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

FAKE REVIEWS DETECTION USING SUPERVISED MACHINE

[1]Ms. R Abinaya, [2]Jose Pravin A, [3]Manoej KR, [4]Nithin DS

[1] Assistant Professor, [2]Student, [3]Student, [4]Student
[1234]Department of Computer Science and Engineering,
[1234]Sri Ramakrishna Institute of Technology, Coimbatore, India

Abstract : With increasing use of internet, online business platforms are becoming the largest market place of the world. Purchase of online product is heavily dependent on user reviews. Some dishonest groups of people misuse this fact by posting fake reviews to promote their own products or demote their competitors. Detection of fake online reviews can be considered as a binary classification task that models a classifier to tell whether review is fake or true. Developed an effective supervised machine learning approach to classify fake online reviews using a dataset that contains hotel reviews from online websites. Online reviews, supervised learning, support vector machine, naive Bayes, logistic regression. Online reviews have great impact on today's business and commerce. Decision making for purchase of online products mostly depends on reviews given by the users.

IndexTerms **– Support vector machine, naive Bayes , logistic regression.**

I. Introduction

Purchasing products online has become an integral part of our daily routine, with platforms like Amazon, AliExpress, eBay, and others offering a wide range of items. Whether it's shopping for goods or planning travel, such as booking hotels and flights, online services cater to diverse needs. However, the absence of physical interaction with products or services makes reliance on online reviews paramount in the decision-making process.

Before making a purchase, individuals often turn to platforms like Amazon, AliExpress, and eBay to gauge the quality of a product or service through the experiences shared by other users. This heavy reliance on online reviews has, unfortunately, opened avenues for malicious entities seeking to manipulate perceptions through fake comments and reviews. These deceptive practices aim to either promote their own products or tarnish the reputation of competitors. Therefore, the identification and prevention of fake online reviews hold significant importance for consumers seeking genuine feedback and for companies striving to maintain their credibility

II. Problem Statement

Given a dataset of review,each labeled as either genuine or fake,the objective is to develop a machine learning model that can accurately classify reviews as genuine or fake. The model should be trained on feature extracted from the text of the review ,such as sentiment, language patterns, and syntactic structures. The goal is to create a reliable system capable of automatically detecting fraudulent or deceptive reviews,which can help businesses maintain the integrity of their online reputation and assist consumers in making informed decisions.

III. Scope of the project

- Gather diverse datasets of reviews.
- Develop supervised machine learning models.
- Deploy the model for real-time fake review detection.

IV. Literature Survey

[1.] Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training Logistic regression is to establish a cost function in the face of a regression or classification problem, and then iteratively solve the optimal model

parameters through optimization method, and then test and verify the quality of the model. It is the primary approach to data classification in the field of big data and machine learning. Traditional logistic regression uses the gradient descent method to solve the optimal parameter of the missing function. However, swarm intelligence algorithms, such as artificial fish swarms, can replace the traditional penalty function method to some extent to ensure the global convergence of the optimization algorithm. Artificial swarm intelligence algorithm can not only realize simple iterative optimization but also has simple environmental adaptability and system self-regulation.

[2.] Fake review detection on online E-commerce platforms: a systematic literature review The increasing popularity of online review systems motivates malevolent intent in competing sellers and service providers to manipulate consumers by fabricating product/service reviews. Immoral actors use Sybil accounts, bot farms, and purchase authentic accounts to promote products and vilify competitors. Facing the continuous advancement of review spamming techniques, the research community should step back, assess the approaches explored to date to combat fake reviews, and regroup to define new ones. This paper reviews the literature on Fake Review Detection (FRD) on online platforms. It covers both basic research and commercial solutions, and discusses the reasons behind the limited level of success that the current approaches and regulations have had in preventing damage due to deceptive reviews.`

[3.] Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews In recent years, fake review detection has attracted significant attention from both businesses and the research community. For reviews to reflect genuine user experiences and opinions, detecting fake reviews is an important problem. Supervised learning has been one of the main approaches for solving the problem. However, obtaining labeled fake reviews for training is difficult because it is very hard if not impossible to reliably label fake reviews manually. Existing research has used several types of pseudo fake reviews for training. Perhaps, the most interesting type is the pseudo fake reviews generated using the Amazon Mechanical Turk (AMT) crowdsourcing tool. Using AMT crafted fake reviews, [36] reported an accuracy of 89.6% using only word n-gram features. This high accuracy is quite surprising and very encouraging. However, although fake, the AMT generated reviews are not real fake reviews on a commercial website. The Turkers (AMT authors) are not likely to have the same psychological state of mind while writing such reviews as that of the authors of real fake reviews who have real businesses to promote or to demote. Our experiments attest this hypothesis.

[4.] Dynamic knowledge graph based fake-review detection Online product reviews are an important driver of customers' purchasing behavior. Fake reviews seriously mislead consumers, challenging the fairness of the online shopping environment. Although the detection of fake reviews has progressed, several problems remain. First, fake comment recognition ignores the correlation between time and the semantics of the comment texts, which is always hidden in the context of the reviews. Second, the impact of multi-source information on fake comment recognition is not considered, as it constitutes a complex, high-dimensional, heterogeneous relationship between reviewers, reviews, stores and commodities. To overcome these problems, the present paper proposes a dynamic knowledge graph-based method for fake-review detection. Based on the characteristics of online product reviews, it first extracts four types of entities using a developed neural network model called sentence vector/twinword embedding conditioned bidirectional long short-term memory.

[5.] A framework for fake review detection in online consumer electronics retailers The impact of online reviews on businesses has grown significantly during last years, being crucial to determine business success in a wide array of sectors, ranging from restaurants, hotels to e-commerce. Unfortunately, some users use unethical means to improve their online reputation by writing fake reviews of their businesses or competitors. Previous research has addressed fake review detection in a number of domains, such as product or business reviews in restaurants and hotels. However, in spite of its economical interest, the domain of consumer electronics businesses has not yet been thoroughly studied. This article proposes a feature framework for detecting fake reviews that has been evaluated in the consumer electronics domain. The contributions are fourfold: (i) Construction of a dataset for classifying fake reviews in the consumer electronics domain in four different cities based on scraping techniques; (ii) definition of a feature framework for fake review detection; (iii) development of a fake review classification method based on the proposed framework and (iv) evaluation and analysis of the results for each of the cities under study.

[6.] Exploiting Product related review features for fake review detection Product reviews are now widely used by individuals for making their decisions. However, due to the purpose of profit, reviewers game the system by posting fake reviews for promoting or demoting the target products. In the past few years, fake review detection has attracted significant attention from both the industrial organizations and academic communities. However, the issue remains to be a challenging problem due to lacking of labelling materials for supervised learning and evaluation. Current works made many attempts to address this problem from the angles of reviewer and review. However, there has been little discussion about the product related review features which is the main focus of our method. This paper proposes a novel convolutional neural network model to integrate the product related review features through a product word composition model. To reduce overfitting and high variance, a bagging model is introduced to bag the neural network model with two efficient classifiers. Experiments on the real-life Amazon review dataset demonstrate the effectiveness of the proposed approach.

[7.] Creating and detecting fake review of online products Customers increasingly rely on reviews for product information. However, the usefulness of online reviews is impeded by fake reviews that give an untruthful picture of product quality. Therefore, detection of fake reviews is needed. Unfortunately, so far, automatic detection has only had partial success in this challenging task. In this research, we address the creation and detection of fake reviews. First, we experiment with two language models, ULMFiT and GPT-2, to generate fake product reviews based on an Amazon e-commerce dataset. Using the better model, GPT-2, we create a dataset for a classification task of fake review detection. We show that a machine classifier can accomplish this goal near-perfectly, whereas human raters exhibit significantly lower accuracy and agreement than the tested algorithms. The model was also effective on detected human generated fake reviews. The results imply that, while fake review detection is challenging for humans, "machines can fight machines" in the task of detecting fake reviews. Our findings have implications for consumer protection, defense of firms from unfair competition, and responsibility of review platforms.

[8.] From conflicts and confusion to doubts: examining review inconsistency for fake review detection: Inconsistency in online consumer reviews (OCRs) may cause uncertainty and confusion to consumers when they make purchase decisions. However, there is a lack of a systematic and empirical investigation of review inconsistency in the literature. This research

characterizes review inconsistency from multiple aspects, including rating-sentiment, content, and language, and proposes hypotheses inconsistency in the literature. This research characterizes review inconsistency from multiple aspects, including rating-sentiment, content, and language, and proposes hypotheses.

[9.] Intelligent fake review detection based in aspect extraction and analysis using deep learning In the era of social networking and e-commerce sites, users provide their feedback and comments in the form of reviews for any product, topic, or organization. Due to high influence of reviews on users, spammers use fake reviews to promote their product/organization and to demote the competitors. It is estimated that approximately 14% of reviews on any platform are fake reviews. Several researchers have proposed various approaches to detect fake reviews. The limitation of existing approaches is that complete review text is analysed which increases computation time and degrades accuracy. In our proposed approach, aspects are extracted from reviews and only these aspects and respective sentiments are employed for fake reviews detection. Extracted aspects are fed into CNN for aspect replication learning. The replicated aspects are fed into LSTM for fake reviews detection. As per our knowledge, aspects extraction and replication are not applied for fake reviews detection which is our significant contribution due to optimization it offers. Ott and Yelp Filter datasets are used to compare performance with recent approaches. Experiment analysis proves that our proposed approach outperforms recent approaches. Our approach is also compared with traditional machine learning techniques to prove that deep neural networks perform complex computation better than traditional techniques.

[10]. Enhancing NLP techniques for fake review detection We are in the era of internet where people are more techno-savvy and they surf internet before buying a single item. Since buying a product online is easy and convenient these days, people also tending towards it as it save time and sometimes money. Also many branded products can be bought without thinking much about quality as name is enough for branded item. Nowadays various vendors also advertise their products through social media like facebook, whatsapp etc. Thus it is an extremely important to check their reliability before buying product. Buyer or client wants to check the opinion of other buyers regarding their purchase for that product. Most of the times review given by the user is not considered genuine as review was given without buying it. Sometimes review contains unrelated words.

V. METHODOLOGY:

Several classification algorithms are developed for supervised machine learning.The main objective of these algorithms is to find a proper model that disseminates the training data. For example, Support Vector Machines (SVM) is a discriminated classifier that basically separates the given data into classes by finding the best separable hyper-plane which categorizes the given training data[19]. Another Common supervised learning algorithm is Naive Bayes (NB). The key idea of NB relies on Bayes theorem; the probability of event A to happen given the probability of event B which is formed as P(A—B) = P(B—A)*P(A) P(B) [20]. NB calculates a set of probabilities by counting the frequency and the combined values in a given dataset. NB has been successfully applied in several application domains like text classification, spam filtering and recommendation,systems.

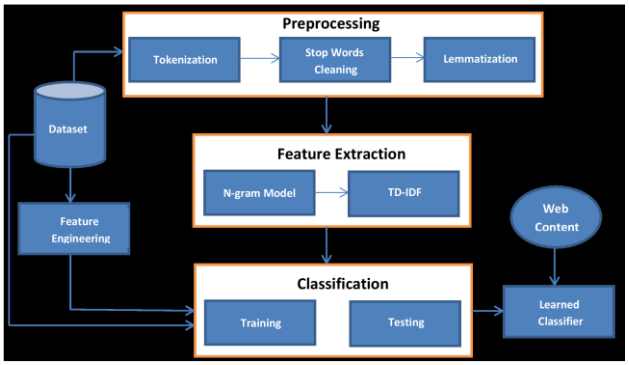

Figure 5.1 Workflow

Algorithm Used:

Multinomial Naive Bayes:

The Multinomial Naive Bayes model is a probabilistic machine learning algorithm that is commonly used for text classification and document categorization tasks. It is a variant of the Naive Bayes algorithm that assumes the features (or words in the case of text classification) are generated from a multinomial distribution. It's particularly well-suited for tasks involving discrete data, such as word counts in text documents. Calculate the prior probabilities for each class (category) by counting the frequency of each class label in the training data. Estimate the conditional probabilities of each word given a class label by counting the frequency of each word within documents belonging to that class.

Support Vector Machine:

Detecting fake reviews using supervised machine learning, specifically Support Vector Machine (SVM), involves building a model that can distinguish between genuine and fraudulent reviews based on various features extracted from the reviews. In this project, the aim is to develop a robust system capable of accurately identifying fake reviews, which are prevalent in online platforms and can mislead consumers.The process begins with data collection, where a dataset containing labeled examples of both genuine and fake reviews is gathered. These reviews can be sourced from e-commerce websites, social media platforms, or any other online review platform. Each review is associated with a label indicating its authenticity.Next, the collected data is preprocessed to extract relevant features that can be used to train the SVM model. Feature extraction is a crucial step and may involve various techniques such as text processing, sentiment analysis, lexical analysis, and syntactic analysis. These techniques help in capturing important characteristics of the reviews, such as the language used, sentiment expressed, grammatical structure, and more. Once the features are extracted, the dataset is divided into training and testing sets. The training set is used to train the SVM classifier using supervised learning techniques. SVM is a powerful algorithm for classification tasks, particularly suitable for binary classification problems like fake review detection. It works by finding the optimal hyperplane that separates the two classes in the feature space with the maximum margin. In conclusion, the project on detecting fake reviews using supervised machine learning with Support Vector Machine involves collecting, preprocessing, and extracting features from a dataset of reviews, training an SVM classifier on labeled data, and evaluating its performance in accurately identifying fake reviews. By leveraging machine learning techniques, this project contributes to combating fraudulent activities in online review systems and enhancing consumer trust and confidence.

VI. EXPERIMENTAL RESULTS

This Python code implements a web application for detecting fraudulent reviews in online consumer platforms using machine learning techniques. It utilizes Streamlit as the framework for building the user interface. The functionality of the application is divided into several components are Loading Pickle Files: The code loads pre-trained machine learning models and vectorizers from pickle files. These files contain the trained logistic regression model and count vectorizer used for text classification. Text Preprocessing: Before classification, the entered review text undergoes preprocessing. This includes correcting spelling errors using TextBlob, removing special characters, converting text to lowercase, removing stopwords, and stemming the words using the PorterStemmer algorithm. Text Classification: The preprocessed text is passed through the trained model for classification. The model predicts whether the review is legitimate or fraudulent based on the extracted features. The result is displayed to the user, indicating whether the review entered is genuine or fake.

Figure 6.1: Scikit Interface

User Interface: The web application provides a user-friendly interface where users can input their reviews and receive instant feedback on their authenticity. includes expandable sections providing an abstract of the project, related links, information about the classifier used, and evaluation results, offering users insights into the methodology and performance of the classifier. Overall, this code creates an interactive tool for detecting fake reviews, empowering users to make informed decisions while navigating online consumer,platform.

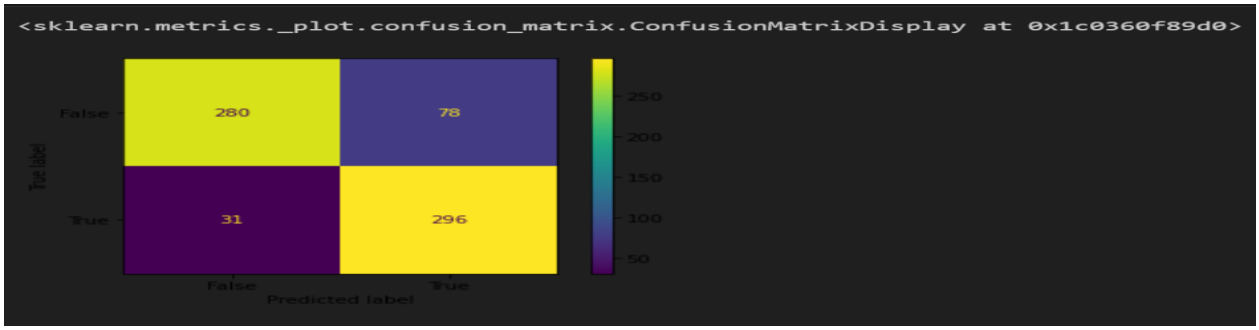

Figure 6.2:Confusion Matrix

The provided code segment employs a Support Vector Machine (SVM) classifier for binary classification tasks. Initially, an SVM instance is created with a linear kernel, denoted as `svm1`. This kernel facilitates linear separation of the data points in the feature space. Subsequently, the SVM model is trained using the `fit` function, where it learns to distinguish between different classes based on the features extracted from the training data (`train_c`) and their corresponding labels (`y_train`). Once trained, the model is deployed to make predictions on unseen data using the `predict` function, which assigns class abels to the test data (`test_c`). This process enables the SVM classifier to classify new instances accurately based on the patterns learned during training.

VII. **CONCLUSION:**

In order to solve the problem that large-scale labeled datasets are difficult to obtain under the full supervision framework, this study proposed a fake review detection model based on the combination of multi-feature fusion and rolling collaborative training. Experimental results show that this method is more effective than traditional algorithms. It uses unlabelled data to improve the performance of the classification system, and has better classification accuracy. At the same time,the consistency of sentiment and score is analyzed, and the feature extraction of the review is carried out through the text representation model, and the feature fusion is combined with the external features of the text, which can effectively improve the classification effect of the classification model.

REFERENCES

[1] G. Lackermair, D. Kailer, and K. Kanmaz, ``Importance of online product reviews from a consumer's perspective," Adv ,Econ.Bu ., vol. 1, no. 1.

[2] D. S. Kostyra, J. Reiner, M. Natter, and D. Klapper, ``Decomposing the effects of online customer reviews on brand, price, and product attributes," *Int. J. Res. Marketing* , vol. 33, no. 1, pp. 11_26.

[3] S. Ullrich and C. B. Brunner, ``Negative online consumer reviews: Effects of different responses," *J. Product Brand Manage.*, vol. 24, no. 1, pp. 66_77.

[4] S. Deng, C.-X. Wan, A.-H. Guan, and H. Chen, ``Deceptive reviews detection of technology products based on behavior and content," *J. Chin. Comput. Syst.*, vol. 36, no. 11, p. 2498, .

[5] D. Radovanovic and B. Krstajic, ``Review spam detection using machine learning," in *Proc. 23rd Int. Sci.-Prof. Conf. Inf. Technol.*,, pp. 1_4.

[6] H. Deng, L. Zhao, N. Luo, Y. Liu, G. Guo, X. Wang, Z. Tan, S. Wang, and F. Zhou, ``Semi-supervised learning based fake review detection," in *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl.*,, pp. 1278_1280.

[7] H. Ahmed, I. Traore, and S. Saad, ``Detecting opinion spams and fake news using text classification," *Secur. Privacy*, vol. 1, no. 1, p. e9.

[8] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, ``Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, p. 23.

[9] N. Jindal and B. Liu, ``Opinion spam and analysis," in *Proc. Int. Conf.Web Search Data Mining*, , pp. 219_230.

[10] K. H. Yoo and U. Gretzel, ``Comparison of deceptive and truthful travel reviews," in *Proc. ENTER*,, pp. 37_47.