Spam and Fake Spam Message Detection Framework Using ML Algorithm

S. Kalaiarasi, Saurabh Rai, N.Hemanth Venkata Sai, Anshuman Chaurasia, B.Sreeshanth Reddy

Abstract: The period of time now occurring social media has generated incomparable amounts of social data. It is a popular communication and also known as knowledge exchange media. Data could be of any type of text, numbers, figures or statistics that are accessed by a computer. Nowadays, many people depends on contents available in social media in their decisions. Sharing of information with peoples has also attracted social spammers to exploit and spread spam messages to promote personal web logs, advertisements, promotions, phishing, scam, frauds and so on. The prospect that anybody will leave a review give a golden chance for spammers to post spit review regarding product and services for various interests and chances. So, we propose a fake message detection framework using ML ,where we use RB, RL, UB, UL mechanisms in order to detect the spam and fake messages on the social media platform.

Keywords: spam detection, machine learning, fake reviews, Baye's Theorem, NLP

I. INTRODUCTION

The predominance of web has extended the use of electronic business arrangement and administrations. The preponderance of the E-business site give an official inspection to the client with the goal that they can post survey of item at dealer site and convey their perspectives. Such solidity contributed by web is named as client created content. This substance frames profitable statistics for traders organizations about their item and different clients, item makers. It tends to be incredibly significant to make business centred and make benefit. In spite of the fact that these surveys are vital origin of data there is no status control on this client produced information, anybody can devise anything on internet which prompts many low quality inspections still more regrettable survey spam which delude clients influencing their procure choices.. Spam audits are Very essential these day in web business locales.

Revised Manuscript Received on April 14, 2019.

S. Kalaiarasi, Assistant Professor(O.G.), Department of Computer Science and Engineering, Ramapuram campus, SRM Institute of Science and Technology

Saurabh Rai, Undergraduate Student, Department of Computer Science and Engineering, Ramapuram campus, SRM Institute of Science and Technology

N.Hemanth Venkata Sai, Undergraduate Student, Department of Computer Science and Engineering, Ramapuram campus, SRM Institute of Science and Technology

Anshuman Chaurasia, Undergraduate Student, Department of Computer Science and Engineering, Ramapuram campus, SRM Institute of Science and Technology

B.Sreeshanth Reddy, Undergraduate Student, Department of Computer Science and Engineering, Ramapuram campus, SRM Institute of Science and Technology

As not every single online survey are veritable and dependable, it is important to construct procedures for distinguishing audit spam. It is conceivable to direct audit spam discovery utilizing different AI strategies by extricating significant highlights from the content utilizing Natural Language Processing (NLP). In Addition to it commentator data separated from the content can be utilized to help this procedure. Greater part of ebb and flow investigate centre around the directed learning strategies, which requires named information, a shortage with regards to online survey spam. Exploration on techniques for Big Data are of enthusiasm, there are a great many web audits, with a lot all the more being created every day.. The essential objective of this paper is to give a solid and far reaching near investigation of momentum look into on recognizing survey spam utilizing different AI systems and to create technique for directing further examination.

II. EXISTING SYSTEM

Spam review content and user behaviour analysis has been used by many researchers to detect fake reviews and spam messages. Earlier methods work on mainly 2 detection methods: Linguistics-based method, Behaviour-based method.

Linguistic-based method mainly focuses on detecting fake reviews based on views/opinions expressed earlier. It uses Bigram and unigram methods to identify fake reviews. In bigram method it takes a sequence of two adjacent elements form a word or string and it is often used for analyzing of data and texts in many applications, including in computational linguistics and text identification. In unigram method the probability of each word only depends on that word's own probability in the document.

Behaviour-based method mainly apply mathematical and statistical proficiency on historical data to predict the future behavior of customers. It uses metadata for processing. Encapsulates basic information regarding the data, which can make effort and working with particular instances of data easier. Metadata will be created manually, or by automated information processing. Manual creation tends to be a lot of correct. Automated information creation tends to be a lot of correct. Automated information creation are going to be much more elementary, typically displaying data like file size, file extension, file creation time and by whom it has been created.

III. PROPOSED SYSTEM

The proposed system is based on 4 methodology. These methodology involves techniques to simulate fake reviews by considering the user frequency in posting the reviews for the same product. These methods are Review-behavioral method, User-Behavioral method, Review-Linguistic method, User-Linguistic method.

A. Review-behavioural method

This element depends on metadata which is organized data that portrays and discloses information so as to make it less demanding to find, recover, use and deal with a data asset. It gives information its unique circumstance and implying that is expected to infer bits of knowledge. There are various opportunities with machine learning and the capacity for utilizing calculations to catch basic leadership decides so it very well may be connected to a bigger informational collection. It is based on two features:

1. Early Time Fame (ETF):- ETF is the time at which it decides the output of a query and give result. This can manipulate result at any point of time before completion of process.
2. Threshold rating deviation:- Maximum condition at Which Algorithm can give maximum result, so that we can achieve the duplicate content on website.

B. User-behavioural method

User behavior analysis is a method of gathering insight into the network events that users generate daily basis. Humans doesn't follow a well-defined logic, we have tendency to create a recurrent patterns. We often behave in similar way and follow similar intuitions. So if we can learn the reviewer's pattern, we may be able to identify the spam reviewer. We use this sort of methods to know how reviewer communicates. It is based on each individual user language. It depends on two features:

1. Shattering of reviews given by a single user on a particular product
2. Average of a users false review given to different products and businesses

C. Review-linguistic method

It is based on review text. Machine learning, Natural Language Processing (NLP), Data Mining are techniques which work together to automatically classify and discover patterns from the reviews. It directly extracts from text of the review. Text extraction analyses whether the sentiment towards any review is positive, negative, or neutral. It is based on two features:

1. Ratio of 1st Personal Pronouns (PPl)
2. Ratio of exclamation! sentences containing (RES)

D. User-linguistic method

It is based on reviews collected by each individuals. The generalization of each user is done by Machine Learning. It is based on two features:

1. Average Content Similarity (ACS)
2. Maximum Content Similarity (MCS)

IV. MATHEMATICAL EQUATIONS

This section provides description about methods and the classification algorithm used in this paper, and the mathematical measures employed within this work.

A. Probability that a review is a spam

Certain words have particular prospects of occurring in spam messages. The filter does not be able to recognize these outcomes in advance, and they must be trained first to do so. To instruct the filter, the user must manually check whether a given message is spam or a true message. For every message which is used to train the machine learning algorithm, will improve the probability of each and every word which is recorded as a spam and saved in the database. Let's suppose the spam message contains a word that has been used maximum. Most people who frequently used to reviewing the products knows that whether the given review is spam review or not. The spam detecting software, however, does not know such facts all it can do is compute probabilities and we have to train that software to detect the fake reviews.

The formula used to determine the probability is:

$$Pr(S|W)=\frac{Pr(W|S).Pr(S)}{Pr(W|S).Pr(S)+Pr(W|H).Pr(H)}$$

B. Combining individual possibilities

Spam detecting algorithms are based on formulas that are applicable only if the texts or words present in the message are free from outside control or an independent events. So we can represent it by the following Baye's theorem formula:

$$p=\frac{p_1 p_2 \cdots p_N}{p_1 p_2 \cdots p_N+(1-p_1)(1-p_2)\cdots(1-p_N)}$$

Where p is the prospect that the suspect message is a spam message, p1 is the probability that the spam message contains the first word, p2 is the possibility that the spam message holds the second word, pN is the probability that the spam message contains the Nth word.

C. Probabilities of un-encountered words

In this case we encounter a word that has never been used during the training period, If there is no information available about the particular word then software tends to discard such kind words. All the more for the most part, the words that were experienced just a couple of times amid the learning stage cause an issue, since it would be a mistake to credence aimlessly the data they had given. Then we have to simply avoid taking such texts or words under consideration.

V. SYSTEM DESCRIPTION

A. Architecture

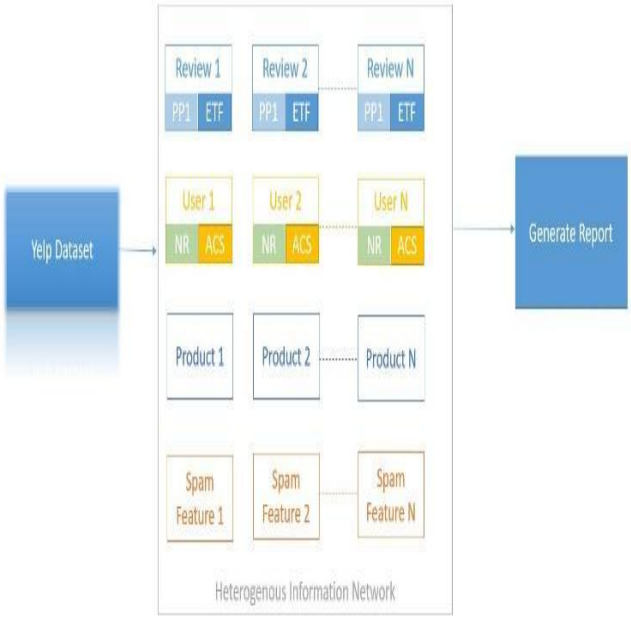

Fig. 1 System Architecture

The Short lived definition of spam is unwanted e-mail or message. The critical point is that an un-labelled data can be labelled as good data that can be used against many other reviews which will give wrong result. Now this wrong result will authenticate many fake and spam reviews which is not a good for review detection. This misleading of review will continue many other reviews and this will keep on increasing. There are many technique which can identify which one is good review are which one is fake review. There is a way to identify the fake and spam reviews by following technique: labeled/unlabeled list, Bayesian classifying algorithms, keyword matching which will search the term can have before or after the phrase and header info survey. Labelled data are those data's on which a user or algo. Can depend on. User can add labeled data anytime and remove also, can also add keywords and also information about the new words. It gives important feature to user in which he/ she can grant/revoke access to data. An Unlabeled data are those data which are not labeled, they are not reliable and cannot give a good result. The review reviewing methods involves a string of rules can be executed as follows. A review will be classified as junk and labeled as a spam if its keywords match with the spam keyword and some reviews are used in training review otherwise it is called as labeled data. Just content calculations that have been displayed with better productivity are as of late utilized for reviewing. Different package build software are used previously on basis of methods. Rule based solution have very low fake reviews. First we extract data from the large data set, extracting only those data which are necessary. After that based upon the behavior of the user the string calculation rule is used, how the user is behaving is it a fake review or a good review by the user. Now the next method is filtering on the basis of user behavior this user give more fake reviews or not. At last we generate the report which contain all the useful information about the reviews.

B. Data-set Extraction

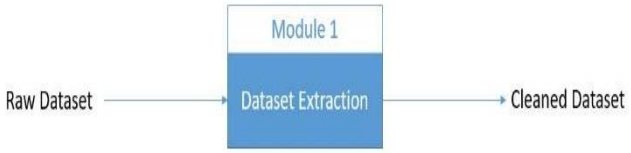

Fig. 2 Data-set Extraction

In this module we extract the data from the yelp server then we make a dataset of all those information for the next process.

C. Review-Behavioural

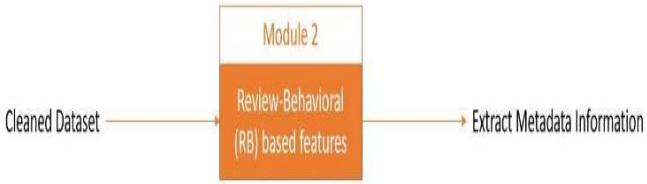

Fig. 3 Review-Behavioural

Now we prepare the true dataset on which algorithm can perform its computation. It is necessary to take only reliable data for true result. We extract the useful information from the reviews and send it to next process.

D. User-Behavioural

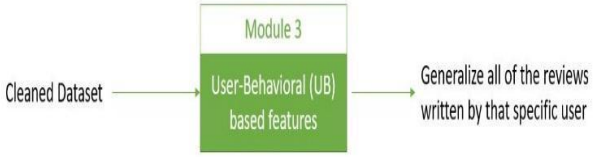

Fig. 4 User-behavioural

From reliable set off information we match the users review. How much user's accuracy is. The content of user's review is reliable or not, we match all the other review for good result.

E. Generate Reports

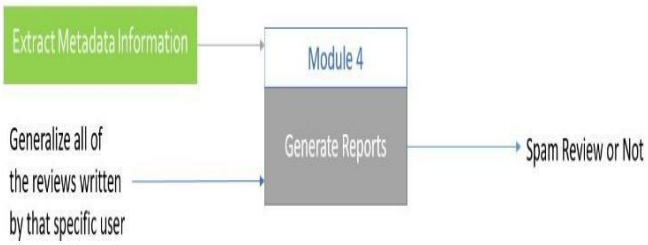

Fig. 5 Generate Reports

In this module we use the result of module 1 and module 2. We take all the reliable data from dataset and generalized data of user's and generate a sets of results from that set of results we take only maximum accuracy result and show it.

FUTURE WORKS

The results attained by our model with some enlightened detection approach which shows that our proposed model produced better results than most of the prior works. Another point is to be noted here is, this approach not only outperform some of the supporting works but also produces a labeled dataset from the unlabeled data in the process which might be used as labeled dataset in future exertions. Future empirical studies may include large-scale datasets from different domains in order to increase the size and diversity of the data to evaluate the heftiness of different classifiers. Diverse sets of tuning and smoothness techniques can be introduced also. The feature set might be improved by using n-gram models (unigrams, bigrams and trigrams) with additional pre-processing techniques. Future research can introduce ensemble methods which involves multiple classifiers to detect review spams more specifically and effectively. This research will continue, for achieving even better performance on reviews of different language and domains using the above mentioned improvement guidelines. The future extend our auditing protocol to support the data dynamic operations, which is efficient and provably secure in the random model.

VI. CONCLUSION

In this paper, we proposed a well organized and brief description of spam detecting methods which has been used from past. As fake reviews ruins the e-commerce business so in this paper, we proposed an efficient and more reliable system to detect the fake reviews given by the spammers. Here Naive Bayes approach used to detect authentic and spam reviews. Furthermore, our computation sustains less cost and less computation cost of the review, which greatly improves the performance and can be applied to large scale reviews. As linguistic methods gives effective results. More Categories of features are added on the basis of behavioral and reviews statics.

REFERENCES

1. Investigation on social media spam detection. http://ieeexplore.ieee.org/document/8275931
2. Spam E-Mail Classification by Utilizing N-Gram Features of Hyperlink Texts https://www.researchgate.net/publication/318861757_Spam_E-Mail_Classification_by_Utilizing_N-Gram_Features_of_Hyperlink_Texts
3. Detecting Algorithmically Generated Malicious Domain Names http://eprints.networks.imdea.org/67/1/Detecting_Algorithmically_Generated_Malicious_Domain_Names_-_2010_EN.pdf
4. Spam E-Mail Classification by Utilizing N-Gram Features of Hyperlink Texts https://pdfs.semanticscholar.org/c5d5/9b58c2d46e692429446eec10407911b0a416.pdf
5. Time-efficient spam e-mail filtering using n-gram models. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.136.2510&rep=rep1&type=pdf
6. Words Vs. Character N-Ggams For Anti-Spam Filtering. http://www.icsd.aegean.gr/Stamatatos/papers/IJAIT-spam
7. Part of Speech Tagging (POS). http://en.wikipedia.org/wiki/Part-of-speech_tagging