Paper [7] defines spam filters effectiveness on the basis of Naïve Bayes and Neural Network. Accuracy and sensitivity matrices are used to evaluate results. According to that paper accurate results can be achieved using feed forward back propagation network algorithm. More accurate results mean it has high accuracy and more sensitivity. Mixed membership model on the basis of assumption at four different levels is used in Bayesian Approach for soft clusters and classification. In paper [8] –[11] to remove advertisement automatic anti-spam filtering becomes a key unit for junk filtering tools. [12] The writer of this paper has used distance measure for numerical and nominal data and then combines into one. In another defined approach firstly all numerical data is converted into nominal data. This nominal data is later used for the calculation of distance measure by using all variables. In this paper data is taken from different application domains and then the complexity and scalability of different algorithm is measured by testing their performance on taken data. Paper [13] defines that the spammers' social networks are identified by using spectral clustering which is based on high behavioral similarity. The data is taken from Project Honey Pot [14]. The conclusion of that paper states that "(1) phishing emails are either sent by spammers or no phishing emails at all, (2) phishing emails are either sent by the most communities of spammers or no phishing emails at all(3) numerous groups of spammers in groups show clear progressive activities by having comparable IP addresses". It is clear from paper [15]that both stated examples are comparable in generalization performance and both methods are assets efficient although large number of training sets are used. Bag of words created from different websites are used for spam classification.

2. EMAIL TEXT MINING

Techniques

Emails are a grouping of structured and unstructured information. Each email includes structured and standard data fields such as key receiver(s); copied recipient(s); sender; date and subject line. The text of the email is essentially unstructured. Minkovand notes that the [3] uniqueness of emails permit us to view a set of semi structured emails as a graph, with nodes representing actors, temporality, subject matter and meetings. We can exploit these characteristics by employing a variety of techniques as illustrated in Figure 1.

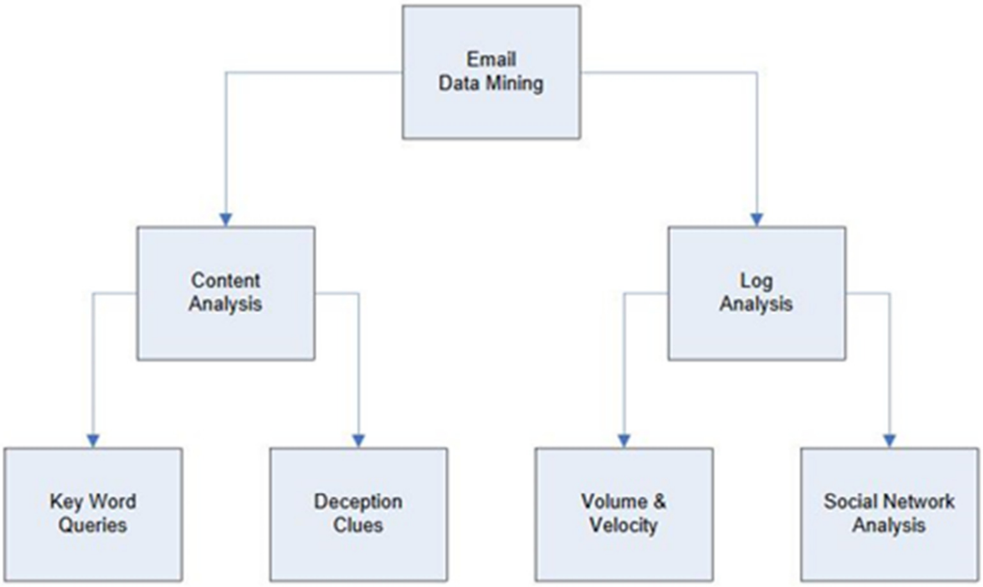

Figure 1 Email data mining framework.

The techniques, which will be argued in the subsequent paragraphs, can be alienated into two wide categories. The first group is those methods that text mines the content of emails and attachments to those emails. The second category spotlight on the email logs as distinct from the content of the emails. These techniques are not mutually exclusive and, as the remainder of this paper discusses, blending these techniques can recover the efficiency and helpfulness of locating fraud.

Content Analysis

Content analysis of printed resources has a long the past in a multiplicity of research areas together with language and literature, sociology, and information systems. Applying that body of research to email contents is an understandable extension of that research. The following paragraphs summarize key word queries and identifying deceptive emails. Weaved in those discussions are comments regarding the use of continuous monitoring as it could apply to email content analysis.

Keyword Searches

Emails could be searched for key words such as finder’s fee, bribe, kickback, and similar words that could indicate questionable actions or overrides of controls. This would be cherry-picking the naive fraudsters. It is hard to believe that a fraudster would use such words in the company email, but they do as found by researchers who have explored the Enron email corpus. Some of these people may falsely believe that company email has the same privacy protection as using the U.S. postal system to mail a letter. A simple query of hundreds of suspicious key words by the auditor would probably produce an overwhelming population of false positives. Like any database query, however, the false positives could be greatly reduced by adding more parameters to the query. For example, extracting emails with the suspicious key words AND sent to domains of vendors that the company does business with AND vendors who were granted a new contact during the fiscal year. This query might extract emails where an employee requested a “finder’s fee” for helping ensure that the company will win the contract. As another example, extracting emails with key words AND sent to generic domains (e.g., AOL, Hotmail, Gmail, etc.), which could indicate someone is trying to disguise his/her company affiliation. Probably the most discussed continuous email monitoring is the *Carnivore* system developed by the FBI to scan emails in the United States. The CIA and NSA are assumed to have similar systems to monitor email traffic outside of the U.S. The difference being that the FBI needs a court order before it can monitor a specific person’s email traffic in the U.S. By the way, companies do not need a court order to monitor employee emails. Currently, most companies automatically scan all emails moving through their email server for viruses and other malware. They also scan the content to capture spam. As such, scanning emails for fraud-related key words could be an extension of the current scanning process. There would be a performance hit because of the extra scanning, but the hit should be manageable even for more compound criteria (e.g., key word AND current vendor).

Deception Clues. One growing body of email mining research is deception research. Deception can take two forms. In the classic form of deception the sender is deceiving the recipient of the email. This form of deception could be an outright lie or could be the *normal* part of a negotiation strategy. The more subtle form of this type of deception is when there is collusion between the sender (e.g., an employee) and recipient (e.g., a vendor) and they are trying to deceive a third person (e.g., the employer) monitoring their emails. The second broad type of deception is when the email sender tries to disguise their identity.

Content Deception. In terms of the first type of deception introduced above, Keila and Skillicorn (2005) state that individuals who are trying to deceive generally include the following in their emails:

- Fewer first-person pronouns to dissociate themselves from their own words
- Fewer exclusive words, such as but and except, to indicate a less complex story
- More negative emotion words because of the sender’s underlying feeling of guilt
- More action verbs to, again, indicate a less complex story

In addition, according Skillicorn (2005) and other researchers, even senders who suspect that their emails may be monitored will alter their emails toward “excessive blandness.” According to these researchers, the deceivers are following these behaviors to reduce the cognitive demand of the deception. They want to disassociate themselves from their statements and keep their story simple because it is hard to remember all the details of a complex story.

The above parameters (first-person pronouns, exclusive words, negative emotion words, and action verbs) can be used to ranked emails based on the relative aggregate scores on those parameters. Then a specific person’s emails can be compared to his/her other emails to see if the scores have changed over time. A person’s emails can also be compared to his/her peers to determine if a person’s parameters are significantly different than the peers. To help conduct this type of deception analysis Pennebaker et al. (2001) developed software called Linguistic Inquiry and Word Count (LIWC)2.

Sender Deception. Another form of deception is the email sender trying to disguise his/her identity—whether in written documents, emails, forum postings, or blogs. Like a person's unique fingerprint, researchers have shown that people have unique writing styles or "writeprints" that includes "vocabulary richness, length of sentence, use of function words, layout of paragraphs, and keywords." (Li et al. 2006). To develop these writeprints, some researchers have developed a body of stylometric research (e.g., see McEnery and Oakes 2000). A wide variety of features can be used to define writeprint characteristics. According to Li et al. (2006) these features can be divided into the four following categories:

1. **Lexical features**. These features relate to characters and words that are used by a writer. For example, the lexical writeprint features could include the lists of words, the relative frequency of the words, and lengths of sentences that appear in a person's writings. Different researchers have used specific lexical characteristics to differentiate or to identify authors: sentence length and vocabulary richness (Yule 1944); a set of 50+ high frequency words (Burrows 1992); and just focus on two- and three-letter words and "vowel word" (words that start with a vowel) (Holmes 1998).

2. **Syntactic features**. These features relate to sentence structure. Part of the analysis of syntactic features includes the use of punctuation and function words. Function words (as opposed content or lexical words) could be articles (or determiners), prepositions, pronouns, auxiliary verbs, conjunctions, and particles. There are about 300 function words in the English language.3 A variety of studies have shown that analyzing syntactic feature was superior to analyzing lexical features alone. Holmes (1998) found that function words had good discriminating capability. Baayen (2002) found that including punctuation in the analysis also improved discriminating capability. Stamatatos et al. (2001) explored passive count and part-of-speech tags.

3. **Structural features**. These features relate the overall structure of the author's writing, which has been shown to be strong evidence of personal writing style. De Vel et al. (2001) achieved a high level of author identification performance analyzing emails.

4. **Content-specific features**. While the other three categories of features are essentially content free, how the author uses content keywords related to a specific topic has been shown to have additional discriminating power. If a researcher attempted to operationalize these categories of features by developing metrics to measure those features, the potential lists would be almost limitless. So, a first major step to comparing emails writeprints is developing a set of features that are discriminatory, measurable, and manageable. [20] used almost 1,000 writeprint features to analyze written materials. [18] created a taxonomy that included 270 features that broke down as follows:

- 87 lexical features
- 158 syntactic features
- 14 structural features
- 11 content-specific features

Adding more features to the analysis does not always improve discriminating power. de Vel et al. (2001) found that the performance of their analysis decreased when they increased the number of function words to 320 from 122.[16] used the generic algorithm form of heuristic search to find the optimum subset of features. Starting with 270 features introduced in the prior paragraph, they found the optimum subset for identifying message authors includes 134 features. Their finding of 134 features is not universal; instead the results will vary depending on the textual materials being analyzed and the language used by the authors. The population of textual messages that the researchers used were messages from the misc. for sale computers newsgroup that involved the selling of pirated software. When the researchers applied the general algorithm to similar messages in Chinese newsgroup messages, they started with 114 features and found the optimum subset was 56 of those features. Various forms of pattern recognition, neural networks, artificial intelligence, and other data mining techniques have been used to reduce very large feature sets down to optimal sub-sets. (See Liu and Motoda (1998) for a summary of feature selection studiesand their resulting general framework) In terms of applying these deception detection techniques to a continuous monitoring environment there will be two major tasks. These techniques use aggregated data to form a baseline and then new emails are compared to that baseline. As such, the first task will be creating that baseline and the main issue for this task is selecting the specific (optimum) set of features that will be used. Once the aggregated baseline is developed each email passing through the email server can be scanning to develop metrics that email that, in turn, will be compared to the baseline metrics.

The Challenges of Email Data Mining

Besides being free form and unstructured, emails are noisy, which makes them challenging for data mining. Although email is a written form of communications, senders rarely subject their emails to the same editing scrutiny that they do to formal written (paper-based) communications. As such, there are a variety of reasons for that noise, for example, including:

- Inconsistent use of abbreviations.
- Inconsistent capitalization of words.
- Misspelled words.
- Numbers sometimes spelled-out (e.g., one, two, thirteen) and other times numerical representations are used (e.g., 1, 2, 13).
- Missing and incorrectly used words.
- Incorrect grammar.
- The sender’s message frequently includes the prior sender’s email as a part of their replying email. Sometimes emails could have the complete discussion tread that includes several generations of replies and replies to replies in the same email.
- Inability to identify the identities of email participants and their relative roles and responsibilities.

Therefore, in general, any content analysis of email, as opposed to analyzing email logs, will have to be proceeded by significant email data cleaning, which will be a major challenge to attempts to create a system for the *continuous* monitoring of emails. The quoted text (replies to replies) within the emails will have to be removed so only the sender’s “new” material in the email is analyzed. Non-text information (e.g., line breaks, extra space, and other control characters) will have to be filtered out. The remaining text will need to be normalized. Tang et al. (2005) recommend a four-pass cascade approach. The first pass is non-text filtering, which is then followed by paragraph normalization, sentence normalization, and word normalization. Then the subsequent content analysis will be both more efficient and effective. This email cleaning that appears to be a required prerequisite to analyzing email content probably means that email content (as opposed to email logs) will not be analyzed on a true real-time basis. Instead, like the FBI’s *Carnivore*, copies of emails will be temporarily stored offline. Offline emails will be then be cleaned and analyzed. This will not be a major problem, meaning that a few minutes will elapse before a suspicious email is flagged and reported to an administrator or auditor.

3. EMAIL TEXT MINING TOOLS

In this part we discuss some text mining tools for monitor emails, and how tool help to monitor continues emails.

Email Content Monitoring

Numeral business text mining tools are developed in recent years to tolerate company security managers to watch the content of emails. the primary category of tools is general purpose network observance tools that are developed for the needs of security observance and assessment. Progressively vendor’s area unit providing email content observance as a by-product of spam or spyware assessment. For instance, eSoft Corporation’s Threat Wall manages virus and spam coming back into the entity however additionally undertakes content filtering “by scanning all emails for admin-defined keywords, phrases or regular expressions.” The package emails violations to directors. the same feature is found in Symantec’s Symantec Mail Security 8x00 Series appliances, that mix hardware and package in an exceedingly single device. As will be notional, given the root of Symantec, the first focus of the appliance is on virus defense and spams turning away, extra plug-ins for content observance also are out there. The second category of tools expressly monitors emails and alternative net communications. nowadays these tools area unit for the most part designed to stop losses of information science or breaches of compliance needs. The latter issue is especially vital for organizations that area unit subject to intensive privacy compliance needs. the wants beneath HIPAA for healthcare suppliers to keep up the privacy of their Corporation’s8 Vericept Content 360º tool is Associate in Nursing example of this comparatively new category of the package. in keeping with Vericept, “Vericept’s Content 360° visibility provides early detection of close at hand threats and corporate executive risk by observance all Internet-based communication and distinguishing areas of immediate money, name and legal risk. By correlating events and analyzing patterns of behavior, Vericept will facilitate organizations to spot an occurrence before it happens and take immediate action to safeguard against serious security breaches which will cause irreversible complete and name harm.” a noteworthy tool inside the Vericept suite is “Email Vericept Self-Compliance,” that permits the sender of the e-mail to spot the e-mail as being applicable. Taking a somewhat totally different approach, Reckoned Corporation9 has developed “iGuard Appliance” that scans networks, as well as emails, for sensitive knowledge. the main focus of the merchandise is that the protection of belongings and compliance. a comparatively new startup that has specifically addressed email observance is InBoxer; INC.10 while the primary merchandise of the corporation were within the anti-spam domain, a lot of recently the corporation has developed its “Anti-Risk Appliance.” in keeping with In Boxer the appliance attracts from the corporation’s “proprietary, subtle “language models” supported the manner words area unit unremarkably employed in order to spot patterns in text. Our technology comes from years of expertise within the speech recognition business. we tend to learned the way to distinguish between words that sound alike by analyzing the whole message.” curiously, In Boxer archives and indexes all email inside the corporation, permitting resultant searches for rhetorical or alternative functions easy. Shoppers area unit Associate in Nursing example of such a crucial compliance demand.

4. IMPLEMENTATION

The steps to implement Email Text Mining and Text Classification

- Data set is prepared by collecting a group of e-mails from the publicly available corpus of legitimate and phishing e-mails. Then the e-mails are labeled as legitimate and phishing correspondingly.

- Tokenization is performed to separate words from the e-mail by using white space (space, tab, newline) as the delimiter.
- Then the words that do not have any significant importance in building the classifier are removed. This is called stop word removal and stop words are words like a, the, that etc.
- Then stemming is performed to remove in flexional ending from the necessary words.
- Finally, the Term-Document-Frequency (TDF) matrix is created where each row in the matrix corresponds to a document (e-mail) and each column corresponds to a term (word) in the document. Each cell represents the frequency (number of occurrence) of the corresponding word in the corresponding document. Thus, each e-mail in the data set has been converted into an equivalent vector.
- Generally prior to the classification, dimensionality reduction techniques are applied to convert the long vector created in step 5. Feature selection or feature extraction techniques are used for dimension reduction and this improves the training time of the classifiers.
- Finally the classification model classifies the dataset into phishing and legitimate.

Steps to Text mining on email data classification

Step1: Load required packages

Step2: Load Email data from email folder to data frame

Step3: Create email corpus

Step4: Create document term and word count

Step5: Create to model to assess the email

Step6: Create corpus for training and test data

Step7: Email classification of spam and ham emails

5. CONCLUSIONS

In this paper, we discussed email data mining challenges and using text mining tools. As the Enron email corpus database are likely to include exchange of information among parties that will provide proof and context for matters that are subject to reassurance. And different email text mining techniques explained, content analysis, Key Word Searches, Structural features, content specific features and Syntactic features.

REFERENCES

[1] Androutsopoulos, I., Koutsias, J., Chandrinos, K. also, Spyropoulos, C. (2000), A trial examination of innocent bayesian and watchword based against spam _ltering with individual email messages, in `Proceedings of the 23rd yearly worldwide Special Interest Group on Information Retrieval (SIGIR) meeting on Research and improvement in data recovery', SIGIR '00, ACM, New York, NY, USA, pp. 160{167.

[2] Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Sakkis, G., Spyropoulos, C. also, Stamatopoulos, P. (2000), `Learning to _lter spam email: A correlation of a credulous bayesian and a memory-based approach', Computing Research Repository (CoRR) cs.CL/0009009.

[3] B alter, O. (2000), Keystroke level investigation of email message association, in Proceedings of the SIGCHI gathering on Human factors in processing frameworks, CHI 00, ACM, New York, NY, USA, pp. 105{112.

[4] Bellotti, V., Ducheneaut, N., Howard, M., Smith, I. what's more, Grinter, R. E. (2005), Quality versus amount: email-driven undertaking administration and its connection with over-burden', Hum.- Comput. Associate. 20, 89{138.

[5] Bickel, S. what's more, Sche_er, T. (2004), Learning from message sets for programmed email replying, in Proceedings of the European Conference on Machine Learning (ECML)', pp. 87{98.

[6] Bird, C., Gourley, A., Devanbu, P., Gertz, M. also, Swaminathan, A. (2006a), Mining email interpersonal organizations, in `Proceedings of the 2006 International Workshop on Mining Software Repositories', MSR '06, ACM, New York, NY, USA, pp. 137{143.

[7] Bird, C., Gourley, A., Devanbu, P., Gertz, M. also, Swaminathan, A. (2006b), Mining email interpersonal organizations in postgres, in `Proceedings of the 2006 International Workshop on Mining Software Repositories', MSR '06, ACM, New York, NY, USA, pp. 185{186.

[8] Blanzieri, E. what's more, Bryl, A. (2008), A overview of learning-based methods of email spam _ltering', Artif. Intell. Rev. 29, 63{92.

[9] Blei, D. M., Ng, A. Y. what's more, Jordan, M. I. (2003), Latent dirichlet portion, J. Mach. Learn. Res. 3, 993{1022.

[10] Boykin, P. O. what's more, Roychowdhury, V. P. (2004), Personal email organizes: An e_ective antispam apparatus, Computing Research Repository (CoRR) cond-tangle/0402143.Bradley, A. (1997), `The utilization of the zone under the ROC bend in the assessment of machine learning calculations', Pattern Recognition 30, 1145{1159.

[11] Breiman, L. (2001), Random timberlands', Mach. Learn. 45, 5{32.

[12] Breiman, L., Friedman, J., Stone, C. J. what's more, Olshen, R. A. (1984), Classi_cation and Regression Trees, 1 edn, Wadsworth and Brooks, Monterey, CA.

[13] Campbell, C. S., Maglio, P. P., Cozzi, A. what's more, Dom, B. (2003), Expertise identi_cation utilizing email interchanges, in Proceedings of the twelfth universal gathering on Information and learning administration, CIKM '03, ACM, New York, NY, USA, pp. 528{531.

[14] Carvalho, V. R. what's more, Cohen, W. W. (2008), Ranking clients for smart message tending to, in Proceedings of the IR look into, 30th European meeting on Advances in data recovery, ECIR'08, Springer-Verlag, Berlin, Heidelberg, pp. 321{333.

[15] Claburn, T. (2005), Spam costs billions, Website: http://www.informationweek.com/news/59300834.

[16] Cohen, W. (1996), Learning decides that group email, in Papers from the Association for the Advancement of Arti_cial Intelligence (AAAI) Spring Symposium on Machine Learning in Information Access, AAAI Press, pp. 1825.

[17] Cohen, W. W. (1995), Fast e_ective rule induction, in Proceedings of the Twelfth International Conference on Machine Learning, Morgan Kaufmann, pp. 115{123.

[18] Cormack, G. and Lynam, T. (2004), A study of supervised spam detection applied to eight months of personal e-mail.

[19] Venolia, G.D. and Neustaedter, C. (2003), Understanding sequence and reply relationships within email conversations: a mixed-model visualization, in Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '03), ACM, New York, NY, USA, pp. 361-368.

[20] Wang, M.-F., Jheng, S.-L., Tsai, M.-F. and Tang, C.-H (2011), Enterprise email classication based on social network features, in Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, 2011, IEEE Computer Society, Washington, DC, USA, pp. 532536.

[21] Myneni Madhu Bala, K. Navya and P. Shruthilaya, Text Mining on Real Time Twitter Data for Disaster Response. International Journal of Civil Engineering and Technology, 8(8), 2017, pp. 20–29.

[22] M. John Basha and Dr. K. P. Kaliyamurthie, Effective Linear-Time Document Clustering In Text Mining Using Web Document Categorization, International Journal of Civil Engineering and Technology, 8(10), 2017, pp. 224–234.

[23] Inje Bhushan V. and Prof. Mrs. Ujwalapatil, A Comparative Study on Different Types of Effective Methods in Text Mining: A Survey. International Journal of Computer Engineering & Technology (IJCET), Volume 4, Issue 2, March – April (2013), pp. 535-542

V.Sudheer Goud, Post Graduated in Master of Computer Application (**MCA**) from OU,1994 , Post Graduated in Master of Business Administration (**MBA**) from OU,2006, Post Graduated in Master of Computer Science & Engineering (**M.Tech**) from IETE , Hyderabad in 2013 and Pursuing Phd in Computer Science in ANU. He is currently working as an Associate Professor, Department of Computer Science in **Holy Mary Institute of Technology and Science** (HITS), (V) Bogaram, (M) Keesara, Medchal .Dist, Telangana, India. He has 23 years of Teaching Experience. His research interests include, Data Mining, Cloud Computing and Information Security.

P. Premchand, He is currently working as a Professor, Department of Computer Science and Engineering, University College of Engineering, Osmania University, Hyderabad, Telangana State, India.