# SPAM USER IDENTIFICATION WITH FAKE REVIEWS USING THRESHOLD LIMIT FOR SOCIAL MARKETING

**J HariBabu#1, R Ramya Sri#2, R Sravanthi#3, R Sindhu Madhavi#4, T Navya Sri#5, P Naga Sowmya#6**

#1Assistant Professor, DeptOf CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)
#2 Student, DeptOf CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)
#3 Student, DeptOf CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)
#4 Student, DeptOf CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)
#5 Student, DeptOf CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)
#6 Student, DeptOf CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)

## ABSTRACT

Many of the people rely on the social media before making a decision to purchase any product. Most of the public look for reviews or feedback. While reading the best review, there can possibilities of a fake review, so people never look for it. It is necessary to help people in identifying spam review, so that they can purchase the best product what they needed. Every website has a comment box given, so that any anomalous user can leave his/her feedback/ opinion ,this gives a grand opportunity for the spammer to compose a spam message. Spammer always used fake review either to promote the product or the demote the competitors. So this becomes very much necessary to filter out the spam review. A Netspam Framework has been introduced which act as heterogeneous information networks used to spam highlights by classifying based on user and review features. In this, we had used Latent Semantic Analysis algorithm and Sentiment Analysis algorithm for identifying positive and negative token view. Another Personalized Recommendation Algorithm is used for a search query.

**Keywords** – Social Media, Heterogeneous Information Network, Fake Review, Spammer, Spam Review, Semantic Analysis

## I. Introduction:

Online Social Media gateways assume an influential part in data proliferation which is considered as an imperative hotspot for makers in their publicizing efforts and additionally for clients in choosing items and administrations[1]. In the previous years, individuals depend a ton on the composed audits in their basic leadership procedures, and positive/negative. reviews empowering/debilitating them in their choice of items and administrations. Moreover, composed surveys additionally help specialist co-ops to improve the nature of their items and administrations. These reviews in this way have turned into an imperative factor in accomplishment of a

business while positive audits can bring benefits for an organization, negative audits can possibly affect validity and cause financial misfortunes [2]. The way that anybody with any personality can leave remarks as spam, gives an enticing chance to spammers to compose counterfeit reviews intended to delude clients' conclusion. These deceptive audits are then duplicated by the sharing capacity of online networking and spread over the web [3]. The reviews written to change clients' impression of how great an item or an administration are considered as spam and are regularly composed in return for cash. The general idea of the proposed structure is to demonstrate a given review dataset as a Heterogeneous Information Network (HIN) [19] and to outline issue of spam recognition into a HIN classification issue. Specifically, here display review dataset as a HIN in which audits are associated through various nodetypes, (for example, highlights and clients). A weighting calculation is then utilized to ascertain each element's significance (or weight). These weights are used to ascertain the final names for reviews utilizing both unsupervised and directed methodologies.

Online Social Media portals play an influential role in information propagation which is considered as an important source for producers in their advertising campaigns as well as for customers in selecting products and services [6] --- [9]. In the past years, people rely a lot on the written reviews in their decision-making processes, and positive/negative reviews encouraging/discouraging them in their selection of products and services. In addition, written reviews also help service providers to enhance the quality of their products and services. These reviews thus have become an important factor in success of a business while positive reviews can bring benefits for a company, negative reviews can potentially impact credibility and cause economic losses. The fact that anyone with any identity can leave comments as review, provides a tempting opportunity for spammers to write fake reviews designed to mislead users' opinion [10]. These misleading reviews are then multiplied by the sharing function of social media and propagation over the web. The reviews written to change users' perception of how good a product or a service are considered as spam , and are often written in exchange for money.

As shown in, 20% of the reviews in the Yelp website are actually spam reviews. On the other hand, a considerable amount of literature has been published on the techniques used to identify spam and spammers as well as different type of analysis on this topic . These techniques can be classified into different

Categories; some using linguistic patterns in text which are mostly based on bigram, and unigram, others are based on behavioral patterns that rely on features extracted from patterns in users' behavior which are mostly meta data basedand even some techniques using graphs and graph-based algorithms and classifiers .

Despite this great deal of efforts, many aspects have been missed or remained unsolved. One of them is a classifier that can calculate feature weights that show each

feature's level of importance in determining spam reviews. The general concept of our proposed framework is to model a given review dataset as a Heterogeneous Information Network (HIN) and to map the problem of spam detection into a HIN classification problem. In particular, we model review dataset as a HIN in which reviews are connected through different node types (such as features and users). A weighting algorithm is then employed to calculate each feature's importance (or weight). once the spam user is identified user can be blocked by the admin . The blocked spam user cannot perform login again so that we can avoid further spam actions from him.

In outline, the fundamental commitments are as per the following: (I) Here propose NetSpam structure that is a novel network - based approach which models reviews organizes as heterogeneous information systems. The classification step utilizes distinctive metapath writes which are creative in the spam identification area. (ii) another weighting technique for spam highlights is proposed to decide the relative significance of each component and shows how viable every one of highlights are in recognizing spams from ordinary surveys. Asclarified in the unsupervised approach, NetSpam can find highlights significance even without ground truth, and just by depending on metapath definition and in view of qualities ascertained for each survey. (iii) NetSpam enhances the precision contrasted with the stateof-the workmanship as far as time many-sided quality, which very depends to the quantity of highlights used to recognize a spam audit; consequently, utilizing highlights with more weights will brought about distinguishing counterfeit surveys less demanding with less time intricacy.

## II. RELATED WORK

In [1] authors going for giving a proficient and compelling strategy to recognize review spammers by consolidating social relations in view of two suspicions that individuals will probably consider reviews from those associated with them as reliable, and review spammers are less inclined to keep up a substantial relationship coordinate with ordinary clients. The commitments of this are twofold: (1) We expound how social connections can be joined into audit rating forecast and propose a trust-based rating expectation demonstrate utilizing nearness as put stock in weight; and (2) We outline a trust-mindful identification showin terms of rating fluctuation which iteratively ascertains client particular general dependability scores as the marker for spamicity. Since not every single online survey are honest and reliable, it is vital to create strategies for recognizing audit spam. By extricating significant highlights from the content utilizing Natural Language Processing (NLP), it is conceivable to lead audit spam discovery utilizing different machine learning procedures used in [2]. Moreover, commentator data, aside from the content itself, can be utilized to help in this procedure. In this paper, we overview the unmistakable machine learning systems that have been proposed to take care of the issue of audit spam recognition and the execution of various methodologies for order and discovery of survey spam.In [3] authors

propose utilizing unsupervised oddity discovery systems over client conduct to recognize possibly awful conduct from typical conduct. Here presenting a procedure in view of Principal Component Analysis (PCA) that models the conduct of typical clients precisely and distinguishes noteworthy deviations fromit as abnormal. It tentatively approved that typical client conduct (e.g., classifications of Facebook pages preferred by a client, rate of like movement, and so forth.) is contained inside a low-dimensional subspace agreeable to the PCA strategy.By utilizing the perplexing conditions among audits, clients and IP addresses, in[ 4] authors initially proposed an aggregate arrangement calculation called Multi-wrote Heterogeneous Collective Classification (MHCC) and afterward extend it to Collective Positive and Unlabeled learning (CPU). Results demonstrate that the proposed models can particularly enhance the F1 scores of solid baselines in both PU and non-PU learning settings. Since the models just utilize dialect free highlights, they can be effectively summed up to different dialects.In [5] authors expect to distinguish clients creating spam audits or review spammers. It recognized a few trademark practices of review spammers and model these practices to identify the spammers. Specifically, try to display the accompanying practices. To start with, spammers may target particularitems or item bunches keeping in mind the end goal to expand their effect. Second, they tend to go amiss from alternate analysts in their appraisals of items. Here propose scoring techniques to quantify the level of spam for every commentator and apply them on an Amazon survey dataset. At that point select a subset of exceedingly suspicious analysts for encourage examination by the client evaluators with the assistance of an online spammer assessment programming uncommonly created for client assessment tests. In [6] authors proposed a novel idea of a heterogeneous review chart to catch the connections among commentators, reviews and stores that the analysts have checked on. Here investigate how communications between hubs in this diagram can uncover the reason for spam and propose an iterative model to distinguish suspicious commentators. This is the first run through such unpredictable connections have been distinguished for survey spam location. It additionally builds up a viable calculation strategy to measure the trustiness of analysts, the genuineness of audits, and the dependability of stores. Unique in relation to existing methodologies, it didn't utilize survey content data. So the model is along these lines integral to existing methodologies and ready to discover more troublesome and unpretentious spamming exercises, which are settled upon by human judges after they assess our outcomes. In [7] authors build up a deliberate technique to consolidation, analyze, and assess surveys from different facilitating locales. It centered around lodging surveys and utilize in excess of 15 million audits from in excess of 3.5 million clients spreading over three noticeable travel destinations. This work comprises of three pushes: (a) create novel highlights equipped for recognizing cross-site disparities adequately, (b) direct seemingly the principal broad investigation of cross-site varieties utilizing genuine

informationand build up a lodging character coordinating strategy with 93% precision, (c) present the TrueView score, as a proof of idea that cross-site examination can better advise the end client. This work is an early exertion that investigates the focal points and the difficulties in utilizing numerous auditing destinations towards more educated basic leadership. In [8] authors adopt an alternate strategy, which abuses the burstiness idea of reviews to distinguish review spammers. Blasts of audits can be either because of sudden prominence of items or spam assaults. Commentators and surveys showing up in a burst are frequently related as in spammers tend to work with different spammers and honest to goodness analysts has a tendency to seem together with other honest to goodness commentators. This prepares for us to manufacture a system of commentators showing up in various bursts. Then display analysts and their co occurrence in blasts as a Markov Random Field (MRF),and utilize the Loopy Belief Propagation (LBP) strategy to deduce whether a commentator is a spammer or not in the chart. It likewise proposed a few highlights and utilize include actuated message going in the LBP structure for arrange surmising. Here further propose a novel assessment strategy to assess the distinguished spammers naturally utilizing administered grouping of their audits. Furthermore, utilize space specialists to play out a human assessment of the recognized spammers and non-spammers. In [9] , exploration is a stage forward in enhancing the precision of recognizing abnormality in an information chart speaking to availability between individuals in an online interpersonal organization. The proposed mixture strategies depend on fluffy machine learning methods using distinctive sorts of auxiliary information highlights. The techniques are exhibited inside a multi-layered structure which gives the full prerequisites expected to discovering irregularities in information charts created from online interpersonal organizations, including information demonstrating and investigation, marking, and assessment.In [10] authors misuse machine learning techniques to recognize survey spam. Around the end, physically fabricate a spam accumulation from crept audits. At first dissect the impact of different highlights in spam distinguishing proof. It likewise watched that the review spammer reliably composes spam. This gives another view to recognize audit spam: it can distinguish if the creator of the survey is spammer. In [11] authors proposed another comprehensive approach called SPEAGLE that uses pieces of information from all metadata (content, timestamp, rating) and in addition social information (system) tackle them all in all under a *unified* structure to spot suspicious clients and surveys, and in addition items focused by spam. In addition, our technique can effectively and flawlessly incorporate semi-supervision, i.e., a (little) arrangement of marks if accessible, without requiring any preparation or changes in its hidden calculation.

## III. LITERATURE SURVEY

### Estimating the Prevalence of Deception in Online Review Communities

Author: MyleOtt,ClaireCardie and Jeff Hancock Year: 2019

Description: Presents a general framework for estimating the prevalence of deception in online review communities.

**Spotting fake reviews via collective PU learning**

**H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao . Year: 2019**

Description: It propose models that can markedly improve the F1 scores of strong baselines in both PU and non-PU learning settings. Since our models only use language independent features, they can be easily generalized to other languages.

**Opinion spam and analysis**

**N. Jindal and B. Liu. Year: 2018**

Description: This paper analyzes spam activities and presents some novel techniques to detect them.

## IV. IMPLEMENTATION

**Admin**

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as adding Categories, Adding Products for that Categories, Viewing and authorizing users, View Spam accounts details,Viewing friend request & response, All recommended posts, All posts with all Reviews, All Positive and Negative Reviews , Removing Products, Viewing All Purchased Products, viewing Positive and Negative Reviews Chart on products.

**Adding Categories**

In this module, the admin adds the category details such as category name. These details will be stored into the database.

**Adding Products**

In this module, the admin adds Product posts for categories which include details such as, product image, product name, cost, description and uses of that product. These details will be stored into the database. These details will be further searched and accessed by the users in order to recommend to their friends and to buy products.

**Authorize Users**

In user's module, the admin can view the list of users who all registered. In this, the admin can view the users' details such as, user name, email, address, phone number and authorize the users.

**Request & Response**

In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user image, requested user name, user name request to, status and time & date. If the user accepts the request then the status will be changed to accepted or else the status will remains as waiting.

**All Recommended Posts**

In this module, the admin can view all the recommended products. If any recommendations happened for particular products, those details will be shown along with products. Details include product name,

recommended user name, user recommended to name and the date.

**View Positive /Negative Comments**

In this, the admin can view all posts with their Positive and Negative Comments posted by users based on their opinions.

**Positive**: If the user comment contains at least one of the word which is listed in positive words, then that comment will be treated as positive comment.

**Negative**: If the user comment contains at least one of the word which is listed in negative words, then that comment will be treated as negative comment.

**All Comments on Products**

In this module, the comments of all posts will be displayed. Comments includes Positive, Negative, Non-Positive and Non-Negative. It includes details such as, commented user name, comment and date.

**All Purchased Products**

In this module, the products which are purchased by users will be displayed. It includes details such as, purchased user name, purchased products, price of the products and the date of purchase.

**Positive Comments Chart**

In this module, the number of positive Reviews got by the particular product will be displayed in a chart.

**Negative Comments Chart**

In this module, the number of negative Reviews got by the particular product will be displayed in a chart.

**Deleting/Removing Products**

In this module, the products which have got the negative comments from more than five users will be listed and removed by the admin.

**User**

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like viewing their profile Account details like Spam or Normal , search users and send friend request, viewing friend requests, searching posts and recommend to friends and viewing all product recommendations sent to him by his friends, commenting on posts, purchasing products and viewing their product search history.

**Search Users**

The user can search the users based on names and the server will give response to the user like User name, user image, E mail id, phone number and date of birth. If you wish to send friend request to particular user then click on “request” button, then request will be send to that particular user.

**Searching Products and Recommend to Friends**

In this, the user searches for products based on the products description. The user can recommend searched products to his friends, comment on post and he can add the products to cart to buy those added products later by using their created account.

**View Friend Requests**

In this module, the user can view the friend requests which are sent by other users. Which includes request sent user details with their tags such as user name, user image, date of birth, E mail ID, phone number and Address and user can accept the request by clicking on the “waiting” link.

**View Product Recommends**

In this module, the user can view all the products which are recommended by his friends. This includes recommended user name and his image, recommended products details.

**View Product Search History**

In this module, the user can view all the searched products names and categories, the keywords which he used to search the products. This includes details such as, searched product, used keyword and date of search.

**View Bank Account Details**

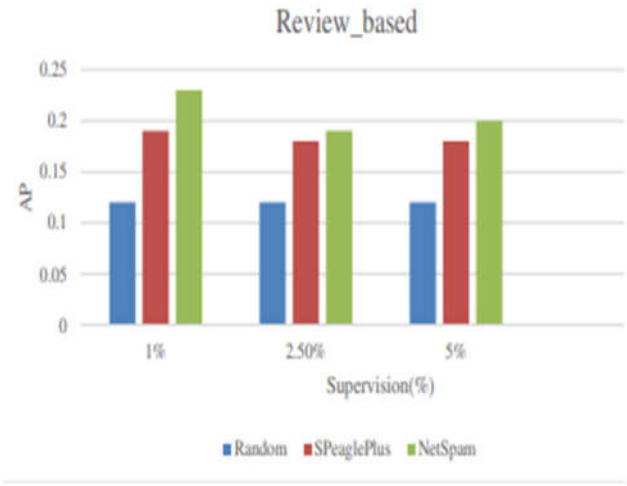
In this module, the user can create his bank account by providing details such as, account number, branch, address, email id. Later he can add money to his account and can view his account details.

## V. Results

In this area, we assess NetSpam from alternate point of view and contrast it and two different methodologies, Random approach and SPeaglePlus . To contrast and the first one, we have built up a system in which surveys are associated with each other haphazardly. Second approach utilize a wellknown chart based calculation called as "LBP" to figure final marks. Our perceptions indicate NetSpam, beats these current strategies. At that point investigation on our perception is performed and finally we will look at our structure in unsupervised mode. Finally, we examine time multifaceted nature of the proposed structure and the effect of camouflage system on its execution [23].

Accuracy: The execution as far as the AP and AUC. As it's appeared in the greater part of the four datasets NetSpam outflanks SPeaglePlus uncommonly when number of highlights increment. What's more extraordinary supervisions have no extensive impact on the metric esteems neither on NetSpam nor SPeaglePlus [22].

Results likewise demonstrate the datasets with higher level of spam surveys have better execution since when portion of spam audits in a specific dataset builds, likelihood for an audit to be a spam audit increments and thus more spam audits will be marked as spam audits and in the aftereffect of AP measure which is exceptionally reliant on spam rate in a dataset. Then again, AUC measure does not fluctuate excessively, on the grounds that this metric isn't reliant on spam audits rate in dataset, however on the final arranged rundown which is computed in view of the final spam likelihood.

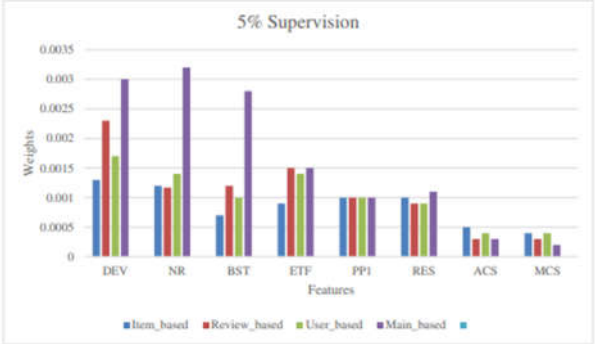



Feature Weights Analysis: Next we talk about highlights weights and their association to decide spamicity. In the first place we investigate the amount AP and AUC are subject to variable number of highlights. Dataset Impression on Spam Detection: As we clarified beforehand, unique datasets yield diverse outcomes in light of their substance. For all datasets and most weighted highlights, there is a sure arrangement for highlights weights. As is demonstrated inFigure3 for four datasets, in all of them,features for the Main dataset have more weights and highlights for Review-based dataset remain in the second position. Third position has a place with User-based dataset and finally Item-based dataset has the base weights (for in any event the four highlights with generally weights).

Feature Weights Importance: There are couple of features which are more weighted than others. Mix of these highlights can be a decent clue for acquiring better execution. The aftereffects of the Main dataset demonstrate all the four behavioral highlights are positioned as first includes in the final general weights. What's more, as appeared in the Review-based and additionally other two datasets, DEV isthemostweighted include. This is likewise same for our second most weighted element, NR. From the third component to the last element there are distinctive request for the specified highlights. The third component for both datasets User-based and Review-based is same, ETF, while for the other dataset, Item-based, PP1 is at rank 3. Going further, we see in the Review-based data set all four most weighted highlights are behavioral-based highlights which demonstrates how much this sort of highlights are essential in recognizing spams as recognized by different functions too [12], [20]. There is a solid relationship between's highlights weights and the exactness. For the Main dataset we can see this relationship is considerably more clear and furthermore pertinent. Ascertaining weights utilizing NetSpam help us to see how much an element is powerful in distinguishing spam audits; since as much as their weights increment two measurements including AP and AUC additionally increment separately and in this way our structure can be useful in identifying spam surveys in light of highlights significance.

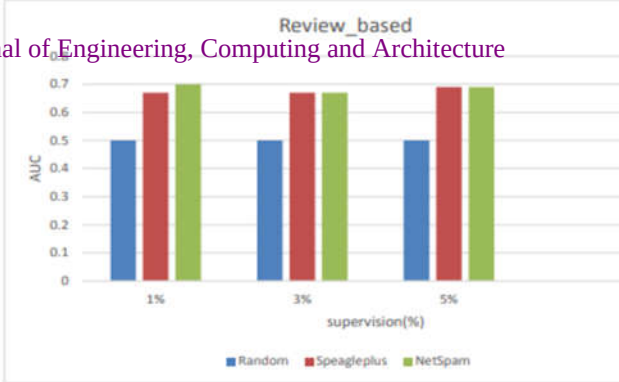
The perceptions show bigger datasets yield better connection between's highlights weights and furthermore its precision in term of AP. Since we have to know each element rank and significance we utilize Spearman's rank relationship for our work. In this experience our fundamental dataset has relationship esteem equivalent to 0.838 (p-value=0.009), while this incentive for our next dataset, User-based one, is equivalent

to 0.715 (p-esteem = 0.046). As much as the extent of dataset gets littler in the analysis, this esteem drops. This issue is more evident in Item and Review-based datasets. For Item-based dataset, relationship esteem is 0.458 which is low, since examining Item-based dataset needs Item-based highlights. The highlights are indistinguishable to every thing and are like client based highlights. At long last the acquired outcomes for our littlest dataset is fulfilling, on the grounds that final comes about considering AP demonstrate a connection close to 0.683 amongst weights and exactness (comparable outcomes for SPeaglePlus too). Weights and exactness (as far as AP) are totally corresponded. We watched values 0.958 (pvalue=0.0001), 0.764 (p=0.0274), 0.711 (p=0.0481) and 0.874 (p=0.0045) for the Main, User-based, Item-based and Reviewbased datasets, separately. This outcome indicates utilizing weight count technique and considering metapath idea can be compelling in deciding the significance of highlights. Comparable outcome for SPeaglePlus additionally demonstrates our weights count strategy can be summed up to different structures and can be utilized as a primary part to finding each element weight.Our outcomes additionally show highlight weights are totally reliant on datasets, considering this reality two most vital highlights in all datasets are same highlights. This implies with the exception of the first two highlights, different highlights weights are exceptionally factor regrading to dataset utilized for extricating weights of highlights.




Unsupervised Method: One of the accomplishment in this examination is that even without utilizing a prepare set, we can in any case find the best arrangement of highlights which respect the best execution. As it is clarified in Sec. IIIAn, in unsupervised approach unique plan is utilized to ascertain central marks and next these names are utilized to compute the highlights' weight and finally audit names. Our perceptions appear there is a decent relationship in the Main dataset in which for NetSpam it is equivalent to 0.78 (p-value=0.0208) and for SPeaglePlus this esteem achieve 0.90 (p=0.0021). As another case for client-based dataset there is a connection equivalent to 0.93 (p=0.0006) for NetSpam, while for SPeagle this esteem is equivalent to 0.89 (p=0.0024). This perception demonstrates NetSpam can organize highlights for the two structures.

Time Complexity: If we consider the Main dataset as contribution to our system, time multifaceted nature with these conditions is equivalent to □(□ଶ□) where □ is number of edges in made system or audits number. It implies we have to check if there is a metapath between a specific hub (audit) with different hubs which is □(□ଶ ) and this checking must be rehashed for extremely highlight. Along these lines, our chance




multifaceted nature for offline mode in which we give the Main dataset to system and ascertain spamicity of entire surveys, is □(□𝑚□) where m is number of highlights. In online mode, an audit is given to NetSpam to see whether it is spam or not, we have to check if there is a metapath between given survey with different surveys, which is in □(□), and like offline mode it must be rehashed for each component and each esteem.

The Impact of Camouflage Strategy: One of the difficulties that spam identification approaches confront is that spammers frequently compose non-spam audits to conceal their actual character known as camouflage [10], [12], [14]. For instance they compose positive surveys for good restaurant or negative reviews for low-quality ones; hence each spam indicator framework neglects to distinguish this sort of spammers or possibly has some inconvenience to spot them. In the past examinations, there are diverse methodologies for dealing with this issue. For instance, the writers expect there is dependably a little likelihood that a decent audit composed by a spammer and put this suspicion in its similarity lattice. In this examination, we endeavored to deal with this issue by utilizing weighted metapaths.

## VI. CONCLUSION

This novel spam identification system to be specific NetSpam in light of a metapath idea and another chart-based strategy to mark surveys depending on a rank-based naming methodology. The execution of the proposed system is assessed by utilizing two certifiable marked datasets of Yelp and Amazon sites. The perceptions demonstrate that ascertained weights by utilizing this metapath idea can be exceptionally viable in distinguishing spam audits and prompts a superior execution. Likewise, it is discovered that even without a prepare set, NetSpam can figure the significance of each component and it yields better execution in the highlights expansion process, and performs superior to anything past works, with just few highlights. Additionally, in the wake of defining four principle classes for highlights our perceptions demonstrate that the surveys behavioral classification performs superior to anything different classes, regarding AP, AUC and in addition in the figured weights. The outcomes likewise confirm that utilizing diverse supervisions, like the semi-administered technique, have no perceptible impact on deciding the vast majority of the weighted highlights, similarly as in various datasets.

## REFERENCES

[1] H. Xue, F. Li, H. Seo, and R. Pluretti, “Trust-Aware Review Spam Detection”, IEEE Trustcom/ISPA, 2015

[2] M. Crawford, T. D. Khoshgoftar, J. N. Prusa, A. Al. Ritcher, and H. Najada, “Survey of Review Spam Detection Using Machine Learning Techniques”,Journal of Big Data. 2015..

[3] B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks", In USENIX, 2014 .

[4] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective PU learning",In ICDM, 2014.

[5] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors",In ACM CIKM, 2010.

[6] G. Wang, S. Xie, B. Liu, and P. S, " Review graph based online store review spammer detection",IEEE ICDM, 2011.

[7] A. J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, " Trueview: Harnessing the power of multiple review sites", In ACM WWW, 2015.

[8] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh ," Exploiting burstiness in reviews for review spammer detection", In ICWSM, 2013.

[ 9] R. Hassanzadeh, "Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic", Queensland University of Technology, Nov. 2014.

[10] F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam",Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.

[11] R. Shebuti and L. Akoglu, "Collective opinion spam detection: bridging review networks and metadata",In ACM KDD, 2015.

[12] J. Fontanarava, G. Pasi , and M. Viviani, "Feature Analysis for Fake Review Detection through Supervised Classification," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 2017, pp. 658-666.

[13] S. K. Chauhan, A. Goel, P. Goel, A. Chauhan and M. K. Gurve, "Research on product review analysis and spam review detection," 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, 2017, pp. 390-393.

[14] Peng, Qingxi&Zhong, Ming. (2014). Detecting Spam Review through Sentiment Analysis. Journal of Software. 9. . 10.4304/jsw.9.8.2065-2072.

[15] S. P. Rajamohana, K. Umamaheswari, M. Dharani and R. Vedackshya, "A survey on online review SPAM detection techniques," 2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), Coimbatore, 2017, pp. 1-5

[16] S. Shehnepoor, M. Salehi, R. Farahbakhsh , and N. Crespi, "NetSpam: A Network-Based Spam Detection Framework for Reviews in Online Social Media," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 7, pp. 1585-1595, July 2017.

[17]. S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In ACM KDD, 2012.

[18]. G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. IEEE ICDM, 2011.

[19]. Y. Sun and J. Han. Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE, 2012.

[20]. A. Mukerjee, V. Venkataraman, B. Liu, and N. Glance. What Yelp Fake Review Filter Might Be Doing?, In ICWSM, 2013.

[21] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. Int. Conf. Comput. Commun. (INFOCOM), Apr. 2011, pp. 829–837.

[22] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Systems, vol. 23, no. 8, pp. 1467–1479, 2012.

[23] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE Int. Conf. Distrib. Comput. Syst. (ICDCS), Jun. 2010, pp. 253–262.

**Authors Profile**

**J HariBabu, M.Tech.,** working as an Asst.Professor in the Department of Computer Science & Engineering, Qis Institute of Technology, Ongole, Andhra Pradesh, India.

**R Ramya Sri,** Pursuing B Tech in the Department of Computer Science & Engineering, Qis Institute of Technology, Ongole, Andhra Pradesh, India.

**R Sravanthi, P**ursuing B Tech in the Department of Computer Science & Engineering, Qis Institute of Technology, Ongole, Andhra Pradesh, India.

**R SindhuMadhavi,**Pursuing B Tech in the Department of Computer Science & Engineering, Qis Institute of Technology, Ongole, Andhra Pradesh, India

**T Navya Sri,** Pursuing B Tech in the Department of Computer Science & Engineering, Qis Institute of Technology, Ongole, Andhra Pradesh, India.

**P Naga Sowmya,** pursuing B Tech in the Department of Computer Science & Engineering, Qis Institute of Technology, Ongole, Andhra Pradesh, India.