JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Block Chain Based Secure Spam Review Detection Using Machine Learning Techniques

Author 1 Sinthuja M, Author 2 Mr.Jayakumar D, Author 3 Mr. Ram Kumar M.O

Designation 1
Student
IFET College of Engineering Villupuram
Designation 2
Associate Professor
IFET College of Engineering, Villupuram
Designation 3
Associate Professor
IFET College Of Engineering, Villupuram

ABSTRACT

Online reviews of various products or have become an important source for determining public opinion. Therefore, tradesmen and supplier are anxious about testimonials, because they directly affect their business. Unfortunately, Over the past few years, the problem of spam review detection has received a lot of observation among communities and researchers, but experiments with real, bulk amounts of spam datasets are still needed. This can encourage to examine the review which is widespread in online reviews. In this work, two different spam detection methods are proposed: (1) Spam audit detection using client behaviours it uses the behavioral characteristics of thirteen different spammers to calculate a spam score, which is then used to identify spammers and spam. reviews and (2) Spam Review Detection Using a Linguistic Method (SRD-LM) operates on review content and uses transformations, feature selection, and classification to identify spam reviews. Experimental evaluations are performed using a real Amazon review dataset, analyzing 30.2 million analysis and 20 million analysis. Research show thattwo advanced models notably improved the process of detecting junk evaluations. In particular, behavior method achieved an accuracy of 95%, while the linguistic method achieved a spam scan detection accuracy of 80.1%. In comparison, Behavior method achieved better accuracy because it uses the rich features of spammers' behavior in the validation dataset, which provides a comprehensive analysis of spammers' behavior. To our knowledge, this is the first study of its kindthat uses real-world survey data to analyze the behavioral characteristics and linguistic methods of different spammers using different available classifications.

Keywords— Internet analysis, feature detection, linguistic characteristics, and behavioral characteristics.

1 INTRODUCTION

The information superhighway has emerged as the primary platform for human expression in the modernera. Using e-commerce sites, forums, and blogs, people may quickly express their thoughts on any product or service. Nowadays, everyone online is aware of how crucial these online reviews are for buyers and sellers. The majority of consumers read product reviews and services before investing on anything. The tendency of spamming have been increased recently since every user can write their ownreviews and post them online without any restrictions. Anybody may book someone to fabricate reviewsof their products or services.. Spammers are those folks. Most of the reviews are made for financial gainor to advertise a certain item or service. The main issue with review websites is that it's simple for spammers to generate buzz about a product by posting spam reviews. Email and web spam are typicallyreferred to as review spam. By changing the content of a website so that search engines rank the site highly, spam on the internet is used to entice users. The major reason email spam is sent out is to advertise. Spam reviews, on the other hand, vary in that they express a fake opinion about a good or service, and they are highly challenging to identify mechanically. Thus, spam control detection cannot

be accomplished using current web spam or email spam detection methods. In this study, Spam Checkerresearch is based on

30.5 million Amazon.com reviews and 20 million reviewers. However, the main limit of this domain is that the available data is not labeled, just like the Amazon data set. The suggested solution first develops a spam-checking detection algorithm employing behavioral approaches (SRD- BM) to provide a named dataset in order to address this issue. After that, classifiers using Spam Inspection Detection Using a Linguistic Approach are trained using this labeled dataset (SRD-LM). To create a spam survey, the suggested methodologies specifically included linguistic elements like shinglestechniques and a variety of spammers' be behavioral traits including activity window, number of analysis, positive evaluation ratio, negative evaluation ratio, initial analysis ratio, and review length. Model for detection. In earlier studies, these language were not fully utilised.

This work produced the next scientific input:
1. The proposed methods used in genuine data file
2. The suggested behaviour technique, which utilised 12 distinct behavioural cues to recognise spammersand unsolicited mail
3. The suggested SRD-BM SRD-LM that employed classifiers and linguistic aspects to recognise spamreviews
4. A comparison and analysis of the planned SRD-BM and SRD-accuracy LM's

2 LITERATURE REVIEW

The literature survey or review only included publications from the last several years that use machine learning to predict spam reviews. In the recent decade, the penetration of bio-inspired algorithms gives tremendous advantages to many industries and applications. Bio-inspired algorithms are a special type of computing procedure in which, the behavior of natural living specimens is imitated. These bio-inspiredalgorithms are mostly used to solve complex problems in computing. The most commonly used bio- inspired algorithms are evolutionary programming, genetic algorithms, and genetic programming and swarm intelligence algorithms (Kazimipour, B, Li, X. & Qin, A.K, et al. 2014). These approaches can be used for performing a heuristic search in feature selection and classification problems. Thus the naturallyinspired algorithms are used for the effective selection of feature subsets in spam review classification. In recent periods many researchers, involved in the research to develop a suitable approach for Spam review classification. Thus in the literature, many prediction algorithms are proposed to identify spam reviews.

Dong-Her Shih et al. (2008) have put-forward a collaborative study framework for spam filtering. The collaborative learning framework is a comprehensible and simple procedure for spam filtering which achieves more than 90 % accuracy. amount of review or feedback content provided byconsumers is increasingly on the internet (Li, J, Ott, M, Cardie, C, & Hovy, E. H, et al. 2014). This information is veryimportant because many new consumers consider it evidence to take their decision (Mukherjee, A, Venkataraman, V, Liu, B, & Glance, N, et al. 2013). But there is no guarantee for the quality of these reviews; some sellers may give fake reviews to promote their product or demote the competitors' products (Banerjee, S, & Chua, A. Y, et al. 2011). These counterfeit reviewers are denoted as opinion spammers, and their content is termed spam review. Spam reviews are of two types; they are positive spam reviews and junk reviews. The spammer produces positive reviews about any particular product toimprove the popularity of the product and to gain more profit. (Ahmad, I,& e Amin, F, et al. 2015). On the other hand, they may produce some negative feedback about a particular product without the proper knowledge to defame the organization or product These spam reviews lead a new consumer or user to take the wrong decision, hence it is essential to spot the truthfulness of every review and reviewer. An experct approach for spam review detection

G. Huayi Li et al. 2015 presented a study to detect fake reviews in Chinese. (Kuldeep Sharma& King-IpLin et al. 2013) have suggested a method for assessing product reviews and identifying reviewspam.Their approach was based on various domain which included checking junk reviews, review rating, links, surveys,similar analysis. Using the information-theoretic measure KL-divergence and its asymmetric feature, Arjun Mukherjee et al. (2013) have suggested an unique way to identify the precisedistinction between the two types of review data. For learning, they suggested adding a new set of behavioural characteristics about reviewers and the reviews they write.

Donato Hernandez Fusilier et al. (2014) have proposed a system for the detection of opinion spam. Theyutilized the 1600 hotel reviews and used the NB classifier for opinion spam detection. Zhen Hai et al. (2016) have proposed a novel supervised multitask learning method via Laplacian regularized logistic regression to further improve the performance of review spam detection.(Michael Crawford et al. 2016)have introduced filter-based feature rankers and a word-frequency model for the reductionof feature dimension in review spam detection

3 ANALYZE DATABASE

Creating a dataset of labeled reviews is a difficult task when developing a supervised learning model forspam review identification. Instead of gathering filtered spam reviews from various company ratings, the majority of the current expert system spam control systems rely on false reviews. Amazon Mechanical Turk or manual annotation are both used to manufacture phony reviews (AMT). AMT is a free marketplace for crowdsourcing where anyone can simply get paid to post evaluations for their requirements. Manually identifying spam analyses is an extremely difficult Turkeys have the same mindset as real spam reviewers, making spam reviews difficult to spot. Because of this, spam rating revelationmodels build using duplicate comments have higher training accuracy than models based on real data. reviews that are spam. Such models, nevertheless, are ineffective in spotting real spam reviews becausethey were trained on bogus reviews. In light of the aforementioned concerns, our research makes use ofactual Amazon product review data, which takes into account the varied behavior and submission history of reviewers. Table 1 displays a thorough value of all the .NonSpam Spam categories, reviews, reviewers, and goods breakdown of the dataset. The product review data utilized these work was not tagged, whichis a need for training a classifier for spam review identification using a linguistic approach. In order to resolve this issue, researchers first creatlabeledlled dataset using SRD-BM (Division IV), and then traina classifier using SRD-LM using that labelled dataset. The NLTK6 3.0 natural language toolkit, which comes with simple-to-use built-in word processing modules, is used in linguistic method for post processing, recognition, overview evaluation , feature extraction, selection, and classification.

Type	**Total Analysis**	**Total Analyst**	**Total Result**
Jewellery	344665	22604523	3198658
Headphones	5748294	3116944	1138462
Watches	7820765	4200425	4758645
Iron Box	4252542	2545111	4107652
Sports and Outdoor	3267594	1984514	4787859
Toys	2257586	13424583	3275863
Total	26787758	15421610	3148785

TABLE 1.Distribution information for the Amazon dataset used in the suggested methodology

4 PROPOSED METHOD

A) USE OF A BEHAVIOR METHOD FOR SPAM DETECTION

In this part, the suggested strategy for identifying spam reviews is discussed, along with an analysis of howwell it performs overall. The unlabeled dataset may be utilized with unsupervised learning to detect spam decisions since the spammer can be recognized by examining their numerous behavioral traits. Using the Behavioral Method a labelled output is generated from an unlabeled dataset. a dataset that doesn't use spam filtering to find spam. Figure 1 depicts the suggested SSRD framework BM's The first step in the procedureis to locate and compute the behavioral traits of spammers in the dataset of anonymous Amazon reviews. Thenormalized values for each behavioral attribute are used to generate the average score of the relevant reviewsof the whole data set. The spam checker's detection accuracy is then determined using this average score andthe average approach. The procedure then proceeds by removing each behavioral attribute one at a time andrecalculating an updated average score known as a drop score in order to determine the significance of eachbehavioral trait. The accuracy of the drop points is compared to the accuracy attained with the average scoreThis decreased behavior is assigned a weight of "3" if the attained accuracy drops by 10%, else a weight of "2". Also, based on their significance within the dataset, each behavioral trait is given a weight. Next, each review's spam score is determined using the importance each behavior is given. To analyze whether a reviewis a junk or not, the spam score is then compared to a flexible threshold.

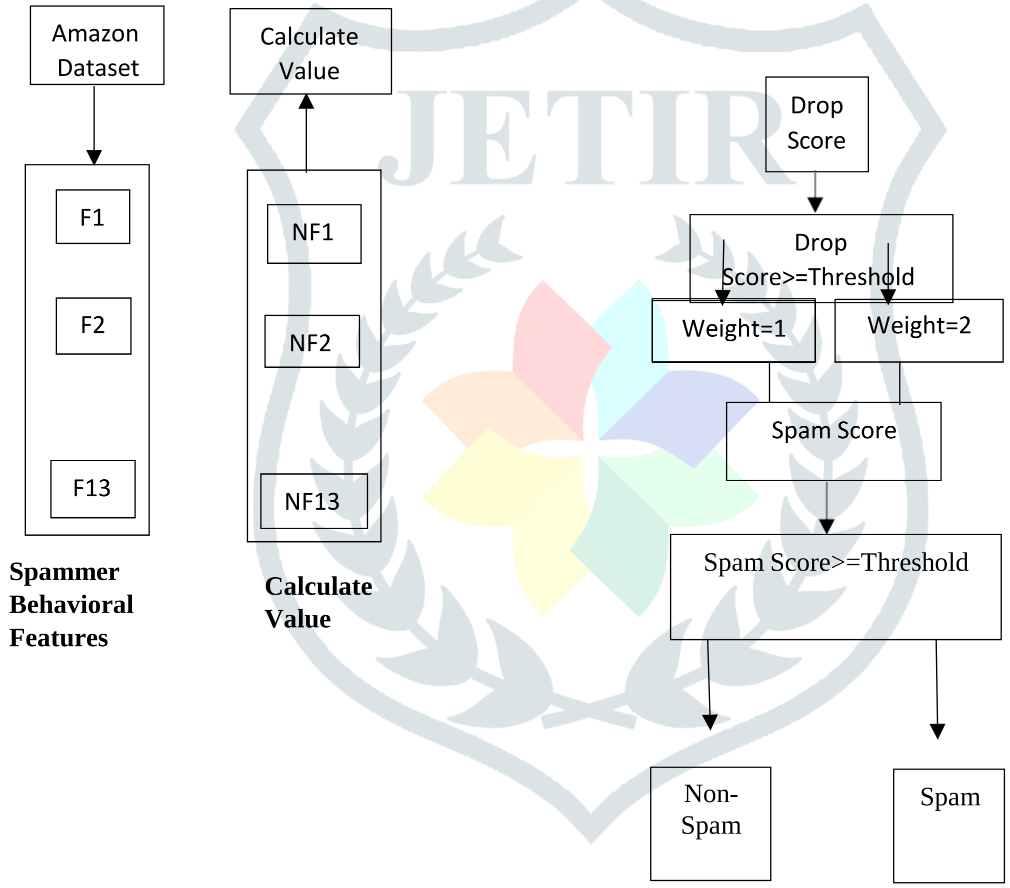

FIGURE1.The framework of the spammer behavioral method

RESULT AND DISCUSSION

The execution of the proposed SRD-BM is evaluated using the assessment metrics of accuracy, recovery, f-measure, and precision. Using a logistic regression classifier, the labeled dataset obtained from the recommended behavior technique is trained and evaluated. This classifier is the approach that performs the best when using the advanced linguistic strategy (Division V). Cross-validation k-fold (k = 5)is then used toassess and test the indicated SRD accuracy. In every iteration, the proposed model a trained using the remaining k-1 segments and tested on one segment. This method is performed k times, exactly utilizing each section once, to train the recommended model. Exact computation The first step in determining accuracy is averaging each review and taking advantage of all the spammers' propensities for misbehavior. The impact of certain spammers' behavioral features is then assessed by altering the drop function approach and examining the audit dataset. The overall accuracy is determined using the spam scoring approach in order toevaluate the effectiveness of the recommended behavior method in scoring spam as opposed to spam.

B) USE OF A LINGUISTIC METHOD FOR SPAM REVIEW DETECTION (SRD-LM)

The linguistic method has been used for the Spam review detection For proposed model, different algorithms were used in this study. However, in the search for greater accuracy, several problems arose withthe algorithm. When the proposed model searched for the best accuracy this time, the system found that although some algorithms have good accuracy. They also have some disadvantages and require time. And the approach understands that a flawless model is needed to detect spam audits. Therefore, the first step of the proposed model is data collection for the linguistic methodology.

The literature discovered that draught text is frequently usedfor spam control detection when using the linguistic way of detection. A scan is typically classed as either "spam" or"not spam" using binary classification. The planned SRD- LM is covered in greater detail in this section. It explains how characteristics are extracted and chosen from the review content. The suggested technique is tested and trained using a variety of classification methods, which are also detailed.

The suggested linguistic approach creates a precise spam control model using a variety of data pretreatment methods, transformations, variable selection, and machine learningalgorithms. The six phases that make up the whole process for the linguistic method are shown in Figure 2.

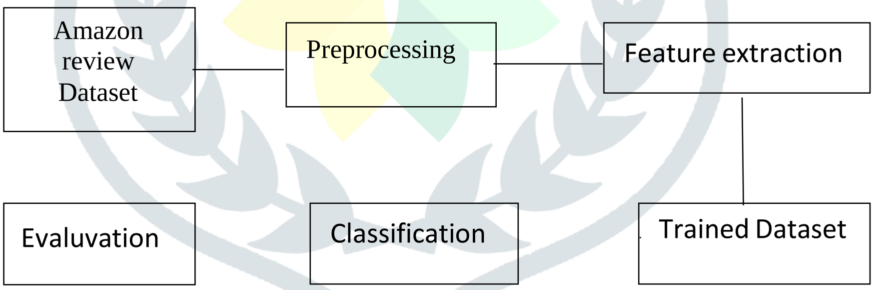

FIGURE 2.Process of SRD-LM

- **Data collection**

This is the foremost step towards the growth of a machine learning model, data collection. It is a important step that affects the quality of the model. The more and better datawe get, the better our model works. There are various techniques to collect data like web scraping, manual intervention, etc. A Comparison of Machine Learning Algorithms for Spam Prediction from Kaggle and Other Sources

- **Dataset**

The data consists of 821 individual data. The dataset has 27 columns, which are described below. The Yelp rating will be a baseline, where suggested reviews related to "genuine" and non-suggested "fake". The power of these data sets are. A large number of user-based reviews allows for taking into account the behavioral characteristics of each user. Multiple units of review ie bus and mobile shop Databases contain only basic information such as the tag, ranking, and occasion of each review associated with the end user who created it.

Method	Precisions	Recall	Accuracy	Auroc
Unigram	0.866	0.847	84.036	0.853
Bigram	0.871	0.853	85.564	0.888
Trigram	0.856	0.838	83.507	0.858
Unigram+ Bigram	0.830	0.810	81.067	0.841
Bigram+ Trigram	0.740	0.729	72.910	0.749
Unigram+ Bigram+ trigram	0.663	0.652	65.554	0.676

TABLE 2. Evaluvation of SRD-LM with svm

- **Data Preparation**

We'll change the data. by eliminating any missing data and certain columns. The column names that we wish to preserve or retain will first be listed. The next step is to eliminate or delete all columns save for the ones we wish to keep. Lastly, we discard or eliminate the rows from the data set that contain missing values.

- **Analyze and Prediction**

They are based on temporal data that further explains how ratings are dispersed over time, including: the user's activity time; the difference between the timestamps of the most recent and earliest reviews for a certain reviewer. Maximum score each day. The number of days between two consecutive pairsof reviews is the data entropy. The supervised machine learning strategies for classifying data, balancing data, and evaluating the classifier are implemented using the following methods.

- **Classification Methods:**

The following four supervised learning techniques are used as input for classification once the textreviews are transformed into document-term matrices.

- Using The Classifier Of Naive Bayes The Nave Bayes (NBC) is utilized for training and classification tasks. Based on Bayes' theorem, this is a probabilistic classification technique. The naive supposition that the characteristics in a dataset are independent of one another is the foundation of naive Bayes classifiers.

- Classifier for Logistic Regression (LR): A standard technique for looking at a datais called logistic regression (LR). To identify the outcome, this classifier makes use of oneor more independent factors. With a binary variable, the outcomes are quantified as either 0 or 1. According to LR, P(y|x), the posterior distribution, is assumed to be a expit.Here, xis the collection of characteristics and y is the label.

- Classifier for Support Vector Machine: Modern classifiers increase predicted validity while abruptly avoiding ounderfit of the data thanks to their optimization process.The SVM projects the input data into this kernel space and creates a linear model there.TheSupport vector network is the state of technique for data mining and machine learning algorithms.

- Random Forest (RF) classifier: A Random Forest (RF) is a meta-estimator that fitsseveral decision tree classifiers on various subsamples of the dataset while using averagingto improve prediction accuracy and decrease overfitting. It seems to be quite well-liked these days because of its numerous advantages, such as being speedier and more extendable than other machine learning models.

5 RESULT AND COMPARISON

In order to pick out spam reviews in the dataset, the proposed behavior method and linguistic method are compared in this section. The comparison's outcomes ofdetected junk reviews are shown in Figure 3.In the dataset, the behavior method distinguished large no of spam reviews with higher accuracy than linguistics. The Amazon dataset had 30 million reviews in all. The behavior method effectively classified8,500,4762 reviews, or 32% of thetotal reviews, as spam, according to observations. Reviews from the remaining 68% are classified as non-spam. SRD- LM, using the identical Amazon dataset, on the other hand, detected 5,458,884 reviews as spam, or 31% of the entire review dataset. This demonstrates that inlarge-scale real- world Amazon datasets, the behavior method is more reliable than the linguistic methodin identifying bogus reviews. The experimental assessment of this has also shown that SRD-BM obtained 33.1% accuracy whilelinguistic achieved 53.8% accuracy in the detection of spam reviews.

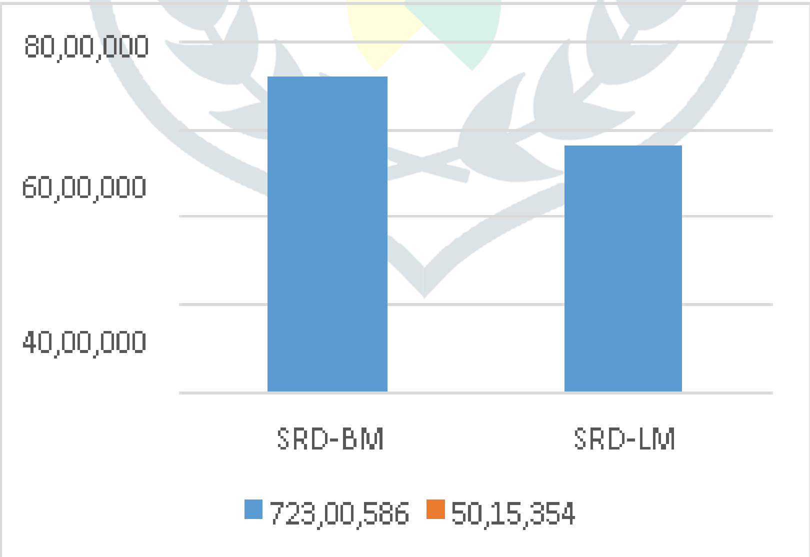

Figure 3: Comparision of behavior and linguistic method

6 CONCLUSION

Spam in online reviews is an increasing issue. Due to the difficulty in separating spam reviews from legitimate evaluations, spam review detection (SRD) is a crucial yet challenging undertaking. Several academics have attempted to identify spammers and spam reviews up until this point, but their efforts have fallen short of providing a comprehensive solution. The article developed the behavior and linguistic algorithms for spam audit identification using behavioral and linguistic approaches after conducting an extensive investigation on the Amazon real-world dataset utilizing the behavioral characteristics of spammers.

As far as the researcher is aware, this is the first study to examine and apply spammer's typical behaviors to a sizable existing audit dataset. Also, the experimental assessment demonstrated that behavioral characteristics like content comparable to a large number of comments, positive review ratio, isolated productreview, activity window, and review span features considerably increased the accuracy of the suggested detection. On the other hand, the suggested linguistic approach linguistic method analyzed the dataset for theidentification of spam web using feature selection, modification, and several classifying algorithms. Each classifier's performance was assessed, and it was discovered that Logistic Regression outperformed SupportVector Machine, Random Forest, and Naive Bayes. Because behavior method the behavioral aspects of the dataset, including timestamps and ratings, which give extra support for spam identification and therefore spam ratings, it outperformed linguistic method in terms of accuracy.

The findings of this study have applications for enhancing the dependability of online platforms for evaluating goods and services. Applications of the study include spotting fake reviews on e-commerce and product/service websites like Amazon, Flipkart, purple, and meesho among others. The accessibility of standard labeled datasets for training classifiers will be the focus of the next research. The dataset is also expanded with new features to boost the precision and dependability of algorithms for spam inspection identification. This might include the unique address of the spammer and how many email IDs have been registered for the particular IP address, the reviewer's email address, and the login location. The identification of spam reviews using a multi-linguistic review dataset and the identification of the spammer usingsuggestive analysis of reviews published by other users are possible further paths for the future. Using deep learning classifiers to implement this challenge is a key future topic for this research.

7 SAMPLE OUTPUT

REFERENCE

[1] J. Huang, T. Qian, G. He, M. Zhong ja Q. Peng, "Detecting Professional Spam Checkers," Proc. Int. Conf. Adv. Applied for mined date. Berlin, Germany: Springer, 2013, p. 288–299.

[2] S. Bajaj, N. Garg and S. K. Singh, “A novel user-basedspam detection,” Procedia Comput. Sci., vol. 122, pp. 1009- 1015, January 2017.

[3] A. Mukherjee, V. Venkataraman, B. Liu and N. S. Pilk: "What Could Yelp's Fake Review Filter Do?" Proc. International Conf. Online Social Media, July. 2013, pp 409–418.

[4] Y. Ren kaj D. Ji, "Learning to detect deceptive opinionspam: Enketo", IEEE Access, vol. 7, s. 42934– 42945

[5] T. Ong, M. Mannino. Gregg, "Linguistic features of Shillreviews", Electron. Business Res. App., vol. 13, no. 2,p. 69-78, March 2014.

[6] I. Dematis, E. Karapistoli and A.S. Vakali, "Fake ReviewDetection Using Spam Indicators and Reviewer Behavioral Characteristics Methodology", Proc. International Conf. Current trend theory practice. to report Cham, Switzerland: Edizioni Della Normale, 2018, pp. 581–595. 53814 Volume 8, 2020 N. Hussain et al.: SRD Using Linguistic and Spammer Behavioral Methods

[7] H. Li, Z. Chen, A. Mukherjee, B. Liu ja J.'', Proc. Int. Konf.Web Social Media, 2015, pp. 634-637.

[8] B. Liu and L. Zhang, "A survey of opinion and sentimentanalyses", Mining Text Data. Boston, MA, Usono: Springer,2012, s. 415–463.

[9] S. Zhou, Z. Qiao, Q. Du, G. A. Wang, W. Fan, kaj X. Yan, "Measuring customer responsiveness from online reviews using big data text analysis," J . Control. Info Syst., Vol. 35, no. 2, lk. 510–539, April 2018.

[10] L. Chen, L. Chun, L . Ziyu and Z. Quan, “Hybrid pseudorelevance feedback for microblog retrieval,” J. Inf. Sci., vol. 39, no. 6, pp 773–788, December 2013.

[11] C. Lin, Z . Huang, F. Yang and Q. Zou, "Identifying content quality in an online social network", IET Commun., voi. 6, hey. 12, s. 1618-1624, 2012.

[12] M. Chakraborty, S. Pal, R. Pramanik and C. R. Chowdary, "Recent Developments in Social Spam Detectionand Prevention Techniques: A Survey," Inf. The process. Check it out, oh. 52, no. 6, pp. 1053-1073, November 2016.

[13] A. Mukherjee, A.S. Kumar, B. Liu, J. Wang, M. Hsu, M.Castellanos kaj R. Ghosh, "Spotting Opinion on Spammers Using Behavioral Traces", Proc. 19-a ACM SIGKDD Int. Konf. Sciu. Discovery Data Mining (KDD), 2013, p. 632–640.

[14] A. Heydari, M. Tavakoli, kaj N. Salim, "False opinion detection using time series methodology", Expert Syst. Appl.,vol. 58, pp 83–92, October 2016.

[15] S.K. Kc and A. Mukherjee, "The Temporal Dynamics ofOpinion Spamming Method: Case Studies from Yelp", Proc.25th Int. Conf. World Wide Web (WWW), 2016, p. 369-379-a

[16] H. Li, G. Fei, S. Wang, B. Liu, W. Shao, A. Mukherjee,kaj J. Shao, "Bimodal propagation and co-burst in online spam detection," Proc. 26-a Int. Conf. World Wide Web (WWW),2017, pp. 1063–1072.

[17] J. G. Biradar, S. P. Algur kaj N.H. Ayachit "Exponential Distribution Model for Review SpamDetection", Int. J. Adv. Computer Res. Science, you. 8, hello.3, s. 938–947, 2017.

[18]N. Jindal and B. Liu, ‘‘Analyzing and detecting review spam,’’ in Proc. 7th IEEE Int. Conf. Data Mining (ICDM), Oct. 2007, pp. 547–552.

[19]R. Y. Lau, S. Y. Liao, R.C.-W. Kwok, K. Xu, Y. Xia kajY. Li, "Text mining and probabilistic language modeling for web spam scanning detection," ACM Trans. Control. Info Syst., Vol. 2, no. 4, s. 1–30, 2011.

[20]F. H. Li, M. Huang, Y. Yang, and X. Zhu, “Learning to detect spam,” Proc. International Joint negotiation Artif. Intel. (IJCAI), 2011, kd. 22, nr. 3, lk. 2488.

[21]D. H. Fusilier, M. Montes-y-Gómez, P. Rosso, and R. G. Cabrera, "Opinion spam detection usingn-grams," Proc. International Conf. Intel. Text process. Calculate. Linguistics. Cham, Switzerland: Springer, April 2015, s. 285–294.

[22]M. Ott, Y. Choi, C. Cardie kaj J. T. Hancock, "Finding fraudulent reviews spam using imagination",Proc. 49. Annu. Kohtumine Assoc. Calculate. Linguistics, Hum. language Tech. , 2011, pp. 309-319.

[23]M. Hazim, N. B. Anuar, M. F. A. Razak and N. A. Abdullah, “Opinion spam detection using a supervised learning method,” PLoS ONE , vol. 13, no. 6, 2018, art. ei. e0198884.

[24]N. Kumar, D. Venugopal, L. Qiu, kaj S. Kumar, "Manipulation Detection on Online Platforms Usinga Hierarchical Supervised Learning Method", J. Manage. Informoj Syst., Vol. 35, no. 1, pp. 350–380,January 2018.

[25]D. Zhang, L. Zhou, J. L. Kehoe ja I. Y. Kilic, "Which Online Reviewer Behavior Matters? The Effects of Verbal and Nonverbal Behavior on the Detection of Fake Online Reviews, J. Manage. Inf. Syst., Vol.33, no. 2, lk 456–481, April 2016.

[26]S. Ahmed and A. Danti, "Effective sentiment analysis and opinion mining of online reviews using rule-based classifiers," in Computational Intelligence in Data Mining, vol. 1. New Delhi, India: Springer, 2016, p. 171-179.

[27]Y. Lin, T. Zhu, H. Wu, J. Zhang, X. Wang, kaj A. Zhou, "Towards online review spam detection: Identifying fake reviews from the review cycle," Proc. IEEE/ACM Int. Konf. Adv. Social Network Anal.Mining (ASONAM), August 2014, p. 261-264.

[28]J. Li, M. Ott, C. Cardie and E. Hovy, "Towards a General Rule for Detecting Misleading Opinion Spam", Proc. 52. Ann. Meeting Assoc. Calculate. Linguistics, vol. 1, 2014, pp 1566– 1576.

[29]H. Ge, J. Caverlee and K.S. Lee, "Crowds, gis and super sellers: a measurement study to detect spam in the supply-based crowdsourcing market", Proc. ICWSM, 2015, pp. 120- 129.