Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 2, June : 2023

FAKE PRODUCT REVIEW DETECTION USING MACHINE LEARNING

K. Sagar, B.Tech Student, Electronic and Communication Engineering, Sreenidhi institute of science and technology, Hyderabad, Telangana, India.
N.Kalyn, B.Tech Student, Electronic and Communication Engineering, Sreenidhi institute of science and technology, Hyderabad, Telangana, India.
R.Venkatesh, B.Tech Student, Electronic and Communication Engineering, Sreenidhi institute of science and technology, Hyderabad, Telangana, India.
P.Anvesh, Assistant professor, Electronic and Communication Engineering, Sreenidhi institute of science and technology, Hyderabad, Telangana, India.
Email: sagarkotrangi123@gmail.com, kalyanyadav1604@gmail.com
rathlavathvenkatesh26@gmail.com, anveshp@sreenidhi.edu.in,

Abstract: In light of the increasing proliferation of e-commerce systems, online reviews are mainly seen as a crucial component for creating and maintaining a solid reputation. They also have a huge impact on how end users make decisions. A positive review for a single target item often attracts additional customers and significantly increases sales. Reviews that are fake or misleading are purposely written in order to build an online reputation and attract new clients. As a result, detecting fraudulent reviews is an important and developing research subject. The ability to detect fraudulent reviews is dependent on both the key qualities of the comments and the reviewers' behaviour. This paper suggests employing machine learning to detect fake reviews. To extract distinct reviewer behaviours, this study employs a variety of methods for feature engineering in addition to the feature extraction procedure used in the reviews. The study compares the performance of several tests performed on authentic Yelp data of reviews for restaurants with and without variables obtained from user behaviour. In both instances, we compare the results of various classifiers such as KNN, SVM, Logistic Regression, and Random Forest. Several n-gram models of language, particularly bi-gram and tri-gram, are also considered in the evaluations. In terms of accuracy, the data show that random forest outperforms the other classifier.
Keywords: Data mining,Fake review detection,Feature Engineering,Logistic Regression,Random forest classifier,Supervisied machine learning

I.INTRODUCTION

Without any trustworthy external oversight, user created content is becoming more and more popular on social media platforms, making it impossible to determine which user-generated content is credible or even which source is genuine. Spreading such false information has serious repercussions that hurt both users and businesses. the numerous subsets of traits, or features, that are frequently taken into account by different methodologies related to reviews and reviews as well as the network structure connecting various things on the evaluation in an exam. Analysis of the primary review and review is the primary goal. -Several features, particularly those using supervised machine learning techniques, have been suggested to detect bogus reviews. Opinions detection for spam may identify phoney evaluations, bogus accounts, bogus blogs, bogus social media postings, and bogus communications.Review-focused websites such as Yelp may be used to detect false reviews. Unsupervised approaches that are centred on visual methods but are not particularly reliable have been employed to detect fraudulent reviews up to this point. The supervised techniques consider numerous features derived from the assessments as well as the conduct of the reviewer. Yelp reviews were considered as a freely accessible large-scale and produced dataset. These assessments are categorised using a few renowned supervised classifiers, which classify them as true or false based on various data features. Reviews are declarations that convey a person's idea, opinion, or experience regarding any product on the market. Users post reviews of products on e-commerce websites to share advice or experience with existing customers and product sellers. By examining the suggestions, the user

experience can assist any organisation in growing and improving. The polarity of evaluations affects how much money a product supplier makes or loses. On the other hand, reviews affect potential buyers when they decide whether to buy a specific product. Reviews have varied effects on users and businesses, it might be determined. Considering this, several companies that sell products employ agents to create fictitious reviews in an effort to boost their sales and reputation. As a result, users take improper product selection decision. Online purchasing is becoming more and more popular. Websites for online commerce brought up new channels for buying and selling goods. E-commerce websites make it easy for customers to buy goods or use any service. After utilising a product or using a service, users frequently express their experiences in reviews on e-commerce websites. BCI offers assistance by utilising the brain's ideas as incoming signal for equipment like wheelchairs, robotic arms, and cursor control. Posting phoney reviews is an immoral practise known as opinion spamming. Opinion spamming aims to lead review readers astray. A user who engages in spamming behaviour is referred to as a "spammer". A spammer's job is to create bogus reviews that will help a company's reputation.

II.LITERATURE SURVEY

The detection of false evaluations of web contents can greatly Machine learning algorithms can substantially aid in the detection of fraudulent web content evaluations.To locate and extract valuable information, web mining approaches generally employ a range of machine learning algorithms. Web mining duties include content mining. [1] .

Opinion mining is a common example of content mining in which a classification algorithm is trained to assess the qualities of the reviews as well as the feelings, and which is focused with detecting the sentiment in text (positive or negative) through deep learning. The detection of fake reviews often focuses on criteria unrelated to the text as well as the review category. Natural language processing and text NLP is commonly utilised while developing review feature sets. However, fake reviews may need the development of additional data relating to the reviewer, such as reviewing time/date or writing styles. [2]

The extraction of important features from reviewers is thus critical to the successful identification of bogus reviews. This study employs a variety of machine classification models to detect fake assessments based on the reviewers' own characteristics as well as the subject matter of the reviews.We apply the classifiers to a real-world dataset of Yelp reviews . [3]

The study applies a variety of feature engineering methods to the corpus, in addition to normal natural language processing, to identify and give review characteristics to classifiers. These strategies aid in the extraction of various reviewer behaviours. The essay evaluates the effects of classifiers taking into account reviewers' extracted attributes. The study compares the results of two distinct models of language, TF-IDF with gender flexibility and TF-IDF with tri-grams, with and without the retrieved features. The data suggest that the built-in characteristics boost the method's efficacy in detecting fraudulent reviews. The problem of false comment identification has been addressed since 2007 . [4]

Textual and behavioural components have been extensively exploited in the identification of fake reviews study. Texts refer to the language aspects of a review activity. In other words, features are primarily determined by the content of the reviews. Nonverbal characteristics of the reviews are described in Behavioural Features. They are heavily influenced by the behaviours of the reviewers, such as the way they write, gestures, and how frequently they review writing. While tackling textual features is tough and vital, addressing behavioural aspects is equally important and cannot be overlooked because they have a big impact on how well the bogus comment detection procedure works. A number of investigations on the identification of fraudulent reviews rely heavily on textual elements. The authors of used supervised machine learning algorithms to identify fake reviews in . [5]

Five classifiers are used: SVM, Naive-bayes, KNN, k-star, and a decision tree. Three versions of the tagged film evaluation dataset have been simulation tested, each containing 1400, 2000, and load repeated movie reviews. In addition, the authors of used classifiers such as Naïve Bayesian, Logistic Regression, SVM, random forest, and maximal entropy to detect bogus reviews in their dataset. The

collection includes almost 10,000 negative tweets about Samsung goods and services. The writers of used SVM as well as Idealistic basis classifiers. [6]
The resulting dataset, which contains 1600 reviews from 20 renowned Chicago hotels, was used by the authors. The authors of used neural and category algorithms with summed up, CNN, RNN, and GRNN, mean GRNN, and semi mean GRNN classifiers to detect false opinion spam. [7]
They utilised a data set from with honest and dishonest assessments from three distinct sectors: lodging, dining, and physicians. All of the above research initiatives focused solely on textual aspects, with no attempt to account for behavioural characteristics. In previous works, behavioural traits were taken into consideration in the fake review detection approach. [8]
takes into account some behavioural features of Amazon feedback, such as the mean rating and the proportion of the amount of reviews posted by reviewer. [9]
Another piece of art The authors investigated the effects of linguistic and behavioural aspects on the fraudulent review identification technique in the dining or lodging industries. [10]

III.PROPOSED METHODOLOGY

Several techniques for classification are developed for supervised machine learning. The primary purpose of these algorithms is to find acceptable models that propagate the training data.SVM is a differential classification that, in essence, separates the input data used for training into classes by selecting the best separated hyper-plane. The K-Nearest Neighbours algorithm (or KNN) is one of the simplest yet most effective classification approaches. KNN is most typically used in analysis and pattern recognition. The primary idea behind KNN is to classify instance requests according to the votes cast of a group of like classified examples. To calculate similarity, a distance function is commonly used.A decision-tree classifier is another machine learning classifier that employs a tree to represent an assessment of data used for training instances. The method starts iteratively building the tree according to the optimal feature split. Predetermined criteria such as entropy, mutual knowledge, knowledge gain, or the Gini index are used to select the optimal qualities.A random forest is a powerful strategy for dealing with overfitting difficulties that develop within decision trees. The core premise of random forest is to create a bag of branches from distinct dataset samples. Instead of generating the tree from all attributes, Random Forest chooses a tiny random quantity of qualities for every branch in the forest.Logistic regression is an additional straightforward machine learning-supervised classifier. Choosing a hyperplane that categorises the data is essential.

IV.IMPLEMENTATION

Figure 1 depicts the recommended strategy, which is detailed in depth in this section. In order to develop the most effective approach for fake review identification, the proposed method comprises three key steps. These phases are explained in more detail below:

A. DATA PREPROCESSING:

The first step in the proposed technique is data processing, which is one of the most important phases in learning methodologies. Data preparation is vital as the world's information is never usable. The raw data from the Yelp dataset has been preprocessed in the present investigation through a variety of phases to prepare it for computational tasks. The following is a synopsis:

1. TOKENIZATION:

Tokenization is a prominent method for natural language processing. It is a necessary step before proceeding with any other preparatory processes. Tokens are individual words that comprise the text. Tokenization, for example, will break down the line "wearing a helmet is an absolute necessity for pedal cyclists" into tokens that represent "wearing", "helmets", "is", "a", "must", "for", "pedal", and "cyclists".

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 2, June : 2023

2. Stop Word cleaning:

Considering the fact they may have no meaning, stop phrases are the most commonly used keywords. Stop - word examples include (an, a, the, and this). All data in this study are cleansed of stop words before proceeding to the technique for identifying fake reviews.

3. Lemmatization:

Use the lemmatization method to transform the number of form to a singular one. It seeks to remove just the inflectional endings and return the word to its dictionary-base form. For example, replace "plays" with "play".

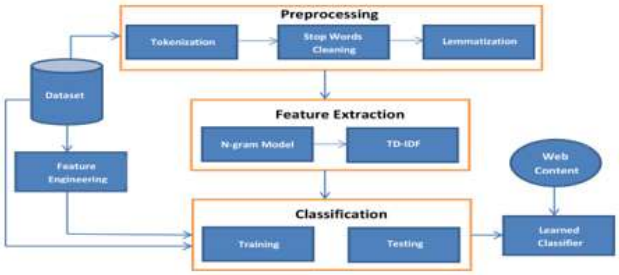

Fig. 1. The Proposed Framework.

B. Extraction of Characteristics:

The purpose of the extraction of features is to increase the performance of a pattern recognition or machine learning system. Feature extraction involves narrowing the input to its main features in order to deliver significantly more valuable data to machine learning and deep learning models. It is basically a way of removing extraneous characteristics from the data in order to improve the model's accuracy.

A number of strategies for extracting features for fake comment detection have been developed in the literature. Using textual features is a popular method. It contains a classification of feelings, which depends on the proportion of both positive and negative phrases in the review (such as "good" and "weak").

The The process of convolution is also considered. The cosine similarity is calculated by dividing the dot sum of the lengths of both vectors by the sin of the angle among the two n-dimensional vectors in n-dimensional space (ormagnitudes). to that with the with the to (IDF). Every word has a distinct TF and IDF score, and the term's TF-IDF weight is the sum of these two scores. A confusion matrix was used to break down the evaluations into four findings; TN: True Negative This category includes actual occurrences. False Reading (FP): True occurrences are classified as false, True Positive (TP): False actions are classified as false, and False Statement (FN): False events are classified as true.

The individual's behaviours and personal profile characteristics are ranked second. These two traits are used to identify spammers. Whether the individual produces repeated evaluations and has no link to the intended site, or if the user's remarks are more numerous and distinct from those of other frequent visitors. We employ TF-IDF to derive the principal bi- and tri-gram characteristics from the material of both language models in this work. After eliminating the variables that indicate user behaviour, we apply the expanded data to both language models.

C.FEATURE ENGINEERING :

Several descriptors of reviewer behaviour during review authoring are known to be present in fake reviews. We will look at a couple of these characteristics along with how they influence how effectively the fake reviews detection algorithm works in this study.We take into account behavioural aspects like emojis, punctuation, and caps-count. The overall number of capital letters, punctuation marks, and emojis used in each review is indicated by the caps-count, punct-count, and emojis fields, respectively. Furthermore, we used statistical analysis to investigate reviewer behaviour by utilising

the "groupby" tool, which determines the quantity of fake or legitimate reviews each reviewer has posted on a certain day and for each hotel. To determine the impact of user actions on the effectiveness of the classifiers, these all parameters are taken into account.

V.RESULTS AND DISCUSSION

We used the Yelp dataset to evaluate our suggested system. This dataset has 38, 063 reviews that total 5853 for 201 hotels in Chicago. The reviews are divided between 4,709 reviews that have been deemed legitimate and 1,144 reviews that have been deemed false. The reviews on Yelp have been divided into real and bogus ones. Furthermore, we investigated reviewer conduct using statistical analysis, namely the "groupby" tool, which estimates the number of fraudulent or authentic reviews each reviewer has made on a specific day and for each hotel.The data contains reviews with a maximum word count of 875, a minimum word count of 4, an average word count of 439.5, a total word count of 103052 for the data's tokens, and a total word count of 102739 for its unique words.In addition to the database and its statistics, we extracted additional elements that depict the actions made by reviewers when writing their reviews. Caps-count displays how many letters in capitals an examiner used overall, punctuation-count indicates how many punctuation points were utilised overall in each assessment, and emojis shows how many emoji were used altogether in each review.All of these characteristics will be taken into account as we examine how user behaviour affects the effectiveness of the classifiers.

Total number of reviews	5833 review
Number of fake reviews	1144 review
Number of real reviews	4709 review
Number of distinct words	102739 word
Total number of tokens	103052 token
The maximum review length	875 word
The minimum review length	4 word
The Average review length	439.5

Table 1:Summary of the Data

This section presents the results of several trials as well as a review of them using five different machine-learning classifiers. Then, from the material of the two language models, we utilise TF-IDF to extract the key bi- and tri-gram features. We apply the expanded data to both language models after deleting the variables that reflect the user behaviours indicated in the previous section. Because of the unequal distribution of positive and negative labels in the dataset, we also analyse accuracy and recall; as a result, f1-score is used as a performance metric in addition to accuracy.The database is utilised for testing 30% of the time and training 70% of the time. The classifiers are evaluated in both the event of and lack of the user activities gathered as features.

Fleiss's Kappa is an able to categorize of inter-rater consistency or agreement that assesses how well different raters agree on how to categorise or rank the same set of things. The Kappa statistic gives a standardised measure of cooperation that runs from -1 to 1, while also accounting for the potential of agreement arising by coincidence. There is no agreement outside chance when the Fleiss's Kappa value

is 0, while a value of 1 denotes perfect agreement. If the value is negative, there's less agreement than would be predicted by chance, which raises the possibility that the raters are competing with one another. In general, moderate to good agreement is defined as a Fleiss's Kappa value of 0.40 or higher, whereas poor agreement is defined as a value lower than 0.40. It's vital to remember that Fleiss's Kappa only works with categorical data, meaning that the ratings or categories must be distinct and incompatible.

```
Shape of Data after TFIDF: (1600, 3000)
Data Information:
  (0, 166) 0.13172930131568297
  (0, 2141) 0.10660780168260178
  (0, 2975) 0.06801677057022719
  (0, 2015) 0.1341834923690232
  (0, 1378) 0.1945432554999002
  (0, 306) 0.12210833167212995
  (0, 60)  0.149707522114893
  (0, 139) 0.154869457625446
  (0, 2327) 0.07536231362558732
  (0, 1316) 0.043214352008951934
  (0, 1509) 0.08969951756453394
  (0, 2342) 0.3890865109998004
  (0, 2416) 0.1635671569438979
  (0, 1572) 0.2800185827494882
  (0, 2458) 0.1361368967464473 8
  (0, 2104) 0.1392744928282928
  (0, 300) 0.10114735335217438
  (0, 2515) 0.21504809716192214
  (0, 1118) 0.1543197273630089
  (0, 2700) 0.19144928801100017
  (0, 2510) 0.07791772952665088
  (0, 671) 0.08579956292083955
  (0, 2930) 0.17219886635695233
  (0, 1109) 0.12031973898581974
  (0, 190) 0.2219126859886043
  :  :
  (1599, 2219) 0.16861639489909463
  (1599, 2360) 0.18031167610986085
```

Fig.2. TF-IDF values of the review

It is inappropriate for data that is continuous or interval.Fleiss's Kappa is a useful tool for assessing the consistency and correctness of ratings or categorizations produced by numerous raters since it provides an overall measure of inter-rater reliability

Table 2. Interpretation of Fleiss' Kappa (Richard Landis and Koch, 1977).

value	Interpretation
0	Poor agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–1.00	Almost perfect agreement

Table III lists the classifiers' accuracy when the users' extracted behaviours are taken into account in the bilingual models. The outcomes show that SVM, with a score of 0.89,Regression analysis has a accuracy of 1.0 ,KNN with an accuracy of 0.76 and random forest with an accuracy of 0.63.Thus,we get the highest accuracy of 1.0 in case of logistic regression.

Table 3: Accuracy of algorithms in the presence of extracted features behaviors

Classification Algorithm	Accuracy
Logistic regression	1.0
KNN	0.76
SVM	0.89
Random forest	0.63

VI.CONCLUSION

In this essay, we discussed the importance of reviews and how they affect almost all elements of web-based data. People's selections are certainly influenced by reviews. As a result, detecting fraudulent comments is an important and ongoing research subject. This paper describes a machine learning strategy for detecting fake reviews.Both the characteristics of the evaluations and the reviewers'

behavioural characteristics are taken into account in the suggested approach. The proposed approach is evaluated utilising the Yelp dataset. The developed technique employs a number of classifiers.

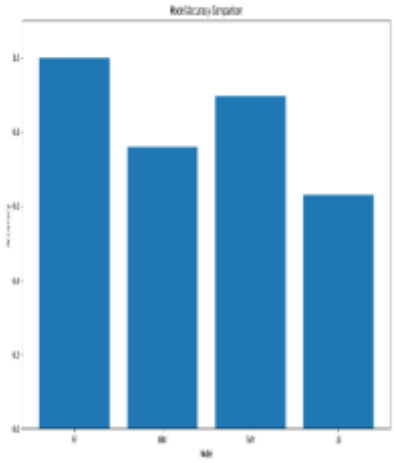

Fig. 3. Model accuracy comparision

The findings also indicate that taking into account the reviewers' behavioural characteristics.. Not all of the reviewers' behavioural traits were considered in the current study.Future research may take into account integrating more behavioural elements, such as those that depend on how frequently reviewers perform reviews, how long it takes them to finish reviews, and how frequently they submit good or negative evaluations. Adding more behavioural characteristics to the technique for detecting fraudulent reviews is expected to increase its performance.

REFERENCES

1)R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," *Information Processing & Management*, vol. 56, no. 4, pp. 1234 – 1244, 2019.

2)S. Tadelis, "The economics of reputation and feedback systems in e- commerce marketplaces," *IEEE Internet Computing*, vol. 20, no. 1, pp. 12–19, 2016.

3)M. J. H. Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *Information Retrieval*, vol. 9, no. 6, 2018.

4)C. C. Aggarwal, "Opinion mining and sentiment analysis," in *Machine Learning for Text*. Springer, 2018, pp. 413–434.

5)A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" in *Seventh international AAAI conference on weblogs and social media*, 2013.

6)N. Jindal and B. Liu, "Review spam detection," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07, 2007.

7)E. Elmurngi and A. Gherbi, *Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques*. IARIA/DATA ANA- LYTICS, 2017.

8)V. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews and blog posts," in *Advance Computing Conference (IACC)*, 2013, pp. 893–898.

9)A. Molla, Y. Biadgie, and K.-A. Sohn, "Detecting Negative Deceptive Opinion from Tweets." in *International Conference on Mobile and Wireless Technology*. Singapore: Springer, 2017.

10)S. Shojaee *et al.*, "Detecting deceptive reviews using lexical and syntactic features." 2013.

11)1Y. Ren and D. Ji, "Neural networks for deceptive opinion spam detection: An empirical study," *Information Sciences*, vol. 385, pp. 213–224, 2017.

12)H. Li *et al.*, "Spotting fake reviews via collective positive-unlabeled learning." 2014.

13)N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 International*

Conference on Web Search and Data Mining, ser. WSDM '08, 2008, pp. 219–230.
14) D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, "What online reviewer behaviors really matter? effects of verbal and nonverbal behaviors on detection of fake online reviews," *Journal of Management Information Systems*, vol. 33, no. 2, pp. 456–481, 2016.
15) E. D. Wahyuni and A. Djunaidy, "Fake review detection from a product review using modified method of iterative computation framework." 2016.
16) D. Michie, D. J. Spiegelhalter, C. Taylor *et al.*, "Machine learning,"*Neural and Statistical Classification*, vol. 13, 1994.
T. O. Ayodele, "Types of machine learning algorithms," in *New ad- vances in machine learning*. InTech, 2010.
17) F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002.
18) T. Joachims, "Text categorization with support vector machines: Learn- ing with many relevant features." 1998.