International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fraudulent Review Detection using Machine Learning Techniques

Bodem Vinay[1]

B. Tech Student, Department of CSE-AI&DS, GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India
Email: 21341A4510@gmrit.edu.in[1]

ABSTRACT

Machine learning is a fancy way for computers to learn from past experiences and use that knowledge to make predictions about the future. This paper is all about finding and getting rid of fake reviews that people sometimes write online. When we shop online, we often read reviews to decide if a product is good or not. But some reviews are fake and not honest. This paper wants to create a system that can tell if a review is fake and remove it. This helps protect people, products, and online stores. The main focus is on using a special kind of computer program that can read the feeling behind a review and figure out if it's genuine or fake. The proposal entails the development of a supervised learning-based approach, specifically Sentiment Analysis, to identify inauthentic reviews. The research primarily targets the E-commerce sector, with a focus on assessing the efficacy of three supervised machine learning algorithms: linear regression, logistic regression , Decision Tree, Support vector machine, K-Nearest Neighbours (KNN) . These algorithms are employed to decipher patterns and similarities within a dataset containing consumer reviews for smartphone products and to increase the accuracy to detect the fake review.

Keywords: Fake review detection, Sentiment Analysis, Supervised learning, machine learning, E-commerce, spam review, Review authenticity.

Introduction

In today's digital age, the Internet has become an integral part of our lives and has revolutionized the way we shop, socialize and search for information. Among the many conveniences it offers, online shopping has gained immense popularity due to its ease and accessibility. When shopping online, consumers rely heavily on product reviews to make informed decisions about the quality and suitability of the products they intend to purchase, and not just online shopping everywhere, where reviews play a major role. However, this reliance on online reviews has created a significant challenge – fake reviews.

Fake reviews, also known as spam reviews, are deceptive and dishonest reviews posted by individuals or entities with malicious intent. These reviews are designed to manipulate consumer perceptions, increase product ratings, and ultimately increase sales, often at the expense of actual consumer feedback. Detecting and mitigating the spread of fake reviews has become a critical issue for e-commerce platforms, consumers and businesses alike.

Machine learning (ML) has emerged as a powerful tool to tackle the problem of fake reviews. ML uses the vast amount of data available online to learn patterns, make predictions, and uncover hidden insights. In the context of online reviews, ML can be used to identify emotional tone, language patterns, and other subtle cues that differentiate genuine reviews from fraudulent ones. This article delves into the area of fake review detection using ML with a specific focus on sentiment analysis as a supervised learning approach. Sentiment analysis (opinion mining), a subset of ML, is the process of recognizing the sentiment or emotion expressed in text, such as a product review.

Literature Review

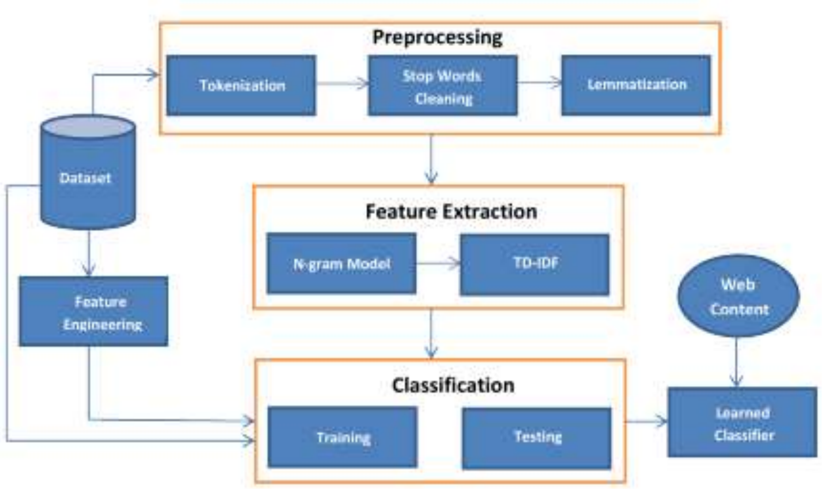
[1] Elmogy- The aim of the paper is to propose a machine learning approach to identify fake reviews on online platforms. The paper focuses on extracting key features from reviews and reviewer behavior to improve the accuracy of classification models. The study involves the use of TF-IDF to extract features from review content, allowing the importance of various words and phrases to be captured. In addition, the study considered both bigram and trigram language models, which can capture different levels of context in reviews. A limitation of the study was the unbalanced nature of the dataset in terms of positive and negative labels, which could affect the performance of the classifiers. Logistic Regression , K-Nearest Neighbors (KNN), Random Forest, Support vector Machine (SVM) algorithms used in this with Logistic Regression accuracy: F-score = 81.41%, KNN:F-score = 83.73%, SVM : F-score = 81.32%, Random Forest: F-score = 81.34%. The paper mentions the inclusion of reviewer behavioral features in the fake review detection process, but does not provide any specific details about these features or how they were extracted. This lack of detail makes it difficult to understand the impact of these behavioral features on classifier performance.

[2] Barbado - The aim of the study was to address the problem of detecting fake reviews in the consumer electronics domain, design a feature framework for social site analysis, develop a dataset for future research, and evaluate the effectiveness of features using classification algorithms. The study includes

the use of user-centric features, which have shown better results compared to review-centric features. The framework divided user-centered functions into four subgroups: Social Functions (S), Personal Functions (P), Trust Functions (T), and Control Functions (RA). Combining different feature subsets resulted in an increase in the F-score, indicating the effectiveness of the framework. The study includes a focus only on the consumer electronics domain and four specific US cities. The study also acknowledges that fake reviewers can mask their invalid reviews, making them difficult to detect based on the text of the review alone. Algorithms used in the study for classification were Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naive Bayes (NB) with performance metric values used in the study were F-scores which were obtained for each experiment was performed in different cities and with different combinations of elements and classification algorithms. F-scores ranged from 0.71 to 0.82, with the highest scores achieved in New York City.

[3] Jindal. N - The aim of the study is to highlight the prevalence of review spam, shed light on the credibility of online reviews, and uncover possible spam activities. The aim of the study is to achieve these objectives through duplicate detection using classification methods. The advantage of using logistic regression for classification is the ability to generate probability estimates for each review that is a spam check, which is necessary for spam detection. A logistic regression model has proven effective in detecting spam reviews. The study includes the difficulty of manually identifying reliable spam reviews, as spam reviews can be made to look similar to innocent reviews. The algorithm used in this involves logistic regression (LR) with the performance metric used to evaluate the classification model being accuracy. The average accuracy value obtained using 10-fold cross-validation is 78%, which is considered a high value. A gap in the study is the lack of reported research on review credibility and the detection of review spam. The aim of the study is to fill this gap by investigating review spam and proposing methods for its detection.

Methodology



Logistic Regression (LR): Logistic regression is a binary classification algorithm that models the probability of an input belonging to one of two classes. It uses a logistic function to transform a linear combination of input features to a value between 0 and 1, making it suitable for probabilistic classification tasks.

K Nearest Neighbor (KNN): The k-Nearest Neighbors algorithm is a non-parametric instance-based classification and regression method. Classifies data points based on the majority class among their k nearest neighbors in the training dataset.

Support Vector Machines (SVM): Support Vector Machines are a powerful supervised learning algorithm used for classification and regression. SVM finds the hyperplane that best divides the data into different classes while maximizing the range between classes.

Decision Tree (DT): Decision trees are a supervised learning algorithm used for classification and regression. They recursively partition the data into subsets based on the values of the input elements, creating a tree structure where the leaves represent class labels or numeric values.

Random Forest (RF): Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. Each tree in the forest is created using a random subset of data and features.

Dataset: The author worked on a revenue dataset consisting of 1600 reviews collected from 20 popular hotels in Chicago. The collected dataset contains approximately 10,000 negative tweets related to Samsung products and services.

Data preprocessing: Tokenization: The text is divided into individual words or tokens.

Stop word cleaning: Common stop words (eg "an", "a", "the", "toto") are removed from the data.

Lemmatization: Plural forms of words are converted to their singular forms to remove inflectional endings.

Feature Extraction: Text Features: This includes sentiment classification, such as determining the percentage of positive and negative words in a review. Cosine similarity is considered as a measure of similarity between reviews. TF-IDF is used to assign weights to terms based on their frequency and importance.

Confusion matrix: This matrix is used to classify reviews into four categories: true negative (TN), true positive (TP), false positive (FP) and false negative (FN).

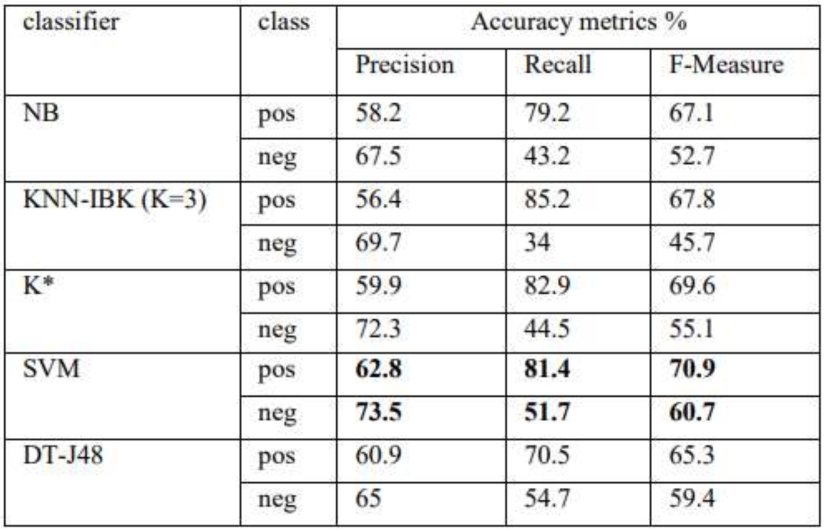
User Personal Profile and Behavioral Features: These features are used to identify spammers by analyzing the timestamp of the user's comments and their uniqueness compared to regular users. Redundant reviews unrelated to the target domain are also considered behavioral features.

Language models: Text features are extracted using bigram and trigram models that capture sequences of two and three words.

Feature engineering: Several features of reviewer behavior are considered during the review writing process. These features include caps-count (total number of capital letters used in the review), punct-count (total number of punctuation marks in the review), and emojis (total number of emoticons in the review).

Evaluation metrics: In the absence of extracted behavioral features, accuracy, precision, recall, and F1-score are reported for each classifier. The average F1 score for all classifiers is considered as the overall performance metric.

Results



classifier	class	Accuracy metrics %		
		Precision	Recall	F-Measure
NB	pos	58.2	79.2	67.1
	neg	67.5	43.2	52.7
KNN-IBK (K=3)	pos	56.4	85.2	67.8
	neg	69.7	34	45.7
K*	pos	59.9	82.9	69.6
	neg	72.3	44.5	55.1
SVM	pos	**62.8**	**81.4**	**70.9**
	neg	**73.5**	**51.7**	**60.7**
DT-J48	pos	60.9	70.5	65.3
	neg	65	54.7	59.4

Discussion

Sentiment Analysis (SA) has gained considerable interest in recent years due to its commercial advantages in analyzing textual data. One of the main problems in SA is detecting fake positive and negative reviews from opinion reviews. The article focuses on the classification of movie reviews into positive or negative polarity using machine learning algorithms and SA methods. It compares the performance of five supervised machine learning algorithms (Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, Logistic Regression and Decision Tree) on three different movie review datasets. The results show that the SVM algorithm outperforms other algorithms in both text classification and fake review detection. The study also highlights the importance of removing ignored words to reduce memory requirements when classifying reviews. Overall, the paper demonstrates the effectiveness of machine learning techniques and SA methods in sentiment classification and fake review detection in movie reviews.

Conclusion

This research focused on analysing a dataset of movie reviews and proposed different methods for sentiment classification. The study used five supervised learning algorithms (NB, K-NN, LR, SVM, and DT) to classify sentiments in two different movie review datasets (V1.0, V2.0, and V3.0). Through rigorous experimental approaches, the accuracy, precision, recall, and F-measure of these algorithms were evaluated. The results indicate that the SVM algorithm demonstrated the highest accuracy in accurately classifying movie reviews. In addition, the research delved into the detection of false positive and false negative reviews and showed the effectiveness of specific processes in identifying such cases. Detection methods have proven to be key to ensuring the reliability of sentiment analysis. Looking ahead, future work is suggested to extend the study to include other data sets, such as those from Amazon or eBay, and to explore different feature selection methods. In addition, it is proposed to incorporate ignored word removal and inference methods along with the use of different tools such as Python or R Studio to further refine and evaluate the sentiment classification algorithms. This comprehensive approach aims to increase the robustness and applicability of research findings to different datasets and analytical methodologies.

References

1. Elmogy, A. M., Tariq, U., Ammar, M., & Ibrahim, A. (2021). Fake reviews detection using supervised machine learning. International Journal of Advanced Computer Science and Applications, 12(1).
2. Barbado, R., Araque , O., & Iglesias, C. A. (2019). A framework for fake review detection in online consumer electronics retailers. Information Processing & Management, 56(4), 1234-1244.
3. Jindal, N., & Liu, B. (2020, May). Review spam detection. In Proceedings of the 16th international conference on World Wide Web (pp. 1189-1195).
4. Gao, Y., Gong, M., Xie, Y., & Qin, A. K. (2020). An attention-based unsupervised adversarial model for movie review spam detection. IEEE transactions on multimedia, 23, 784-796.
5. Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). Fake review detection: Classification and analysis of real and pseudo reviews. UIC-CS-03-2013. Technical Report.
6. Paul, H., & Nikolaev, A. (2021). Fake review detection on online E-commerce platforms: a systematic literature review. Data Mining and Knowledge Discovery, 35(5), 1830-1881.
7. Mohawesh, R., Xu, S., Tran, S. N., Ollington , R., Springer, M., Jararweh , Y., & Maqsood, S. (2021). Fake reviews detection: A survey. IEEE Access, 9, 65771-65802.
8. Salminen, J., Kandpal , C., Kamel, A. M., Jung, S. G., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. Journal of Retailing and Consumer Services, 64, 102771.
9. Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. Journal of Big Data, 2(1), 1-24.
10. Ram, N. C. S., Vakati, G., Nadimpalli, J. V., Sah, Y., & Datla, S. K. (2022). Fake reviews detection using supervised machine learning. Int. J. Res. Appl. Sci. Eng. Technol.(IJRASET), 10, 3718-3727.
11. Chauhan, S. K., Goel, A., Goel, P., Chauhan, A., & Gurve, M. K. (2017, February). Research on product review analysis and spam review detection. In 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 390-393). IEEE
12. Kurkute, G. J. N. D. S., & Shinde, A. R. K. K. (2020). Detection of fake reviews using machine learning algorithm. International Journal of Future Generation Communication and Networking, 13(1), 415-419.
13. Priyanka, G. (2023). By contrasting dimensionality reduction with logistic regression, a novel grading-based genuine review prediction may be made online. Journal of Survey in Fisheries Sciences, 10(1S), 1508-1517
14. Mir, A. Q., Khan, F. Y., & Chishti, M. A. (2023). Online Fake Review Detection Using Supervised Machine Learning And Bert Model. arXiv preprint arXiv:2301.03225.
15. Choi, W., Nam, K., Park, M., Yang, S., Hwang, S., & Oh, H. (2023). Fake review identification and utility evaluation model using machine learning. Frontiers in artificial intelligence, 5, 1064371