Detecting Online Spams through Supervised Learning Techniques

M.S Minu, Kamagiri Mounika, N.Suhasini, Bezawada Tejaswi

Abstract: ***With more customers utilizing on the online review surveys to educate their administration basic leadership, assessment of reviews which economically affect the reality of organizations. Obviously, crafty people or gatherings have endeavored to manhandle or control online review spam to make benefits, etc, and that tricky recognition and counterfeit sentiment surveys is a subject of continuous research intrigue. In this paper, we clarify how supervised learning strategies are utilized to recognize online spam review surveys, preceding showing its utility utilizing an informational index of lodging reviews.***

Keywords- online review surveys, supervised learning, unlabeled data, Naïve bayes algorithm, classifiers, EM algorithm, Bag of Words, Stop word Filtering, Support Vector Machine classifier.

I. INTRODUCTION

Online review spamming is ending up progressively advanced and, in a few cases, sorted out, because of the possibility to benefit the organizations. For instance, a few organizations supposedly enlist online clients to post counterfeit sentiments. These organizations assess the sentiment of the customers about a particular product. These assessments can be utilized to advertise and advance a specific business, spread bits of gossip and harm the notoriety of a contending business. it is trying to distinguish counterfeit suppositions, as one may need to likewise get it the setting of the postings so as to decide if the specific assessment is misleading. For instance, by whatmethod can one dependably decide if the online survey postings about a specific business (for example surveys about an eatery) mirror the genuine encountered the clients who previously posted the surveys. One could, maybe, look into the online reviews and posting history of the clients and assure regardless of whether a specific client is posting.

Online review is an essential job in the present electronic business. It is alluring for a client to make a survey of items or stores before settling on the choice of what or from where to purchase. Because of the deceptive spam reviews, The clients can be misdirected to purchase low-quality The items, while better than average stores can be criticized by noxious surveys. We see that, in all actuality, an extraordinary bit (> 90% in the information we study) of the analysts compose just one survey (singleton audit). These reviews are so huge in number that they can nearly decide a store's appraising and impression. Be that as it may, existing strategies did not analyze this bigger piece of the surveys.

Revised Manuscript Received on November 05, 2019.

M.S Minu, Assistant Professor, Department Of Computer Science And Engineering, Srmist, Ramapuram, Chennai

Kamagiri Mounika, Department Of Computer Science And Engineering, Srmist, Ramapuram, Chennai

N.Suhasini, Department Of Computer Science And Engineering, Srmist, Ramapuram, Chennai

Bezawada Tejaswi, [4]department Of Computer Science And Engineering, Srmist, Ramapuram, Chennai

A question arises, how to recognize spam surveys in singleton reviews? We call this issue singleton review spam location.

To address this issue, we see that the ordinary analysts example is steady and uncorrelated to their rating design transiently. Conversely, spam assaults are normally burst and either emphatically or adversely connected to the rating. In this way, we propose to distinguish such assaults through bizarrely associated fleeting examples. We recognize and develop multidimensional time arrangement dependent on total measurements, so as to portray and mine such connections. Along these lines, the singleton review spam discovery issue is mapped to an anomalous associated design location issue. We propose a various leveled calculation to heartily identify the time windows where such assaults are probably going to have occurred. The calculation likewise pin points such windows in various time goals to encourage quicker human assessment. Trial results demonstrate that the proposed strategy is compelling in recognizing singleton review spam. We find that singleton review is a critical wellspring of spam surveys and to a great extent influences the appraisals of online stores.

As of late, the pattern of spam detection has distended on the grounds that anyone could compose spam surveys and gift them on internet primarily based business sites with no imperative. Some people compose phony surveys for his or her things and administrations; such people are known as as spammers. Spam reviews are commonly written therefore to gain a profit or to advance their things or administrations. One of the principle problems concerning feeling sharing sites is that spammers will while not abundant of a stretch build publicity concerning the precise item by composing spam reviews. In increasing the estimation of the item or administration item on the net, they for the foremost half visit the review phase to suppose alternative purchasers' criticism. Surveys are foremost half positive, the consumer could purchase, else they might not purchase that exact item. This all demonstrates spam surveys have was the first issue in net searching that will lead to a misfortune for each the consumer and therefore the producer. Survey spam will monetarily influence organizations and would possibly cause a sense of doubt within the general open, while not understanding this important issue, on-line survey destinations may grow to spot loaded with falsehoods and all futile .For example, Yelp and Amazon , have effectively gained some ground In any case, there's still a large amount of chance to urge higher in spam review.

Review spam can in like manner unfavorably influence associations as a result of incident in customer trust. The issue is outrageous enough to have pulled in the thought of overwhelming press and governments. For example, the BBC and New York Times have reported that "fake audits are transforming into a normal issue on the Web, and a photography association was starting late presented to

numerous injurious buyer overviews". In 2014, the Canadian Government gave a notice "asking clients to be cautious about fake online backings that give the inclination that they have been made by standard purchasers" and evaluated that 33% of each online review were phony Footnote. As audit spam is an inevitable and hurting issue, making procedures to help associations and buyers perceive fair reviews from fake ones is a critical, anyway testing issue.

In the composition, survey spam has been orchestrated into three social occasions, proposed by Dixit et al. (1) Untruthful Reviews - the essential stress, (2) Reviews on Brands - where the comments are simply stressed over the brand or the vendor of the thing and disregard to audit the thing, and (3) Non-Reviews - those studies that contain either insignificant substance or plugs. The essential order, untruthful studies, is of most stress as they undermine the reliability of the online survey system

Revelation of sort audit spam is a troublesome task as it is problematic, if surely possible, to perceive fake and certified reviews by physically getting them. To diagram the issue of this task, we consider a certifiable and fake model from the dataset made by Ott et al.. As a human judge it is difficult to absolutely figure out which survey is fake and which is authentic.

II. RELATED WORK

The Existing system is designed using the co-training algorithm and EM algorithm which handles large amount of unlabeled data. The recent researches mainly implies on unlabeled data due to the high cost of labeled data. The spam detection is generally resolved using classification techniques. These classification techniques are usually efficient if training set and test set is done on datasets containing both positive and negative review data.

The system of spam detection in wireless sensor networks is designed to reduce the problems with an expense of cryptographic activities to be perform on all messages. In this framework highlights like spam discovery and area mindful message verification instruments which are used to defend spam attacks in antinodes and wireless sensors. This method authentication performs all the time, it is limited to size of the threat. The study of social networks regarding the application of detecting compromised accounts is done and in this study approach, it takes both extroversive and introversive behavior in to the account. This is utilized to identify a bargained record, the social conduct can barely adjust to the legitimate client's social profile. The main drawback of this study is that there is no complete and accurate passion of behavioral profiles for detecting the sample facebook users**[5].** The spam detection is performed in the websites using the classification features and languages models. The spam detection was done in a very efficient way which reduces time and space complexity **[6].** The concept of Language model is used in this detection and the Language model (LM) approach gives the information from the web page to provide high quality indicators of web spam. LMs and QLs approach is used in this study to detect the web spams efficiently. But the current system is not intended to be a real-time application nevertheless analyzing the relationship between a page and those that point to it or disagreement between new source of information to improve performance of LM approach. In the study of half and half approach for identifying robotized spammers by amalgamating network based highlights, the separation intensity of various component classes is likewise broke down. In this both collaboration and network based highlights are seen as most discriminative for spammers discovery. The spammers are recognized when they are at exceptionally propelled stage, and it is hard to get their past logs information **[3].**

The phony sentiment surveys are made by spammers to make financial advantage for the association. The datasets utilized in certain assessments was "more extravagant" than recently utilized dataset as in it contains surveys with both positive and negative feelings. The spam detection is done usually done using minimal meta-data which is more complex**[1].** The spam detection is even done in short messaging service(SMS). These spams create disturbance in the whole SMS system and cause frustration to the users. The spam identification in SMS is normally done by looking at informational indexes and finding their impediments. The fundamental constraint of the exploration datasets that they are applied distinctly for grouping the substance **[2].**

In the hybrid approach for distinguishing computerized spammers in twitter, they present a half breed approach for identifying mechanized spammers by amalgamating metadata, substance and communication based highlights. The main advantage of this study is that the interactive based features are more effective the metadata based features. But examining log information may prompt wrong portrayal **[4]**.

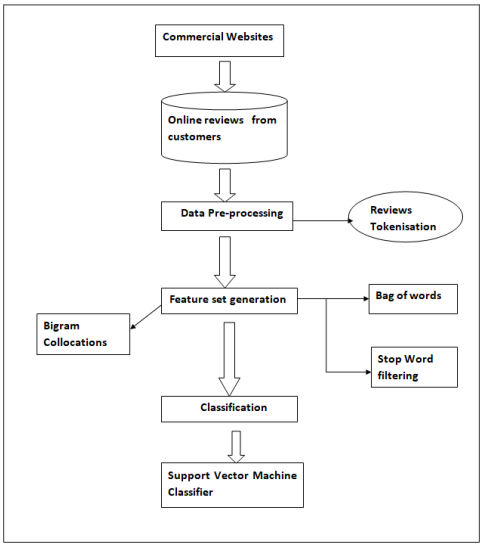

Fig 1.1 Architecture diagram for classification

III. PROPOSED SYSTEM AND EXPERIMENTAL SETUP

A. STEP-BY-STEP APPROACH FOR REVIEW SPAM DETECTION

The Review Spam Detection of the commercial websites has certain steps before they are are classified into spam and Non-spam. The first step towards the spam detection is the collection of data. The reviews from various commercial websites are collected and stored in the form of table in the database. The collected data is then pre processed to remove the anonymous and duplicate data. The pre processed data is then with the feature selection to remove the

frequent words and stop words. The simplified data is then classified using the Supervised Machine learning approach as seen in Fig 1.1 Architecture diagram of classification.
The features to extract the points used for the setup from the collected datasets.
1. Bags of words
2. Stop filtering
3. Bigram collocations
Three principle standard NLP preprocessing steps are considered in this paper including: stemming, accentuation marks evacuation, and stop-words expulsion. In Stemming, we acquire a stem type of each word in the dataset, which is a section of word to which joins can be appended. Stemming calculations are explicit to language and vary concerning execution and precision. Various methodologies could be utilized, for example, Affix evacuation stemming, n-gram stemming, and table query stemming. A significant NLP preprocessing step is accentuation marks expulsion, this marks - used to separate content into sentences, passages and expressions - influences the aftereffects of any content handling approach, particularly what relies upon the event frequencies of words and expressions, since the accentuation marks are utilized as often as possible in content rundown stop-words. The choice for the estimation of the limit worth is significantly influenced by the estimations of exactness and review. In a perfect world, we need both accuracy and review to be 1, yet this only occasionally is the situation. If there should be an occurrence of a Precision-Recall tradeoff we utilize the accompanying the contents.

FEATURE SELECTION APPROACH

Highlight choice is the way toward separating highlights from information so as to utilize them as an order criteria. Various kinds of highlights could be separated from content, for example, Bag-of-words. Stop word filtering and Bi-gram collocations and etymological highlights.

(i) BAG OF WORDS

The bag of words model is a disentangling portrayal utilized in normal language handling and data recovery (IR). In this model, (for example, a sentence or a record) is taken as pack(multi set) of its words, dismissing language and even word request yet keeping assortment. The sack of-words model has likewise been utilized for PC vision.
The bag of words model is ordinarily utilized in strategies for report arrangement where the (recurrence of) event of each word is utilized as an element for preparing a classifier. In Bag-of-words, highlights comprise of individual or gathering of words found in the content. At the point when this gathering of words comprises of n touching consecutive words, it called n-gram highlights. A case of Bag-of-words highlights is appeared in table 1, where each event of a word inside a survey will be spoken to by the recurrence of event of that word in the content. Presentation Business visionaries may push commentators to create incredible reviews about their things or organizations.

STOP WORD FILTERING

A stop word is a commonly used word, (for instance, "the", "an", "an", "in") that a web list has been redone to ignore, both when requesting sections for looking and when recuperating them as the result of a chase request.
We would not require these words consuming room in our database, or involving gainful dealing with time. For this, we can empty them adequately, by taking care of a summary of words that you consider to be stop words. NLTK (Natural Language Toolkit) in python has a summary of stop words set away in 16 unmistakable tongues. You can find them in the nltk_data vault. The stop word separating for the example surveys has been given below:

Sample text with stop words	Without stop words
"The Movie was really amazing"	Movie, really, amazing
"wonderful food and an amazing addition to our community"	Wonderful, food, amazing, addition, community
"The hotel is super luxurious with all comforts"	Hotel, super, luxurious, comforts

Fig 1.2 Example of stop word filtering

(ii) BI-GRAM COLLOCATIONS

Collocations are at least two words that will in general show up as often as possible together, for instance – United States. There are numerous different words that can come after United, for example, the United Kingdom and United Airlines. Similarly as with numerous parts of characteristic language preparing, setting is significant. Furthermore, for collocations, setting is everything.
On account of collocations, the setting will be a record as a rundown of words. Finding collocations in this rundown of words intends to discover basic expressions that happen oftentimes all through the content.

CLASSIFICATION PROCESS

The classification of reviews as spam and non spam is done after the feature selection process. The feature selection process makes the reviews simpler by removing the frequently occurring words and useless stop words.
The Classification process is done to detect the reviews as spam. In comparison among the classifiers for classifying the review spam, certain studies had shown that the Support Vector Machine Classifier is one of the best Classifier among the classifiers. The Support Vector machine classifier gives good performance and good accuracy measure for classification.

SUPPORT VECTOR MACHINE CLASSIFIER

Objective of content grouping is to arrange information into predefined classes. Here they are certain and negative classes. Content grouping is administered learning issue.
Initial phase in content arrangement is changing archive which is in string position into organization reasonable for learning calculation and grouping task. In data recovery it is discovered that word stem functions admirably as portrayal unit. This prompts credited worth portrayal of content. Each word relates to include with, number of times word happens in record, as its worth. Words are considered as highlights just on the off chance that they are not stop words (like "and", "or", and so forth). Scaling the component of highlight with IDF improves the performance.
SVM-Support vector machines are comprehensive students. Groundbreaking property of SVM is that their ability to learn can be free of dimensionality of feature space. SVM measures the multifaceted idea of Hypothesis reliant nervous that disconnects the plane and not number of highlights.
SVM has characterized info and yield position. Info is a vector space and yield is 0 or 1 (positive/negative).

Content archive in unique structure are not reasonable for learning. They are changed into organization which matches into contribution of AI calculation input. For this preprocessing on content records is conveyed out. Then we carryout transformation. Each word will relate to one measurement and indistinguishable words to same measurement. As referenced before we will see TF-IDF for this reason. Presently an AI calculation is utilized for figuring out how to order reports, for example making a model for info yield mappings.SVM has been demonstrated one of the amazing learning calculation for content categorization.

SVM CHARACTERISTICS

ML calculations commonly utilize a vector-space (property estimation) representation of models, for the most part the ascribes relate to words. Anyway word-sets or the situation of a word in the content may have impressive data, and essentially boundlessly numerous highlights can be developed which can improve characterization precision.

2. Classes are double, yet for the most part records are not alloted so correctly. Frequently a record D is said to have a place a little with class X1 and a piece to classification X2, however it doesn't fit well into any of the two. It presumably would require another class, as it isn't like any of the records seen previously.

3. Number of words increment on the off chance that we increment the quantity of reports. Stack's law depicts how the quantity of unmistakable words increment if number of report increments.

4. Portrayals use words as they are in writings. Nonetheless, words may have various implications, and various words may have a similar importance. The correct importance of a word can be controlled by its setting for example each word impacts the importance of its unique situation. In any case, the standard thing (computationally reasonable) portrayal disregard the request for the words. Assignment of SVM is to learn and sum up the info yield mapping. If there should be an occurrence of content arrangement information is set of archives and yield is their particular class. Consider spam channel as model info is an email and yield is 0 or 1(either spam or no spam).The Support vector machine classifier mainly deals with classification of datasets and finding the support vector points from the hyperplane.

SVM BENEFITS

High Dimension Input Space - while content grouping we need to manage numerous highlights (might be all the more then 1000). Since SVM utilizes over fitting protection, which doesn't rely upon number of highlights so they have capacity to deal with huge number of highlights.

Archive Vector Space - in spite of the high dimensionality of the portrayal, every one of the report vectors contain just a couple non-zero element.

More Text Categorization issues are straightly separable.

SVM EVALUATION

Content order frameworks may commit errors. To look at changed content classifiers for choosing which one is better, execution measures are utilized. A portion of these measures the exhibition on one paired class, others total per-classification measures, to give a general presentation. TP, FP, TN, FN are the quantity of genuine/false positives/negatives. The most significant per-classification measures for parallel classifications are

• Precision of svm classifier:

p=TP/(TP+FP)

• Recall measure is given by:

r= TP/(TP+FN)

The most significant midpoints are: miniaturized scale average[13], which checks each archive similarly significant, and full scale normal, which tallies every classification similarly significant.

Algorithm for Support Vector Machine Classifier

INPUT: SIMPLIFIED DATA OF REVIEWS FRO CUSTOMERS
OUTPUT: CLASSIFICATION OF SPAM AND NON-SPAM
Define feature as F and support vector point as S
S=Support vector point+1;
F=feature+1;
for S in range(1,n):
for each F of support vector point
Read String data
Float conversion;
Store values in arr_sv[S][F]
end for
for each F
Read string data
Float conversion
Store values in arr_test[F]
end for
for each S for each F
arr_a[F]+=arr_ay[s]*arr_S[S][F]
end for
Dist_Balue-=b
If(Dist_value>=th) then
return 1
else
return -1
end if

Classification Algorithm	Accuracy	Precision	Recall	F-Measure
Naïve Bayes Classifier	0.812	0.863372093	0.812	0.80511187
Maximum Entropy Classifier	0.784	0.849162011	0.784	0.77342911
Support Vector Machine Classifier	0.884	0.884024578	0.884	0.883

Fig1.3 Study of classifiers of single fold cross validation.

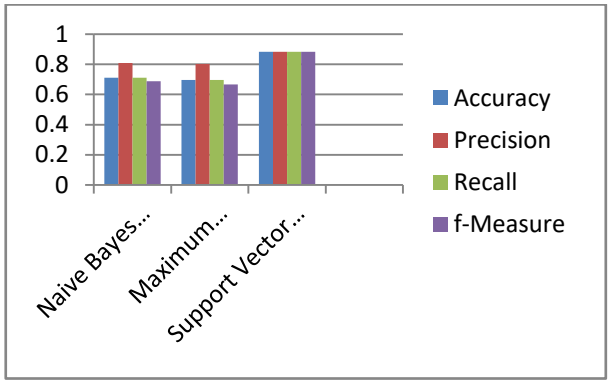

Fig1.4 Comparison of classifiers based on cross validation

Classification Algorithm	Accuracy	Precision	Recall	f-Measure
Naïve Bayes Classifier	0.712	0.808857808858	0.712	0.6875
Maximum Entropy classifier	0.696	0.801753867376	0.696	0.666806958474
Support Vector Machine Classifier	0.884	0.884221311475	0.884	0.883983293594

Fig 1.5 Study of Classifiers of N-fold cross validation

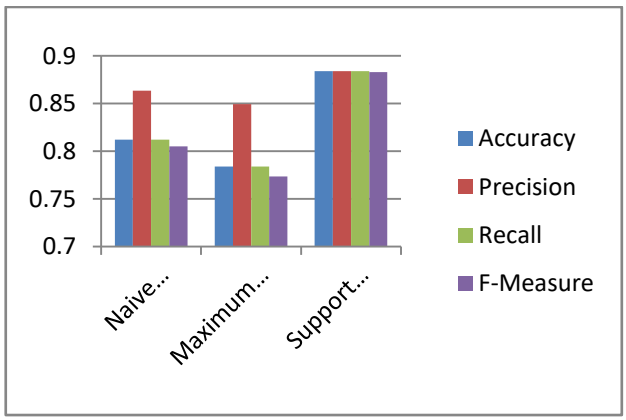

Fig 1.6 Comparison of classifiers based on single fold

IV. CONCLUSION

The outcome can be improved by expanding the preparation dataset size. Presently, the preparation dataset contains of complete 1000 audits (500 positive and 500 negative). This number can be expanded to check whether the addition improves exactness result.

In comparison with the other classifiers for classification of review spams, The above graph shows that the support vector machine classifier is better in terms of accuracy, precision, f measure and recall.

The precision result can likewise be improved by utilizing best words and best bigrams as list of capabilities rather than all words and all bigrams. 'Best' signifies the most as often as possible occuring words or bigrams. This methodology of dispensing with uninformed highlights (or, expelling boisterous information) is a sort of dimensionality decrease. Here is a decent instructional exercise on dispensing with uninformed highlights by making a list of capabilities of best words and best bigrams.

REFERENCES

1. J.K Rout et al, A.Dalmia, Kim-Kwang Raymond Choo,(Senior member ,IEEE),Sambit Bakshi,(Member,IEEE),Sanjay Kumar (Senior Member,IEEE),"Revisiting Semi-Supervised Learning for Online Deceptive Review Detection", in *Proc.special section on security and privacy in applications and services for future internet of things*, vol 5 2017.pp.Jan 18 2017,doi:10.1109/Access.2017.265503.
2. Shafi'I Muhammad Abdulhamid,Muhammad Shafie abd latiff, Haruna chiroma, Gaddafi Abdul-Salaam (Member,IEEE), "A Review on Mobile SMS Spam Filtering Techniques" in *Proc.*pp:Feb13,2017,doi:10.1109/ACCESS.2017.2666785.
3. Xin Ruan Zhenyu Wu, Member IEEE,Haining Wang, Senior Member, IEEE,and Sushil Jajodia, Fellow, IEEE, "Profiling Online Social Behaviors for Compromised Account Detection", in *Proc. IEEE Transactions on Information Forensics and Security*, DOI 10.1109/TIFS.2015.2482465.
4. Mohd Fazil and Muhammad Abulaish, *Senior Member, IEEE "*A Hybrid Approach for detecting Automated Spammers in Twitter" in *Proc. Transactions on Information Forensics and Security*, 1556-6013© 2018 IEEE,doi:10.1109/TIFS.2018.2825958,IEEE.
5. Vedat Coskun, Member, IEEE, Erdal Cayirci, Senior Member, IEEE, Albert Levi, Member, IEEE, and Serdar Sancak "Qurantine Region Scheme to Mitigate Spam Attacks in Wireless Sensor Networks" in *Proc.IEEE Transactions on Mobile Computing*, Vol 5, No. 8, August 2006,doi: 1536-1233/06/$20.00 © 2006 IEEE.
6. [6]Lourdes Araujo and Juan Martinez-Romo, IEEE,"Web Spam Detection: New Classification Features based on Qualified Link Analysis and Language Models" in *Proc.IEEE Transactions on information forensics and security*, vol 5, No.3, September 2010, doi: 10.1109/TIFS.2010.2050767,IEEE.
7. [7] Fei Zhang, *Student Member, IEEE*, "Adversarial Feature Selection Against Evasion Attacks" in *Proc.IEEE Transactions on Cybernetics*, pp 2015 IEEE.doi:10.1109/TYCB.2015.2415032.
8. [8]seGaradi, IEEE,"Email Classification Research Trends: Review and Open Issues" in *Proc. IEEE*, pp: Jun 18,2017, doi: 101109/ACCESS.2017.2702187
9. [9]Chao Chen, Yu Wang, Jun Zhang Senior Member IEEE,"Statistical Features-Based Real Time Detection of Drifted Twitter Spam" in *Proc. IEEE Transactions on Information Forensics and Security*, vol 12, No.4 ,Apr 2017 doi: 10.1109/TIFS.2016.2621888.
10. S.Khattak,N.R. Ramay, K.Riyaz Khan, "A Taxonomy of botnet behavior, Detection, and defense" in *Proc. IEEE Communications surveys and Tutorials*,Vol 16, No.2,Second Quarter 2014,doi:10.1109/SURv.2013.091213.00134.
11. Jian Zhang, Lian-Han Si-Ma, Bin-Qiang Wang, Jian-Kang Zhang and Yan-Yu Zhang, "Low-Complexity Receivers and Energy-Efficient Constellations for SPAD VLC Systems" in *Proc. IEEE Photonics and Technology Letters*, 2016 IEEE, doi:10.1109/LPT.2016.2572300.
12. Peng Zhang, Chuan Zhou, Peng Wang, Byron J.Gao,"E-Tree: An Efficient Indexing Structure for Ensemble Models on Data Streams" in *Proc.IEEE Transactions on knowledge and Data Engineering*, 1041-4347©2013,doi-10.1109/TKDE.2014.2298018.
13. Kentaroh Toyoda, (Member, IEEE), Mirang Park, Tomoaki Ohtsuki(IEEE), "Novel Unsupervised SPIters Detection Scheme by Automatically Solving Unbalanced Situation" in *Proc.IEEE*, pp. June 7,2017.doi:10.1109/ACCESS.2016.2642978.
14. Bartosz Kurlej and Michal Wozniak,"Active Learning Approach to concept drift problem" in *Proc. Oxford University Press* © The author 2011 doi:10.1093/jigpal/jzr011.
15. Zhizhou Yin, Fei Wang, Wei Liu, Member, IEEE," Sparse Feature Attacks in Adversarial Learning" in *Proc. Transactions on Knowledge and Data Engineering* © 2017 doi:10.1109/TKDE.2018.2790928.
16. Stavros Papadopoulos, Anastasios Drosou, Senior Member, IEEE, "A Novel graph Descriptor for the Detection of Biling-related Anomalies in Cellular mobile Networks" in *Proc. IEEE Transactions on Mobile*

Computing,doi:10.1109/TMC.2016.2518668.
17. Chao Chen, Yu Wang, Jun Zhang, Senior Member, IEEE, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam" in *Proc.IEEE Transactions on Information Forensics and Security*, vol 12, No.4, Apr 2017, doi:10.1109/TIFS.2016.2621888.
18. Yafeng Ren and Donghong ji,"Learning to Detect Deceptive Opinion Spam: A Survey" in *Proc.IEEE* April 13, 2019, doi:10.1109/ACCESS.2109.2908495.
19. Adrian Lara, Byrav Ramamurthy,"OpenSec: Policy-based Security using Software-defined Netwrking" in *Proc.IEEE Transactions on Network and Service Management,*© 2015 doi:10.1109/TNSM.2016.2517407.
20. Kuan Zhang, Student Member, IEEE, Xiaohui Liang, Member,IEEE, "PIF:A Personalized Fine Grained Spam Filtering Scheme With Privacy Preservation in Mobile Social Networks" in *Proc. IEEE Transactions on computational Social Systems*, vol 2,No.3 Sep 2015,doi: 10.1109/TCSS.2016.2519819.

AUTHORS PROFILE

M.S MINU is an assistant professor in the Department of Computer Science and Engineering at SRM Instiute of Science and Technology, Ramapuram, Chennai. She received M.E degree in computer science. Her research interests include data science, information security, parallel computing, privacy preservation techniques.

KAMAGIRI MOUNIKA is a B Tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Ramapuram, Chennai. Her research interests include network security and Robotics.

N.SUHASINI is a B Tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Ramapuram, Chennai. Her research interests include system security and mobile application security such as malware analysis/detection.

.

BEZAWADA TEJASWI is a B Tech Student in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Ramapuram, Chennai. Her research interests machine learning and social security and cloud computing**.**