International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 647-651 www.ijaem.net ISSN: 2395-5252

Fake Reviews Detection using Supervised Machine Learning Algorithm

Rohinikhalkar[1], Murari Kumar Jha[2], Divyam Maru[3], Priyanshi Sharma[4]

[1] Assistant Professor, Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be)UniversityCollege ofEngineering, Pune,Maharashtra,India
[2,3,4]Student,DepartmentofComputer Engineering,BharatiVidyapeeth(Deemedtobe) UniversityCollegeof Engineering,Pune, Maharashtra, India

----
Submitted: 01-07-2022 Revised: 10-07-2022 Accepted: 13-07-2022

ABSTRACT—Online reviews are mostly regarded as a critical aspect for establishing and preserving a solid reputation as e-commerce systems continue to advance. Furthermore, they play a significant part in the final decision-making process for users. A favourable review for a target item typically draws additional clients, which increased revenue significantly. Nowadays, Reviews that are intentionally false or misleading are posted to create virtual enhance one's reputation and lure future clients. Thus, recognizing The study of false reviews is active and ongoing. Identifying bogus reviews depend on more than just the reviews main characteristics but also on the reviewers' actions. This essay suggests using machine learning to spot fraudulent reviews. The study evaluates many experiments' results using a real dataset. In this, we used a labelled dataset and performed machine learning algorithms like Logistic Regression, K Nearest Neighbors, Support Vector Classifier, Decision, Tree Classifier, Random Forests Classifier, and Multinomial Naive Bayes. The result reveals that the Support Vector Machine performs best.

Keywords—Fake reviews detection, supervised machine learning, data mining, feature engineering

I. INTRODUCTION

Reviews have evolved into the primary information source for consumers today when making decisions about services or products. Customers might browse evaluations about other people's experiences with hotel services before deciding to book a hotel, for instance. They select whether or not to reserve a room based on the feedback from the reviews. If they found the evaluations to be favourable, they will probably go ahead and reserve the room. As a result, historical analyses rose to the top of many web services' lists of highly reliable information sources. Reviews are viewed as genuine means of exchanging input about products and services, therefore any attempt to skew them by including false or misleading information is viewed as dishonest and is grounds for labelling the review as fake. Such a situation makes us wonder what would happen if not all submitted reviews are trustworthy or sincere. What if any of these testimonials are false? As a result, the detection of fraudulent reviews has become and is currently an active and important study subject.

Machine learning techniques can provide a big contribution to detecting fake reviews of web content. Generally, web mining techniques [1] find and extract useful information using several machine learning algorithms. Content mining is one of the web mining duties. A traditional example of content mining is opinion mining [2] which is concerned with finding the sentiment of text (positive or negative) by machine learning where a classifier is trained to analyse both the features and the sentiments of reviews.

The detection of phoney reviews typically focuses on variables that are not directly related to the content as well as the category of the reviews. Text and natural language processing NLP are typically used while creating review feature sets. However, creating factors relating to the reviewer himself, such as the review time/date or his writing style, may be necessary to create phoney reviews. Therefore, the development of useful feature extraction from the reviewers is key to the successful detection of false reviews.

To achieve this, this study employs several machine learning classifiers to detect fraudulent reviews using both the text of the reviews and some attributes that were retrieved from the reviewers.

The rest of this paper is organized as follows: Section II Summarizes the detecting of fake reviews. Section III states the background about the machine learning techniques. Section IV represents the details about proposed approach..

II. RELATED WORK

The fake reviews detection problem has been tackled since 2007 [3]. Textual and behavioural elements have been extensively used in the research on the detection of fake reviews. The verbal aspect of the review activity is referred to as textual characteristics. In other words, textual features mostly depend on the reviews' content. The nonverbal qualities of the reviews are referred to as behavioural traits. They largely depend on the reviewers' actions, including their writing style, facial expressions, and frequency of review writing. Although addressing textual features is difficult and vital, behavioural features are equally significant and cannot be disregarded because they have a significant impact on how well the fake review detection process works. Textual features have extensively been seen in several fake review detection research papers. In [4], the authors used supervised machine learning approaches for fake review detection. Five classifiers are used which are **SVM**, Naive-Bayes, KNN, k-star and decision tree. Three versions of the labelled movie review dataset [5] with a total of 1400, 2000, and 10662 movie reviews each have undergone simulation testing. Additionally,the authors employed classifiers such as Naive Bayes, Decision Tree, **SVM**, Random Forest, and Maximum Entropy to find fraudulent reviews in their dataset. Against 10,000 unfavourable tweets about Samsung goods & services were dataset. The authors of [6] employed both **SVM** and Naïve bayes classifiers. The authors used the yield dataset which has 1600 reviews gathered from 20 well-known Chicago hotels. To identify false opinion spam, the authors of utilized neural and discrete models with average, CNN, RNN, GRNN, average GRNN, and bi-directional average GRNN deep learning classifiers to detect deceptive opinion spamming.

All the above research works have only considered the textual features without any effort towards the behavioural features.

According to the aforementioned discussion and to the best of our knowledge, no methods have gone particularly deep in identifying attributes that represent the reviewers' behaviour. These characteristics will have a significant impact on how well the bogus review detection method works. This research presents a machine learning method to spot bogus reviews. The proposed solution combines numerous features engineering techniques in addition to the reviews' feature extraction process to extract different reviewer behaviours. New behavioural traits are developed. In addition to textual features, the produced features are used as inputs to the suggested system to detect bogus reviews.

III. BACKGROUND

One of the most significant technology trends, machine learning is the foundation of many essential applications. The major benefit of machine learning is that it enables machines to automatically learn from their past mistakes and advance.

Machine learning algorithms come in a variety of flavours, including supervised, semi-supervised, and unsupervised algorithms. Both input and output data are given in the surprising technique, and the training data must be annotated and categorised. The goal of the unsupervised learning strategy is to determine the most appropriate grouping or classification of the input data as no labels or classifications are provided. As a result, in unsupervised learning, the approach's job is to identify the unlabeled data.

For supervised machine learning, several classification algorithms are created. These algorithms' main goal is to identify a suitable model that disseminates the training data. By selecting the best separable hyper-plane that categorises the provided training data, Support Vector Machines (SVM), for instance, is a discriminating classifier that essentially divides the given data into classes. Naive Bayes is another popular supervised learning algorithm (NB). The core concept of NB is based on the Bayes theorem, which states that

P(A—B) = P(B—A)*P(A) P(B)

is the likelihood of event A occurring given the probability of event B. By counting the frequency and the total values in a dataset, NB determines a set of probabilities. **The K-Nearest Neighbors algorithm (or KNN)** is one of the most simple yet powerful classification algorithms. KNN is used mostly in statistical estimation and pattern recognition techniques. The key idea behind KNN is to classify instance queries based on the voting of a group of similar classified instances. The similarity is usually calculated using the distance function.

Another machine learning classifier that focuses on creating a tree to represent a choice of training data is called a **decision-tree**. Based on the best feasible division among characteristics, the algorithm begins to iteratively build the tree. A predetermined function, such as entropy, information gain, gain ratio, or Gini index, is used to determine which characteristics are the best.

The overfitting issues that arise in the decision tree are successfully addressed by **Random Forest**. Building a bag of trees from various dataset samples is the fundamental principle of random forest. When building each tree in the forest, Random Forest selects a tiny random number of features as opposed to building the tree from all features. Another straightforward supervised machine learning classifier is logistic regression . Finding a hyperplane that categorises the data is essential.

IV. PROPOSED APPROACH

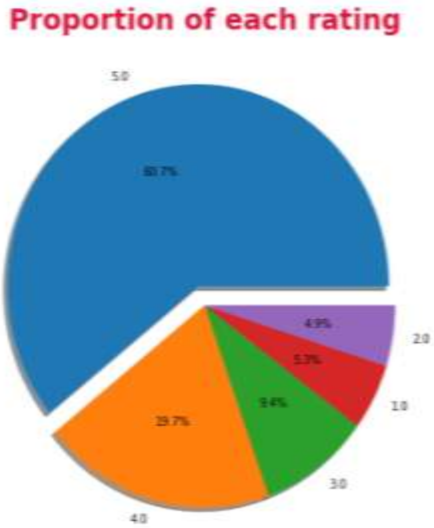
The specifics of the suggested technique are explained in this section. To acquire the optimum model for fake review detection, the suggested approach comprises of three fundamental phases. The following explains these phases:

(A). Data Preprocessing

- Data preprocessing, one of the crucial phases in machine learning methodologies, is the initial step in the suggested approach. Data preparation is essential since the world's data is never suitable for use. The first step in the proposed approach is data preprocessing one of the essential steps in machine learning approaches. Data preparation is essential since the world's data is never suitable for use. So to preprocess data we do the following
- Removing punctuation character
- Transforming text to lower case
- Eliminating stopwords
- Stemming
- Lemmatizing
- Removing digits

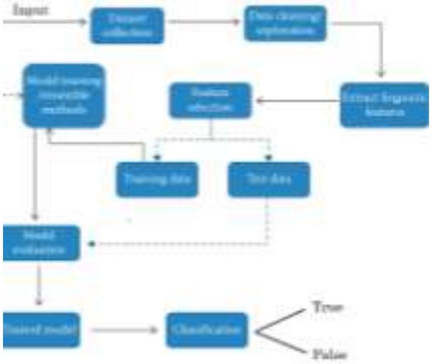
(B). Feature extraction

Feature extraction is a step which aims to increase the performance either for a pattern recognition or machine learning system.In order to provide machine learning and deep learning models with more useful data, feature extraction involves reducing the input to its key features. It is mainly a procedure of removing the unneeded attributes from data that may actually reduce the accuracy of the model.Several approaches have been developed in the literature to extract features for fake review detection. Textual features is one popular approach. It contains sentiment classification which depends on getting the per cent of positive and negative words in the review; e.g. "good", "weak". Also, the Cosine similarity is considered.The cosine similarity is equal to the dot product of the lengths of the two vectors divided by the cosine of the angle formed by the two n-dimensional vectors in the n-dimensional space. The frequency of both true and false (TF) as well as the inverse document are obtained by another textual feature method called TF-IDF (IDF). Each phrase has a unique TF and IDF score, and the sum of these two scores is referred to as the term's TF-IDF weight. A confusion matrix is used to classify the reviews into four results; True Negative (TN): Real events are classified as real events, True Positive (TP): Fake events are classified as fake, False Positive (FP): Real events are classified as fake events, and False Negative (FN): Fake events are classified as real.



(C). Accuracy of different models

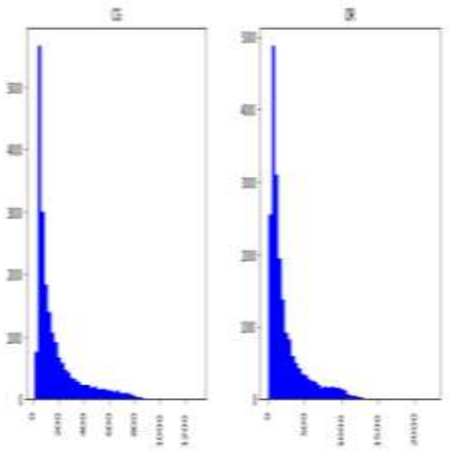
In this, we calculate the classification report, confusion matrix and accuracy score of different models and based on that we select one model and then with flask module of python and HTML,CSS we create a frontend where user enter the review and he/she get the result based on the training of our mode.

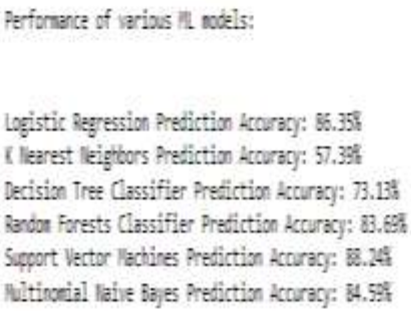


The flow of the module in Fake Reviews Detection System

V. RESULTS

We evaluated our proposed system on a fake review dataset taken from an online source in which there is labelled dataset in which their are two categories CG i.e computer-generated fake reviews and OR i.e original reviews,then by using all supervised learning algorithms we check the accuracy by measuring the confusion matrix, accuracy score and classification report. After checking all the accuracy we came to know that the support vector machine gives the best results.





```
Performance of various ML models:


Logistic Regression Prediction Accuracy: 86.35%
K Nearest Neighbors Prediction Accuracy: 57.39%
Decision Tree Classifier Prediction Accuracy: 73.13%
Random Forests Classifier Prediction Accuracy: 83.69%
Support Vector Machines Prediction Accuracy: 88.24%
Multinomial Naive Bayes Prediction Accuracy: 84.59%
```

VI. CONCLUSION

In this essay, we highlighted the significance of reviews and how they impact nearly every aspect of web-based data. Reviews undoubtedly influence people's choices, as is evident. As a result, the detection of fraudulent reviews is a lively and active study subject.

This research presents a machine learning strategy for detecting bogus reviews. The qualities of the reviews and the reviewers' behavioural characteristics are both taken into account in the proposed approach. The developed technique uses a variety of classifiers. According to the findings, the SVM classifier performs better than the other classifiers at identifying bogus reviews. In the current work, not all reviewers' behavioural characteristics have been taken into account. Future research may take into account integrating more behavioural aspects, such as features 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. It is strongly anticipated that adding more behavioural variables to the strategy for detecting bogus reviews would improve its performance.

ACKNOWLEDGMENT

We want to thank our guide our guide Prof. Rohini Khalkar, who generously provided knowledge and expertise and without whom the project and paperwork could not have been possible.

FUTURE SCOPE

- Used to make a better decision based on original reviews
- It has a wide scope in the coming time as all products, hotels are decided by checking the reviews.
- It is used to calculate the risk associated with any decision.

REFERENCES

[1]. M. J. H. Mughal, “Data mining: Web data mining techniques, tools and algorithms: An overview,” Information Retrieval, vol. 9, no. 6, 2018.

[2]. C. K. Aggarwal, “Opinion mining and sentiment analysis,” in Machine Learning for Text. Springer, 2018, pp. 413–434.

[3]. N. Jind and B. Lie, “Review spam detection,” in Proceedings of the 16th International Conference on World Wide Web, ser. WWW ’07, 2007.

[4]. E. Elmurngi and A. Gerbi, Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. IARIA/DATA ANALYTICS, 2017

[5]. S. Shojaee et al., “Detecting deceptive reviews using lexical and syntactic features.” 2013

[6]. H. Le et al., “Spotting fake reviews via collective positive-unlabeled learning.” 2014

[7]. A. Liaw, M. Wiener et al., "Classification and regression by random forest," R News, vol. 2, no. 3, pp. 18–22, 2002.

[8]. D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, Logistic regression. Springer, 2002.

[9]. G. G. Chowdhury, "Natural language processing," Annual review of information science and technology, vol. 37, no. 1, pp. 51–89, 2003.

[10]. J. J. Webster and C. Kit, "Tokenization as the initial phase in NLP," in Proceedings of the 14th conference on Computational linguistics volume 4. Association for Computational Linguistics, 1992, pp. 1106–1110.

[11]. J. Plisson, M. Lavrac, D. Mladenic´ et al., "A rule-based approach to word lemmatization," 2004.