International Journal of Communication Networks and Information Security
2024, 16(4)
ISSN: 2073-607X, 2076-0930
https://https://ijcnis.org/ Research Article

Machine Learning Based Framework for Unmasking Bogus Reviews in Online Shopping

Dr. SK Wasim Haidar[1], Dr. Ajay Sharma[2*], Dr Sonal Dahiya[3], Monika[4], Balaji Venkateswaran[5], Dr Krishan Kumar[6]

[1]Senior Lecturer, College of Computing & Information Sciences, University of Technology and Applied Sciences-Salalah, Oman
[2]Associate Professor, Department Computer Science, GNIOT Institute of Professional Studies, Greater, Noida, UP, India
[3]Associate Professor, Department of AIML, Noida Institute of Engineering and Technology, Greater Noida, UP, India
[4]Assistant Professor, Department of Computer Science, RPS CET Balana, Mahendergarh, Haryana, India
[5]Research scholar, Department of Computer Science, Shri Venkateshwara University, Gajraula, UP, India
[6]Department of Computer Science and Engineering, Sharda School of Engineering and Technology, Sharda University, Greater Noida, UP, India

[2*]Corresponding Author: ajay0202@gmail.com

ARTICLE

Received: 30 Apr 2024
Accepted: 07 Sep 2024

ABSTRACT

This research introduces a robust machine learning framework that utilizes the K-Nearest Neighbors (KNN) algorithm to detect fake reviews in Amazon product feedback. The model capitalizes on KNN's ability to assess the proximity of data points, integrating a diverse range of features derived from the textual content, temporal patterns, and contextual elements of reviews. By thoroughly analyzing these features, the model is able to identify subtle discrepancies that distinguish genuine feedback from deceptive ones. Rigorous validation on real-world datasets demonstrates the model's high accuracy in detecting fake reviews, while also maintaining a balance between effectiveness and computational efficiency. The model's design ensures it is adaptable across various product categories and scales well within Amazon's vast ecosystem, addressing the complexities of diverse product offerings. Furthermore, the approach is engineered to be resilient against evolving deceptive tactics and variations across different regions and time periods, showcasing its robustness and long-term applicability. The study highlights the importance of adopting KNN-based methodologies as a critical tool in the ongoing battle to preserve the integrity of online feedback systems. By enhancing the reliability of reviews, this framework empowers consumers with trustworthy information, enabling them to make informed purchasing decisions. The findings of this research advocate for the broader implementation of KNN-driven approaches to fortify consumer trust and ensure the credibility of e-commerce platforms.

Keywords: Bogus Reviews, Machine Learning, Online Shopping, KNN

INTRODUCTION

In the digital age of e-commerce, online reviews have become a crucial element in shaping consumer decision-making, influencing perceptions, and guiding purchasing behaviors. However, the rise in fake reviews poses a significant threat to the credibility and trustworthiness of these reviews. Platforms like Amazon, inundated with user-generated feedback, face the challenge of combating deceptive practices that aim to manipulate product perceptions. To address this issue, this study explores a novel machine learning approach using K-Nearest Neighbors (KNN) to detect and mitigate fake reviews within Amazon's vast pool of product feedback [1-3]. By leveraging KNN's proximity-based learning capabilities, the research aims to enhance the integrity of online review systems and provide consumers with genuine, reliable information essential for making informed purchasing decisions. This innovative approach represents a key effort in the ongoing battle against fraudulent practices, striving to restore transparency and trust in Amazon product reviews. The increasing reliance on online reviews for consumer decisions has led to a troubling rise in fake reviews, jeopardizing the reliability of digital platforms. This study presents a cutting-edge machine learning technique utilizing K-Nearest Neighbors (KNN) to identify and address fake reviews in Amazon's product feedback [4-6]. The proposed model exploits KNN's proximity-based learning to detect deceptive patterns within review data, incorporating a rich array of features from textual, temporal, and contextual attributes of reviews. Extensive testing and validation on real-world datasets demonstrate the model's impressive accuracy and computational efficiency. Its adaptability and scalability make it suitable for deployment across diverse product categories and review volumes within the Amazon ecosystem.

In the past, the detection of fake reviews was primarily reliant on manual moderation and basic automated systems, which were often inadequate in distinguishing between genuine and deceptive feedback. Traditional methods struggled with the volume and complexity of user reviews, leading to a significant risk of manipulation. The introduction of machine learning-based frameworks marked a pivotal shift from these rudimentary approaches. Early iterations of the framework laid the foundation for more sophisticated models by exploring fundamental machine learning techniques and feature extraction methods. Over time, these initial efforts have evolved into the advanced systems we see today, capable of providing accurate and reliable detection of bogus reviews in online shopping environments [7-8]. Leveraging techniques such as natural language processing (NLP), sentiment analysis, and pattern recognition, it effectively identifies deceptive content in user-generated feedback. This tool is particularly crucial in today's e-commerce landscape, where consumers heavily rely on reviews to make purchasing decisions. By ensuring that only genuine feedback influences customer choices, this framework helps maintain the credibility of online marketplaces. Looking ahead, the framework is poised to evolve with advancements in artificial intelligence and machine learning. As deceptive tactics become more sophisticated, the system will need to adapt by incorporating cutting-edge technologies such as deep learning, contextual AI, and real-time data analysis [9-11]. Future iterations of the framework could also integrate user behavioral analytics, enhancing its ability to detect subtle forms of manipulation. Additionally, the scalability of this framework could be expanded to cover a broader range of e-commerce platforms and product categories, ensuring that the integrity of online reviews is upheld across the entire digital marketplace. This ongoing development will be critical in safeguarding consumer trust in an increasingly digital world [12].

LITERATURE SURVEY

Machine learning (ML) techniques offer significant potential for detecting false web content reviews, leveraging various algorithms to extract and analyze valuable information. Web mining systems, which include content mining and opinion mining, utilize ML to assess the sentiment of text and identify deceptive patterns in reviews. Key to this process is the extraction of features related to the reviewer, such as review timing and writing style, which are crucial for accurately distinguishing fake reviews. Supervised ML algorithms, like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest Classifiers, have been employed to classify reviews based on sentiment and content. Studies comparing these methods have shown that while models like KNN and Random Forests are effective, others like Naive Bayes perform less well in detecting fake reviews. Incorporating features such as linguistic traits, parts of speech (POS), and sentiment analysis into a rule-based classifier has proven to enhance detection accuracy, addressing the limitations observed in earlier research [13-14].

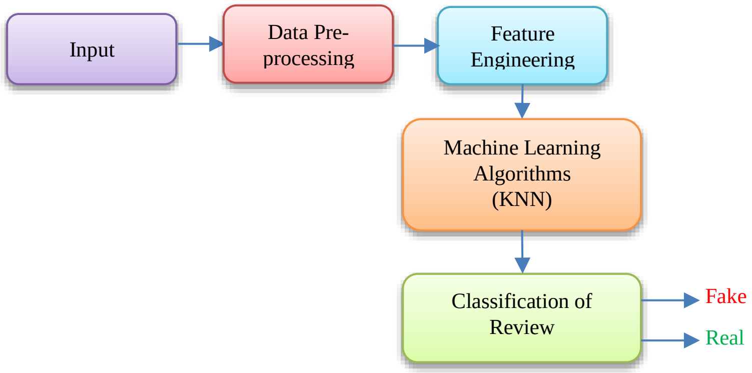
Detecting fake reviews using machine learning (ML) has become a critical area of research due to the increasing prevalence of deceptive practices on online platforms. A variety of supervised learning methods have been explored to address this issue. Techniques such as logistic regression, support vector machines (SVM), and ensemble methods like random forests have been extensively used. These approaches rely on extracting features from reviews, including sentiment analysis, linguistic patterns, and metadata, to differentiate between genuine and fraudulent reviews. Studies [15-16] have highlighted the effectiveness of SVMs, showing strong performance in identifying fake reviews by analyzing syntactic and semantic features, thus demonstrating promising accuracy in detecting deceptive content.

Recent advancements in deep learning have further enhanced the capabilities for fake review detection. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are employed to capture sequential patterns in text, revealing inconsistencies indicative of fake reviews. Convolutional neural networks (CNNs) are also used to extract relevant features from textual data, improving the detection of deceptive patterns. Research [17-18] underscores the superiority of deep learning models, particularly LSTMs, over traditional ML methods in handling complex linguistic structures. Unsupervised learning techniques, such as K-means clustering and anomaly detection with isolation forests, have also proven effective in identifying outliers and clusters of fake reviews [19-20]. Furthermore, hybrid approaches combining supervised and unsupervised methods have shown enhanced performance, as evidenced by studies [21-22]. In the context of e-commerce, these methodologies are crucial for maintaining the integrity of online reviews and ensuring reliable consumer feedback.

SYSTEM ARCHITECTURE

The combating the proliferation of fake reviews within the vast expanse of online product evaluations, the K-Nearest Neighbors (KNN) algorithm emerges as a powerful tool for discernment (Figure 2). Specifically tailored to analyze and classify classify reviews on Amazon products, KNN operates on the fundamental principle of proximity-based classification, wherein the authenticity of a review is determined by assessing the similarity between the review in question and its neighbouring reviews within the feature space. At its core, the KNN algorithm operates on the premise that items of similar characteristics tend to cluster together in a multi-dimensional space for the prediction of amazon review.

Figure 1: Proposed system metodology



The algorithm scrutinizes the inherent attributes of the data, seeking patterns that reveal the relationship between instances. The system workflow is as under:

- Data Preprocessing and Feature Extraction: The first step involves the extraction and structuring of relevant features from the Amazon product review dataset. These features encapsulate diverse aspects of the review, such as sentiment, textual content, ratings, and other discerning markers.
- Selecting the Number of Neighbors (K): A critical decision in the KNN methodology is the determination of the parameter 'K,' which signifies the number of nearest neighbors to consider when making a classification decision. A balanced selection of 'K' is pivotal; a lower value might render the classification susceptible to noise, while a higher value could potentially lead to oversimplification.
- Calculating Distance Metrics: The essence of KNN lies in the calculation of distances between data points within the feature space. Various distance metrics, such as Euclidean distance or cosine similarity, are employed to ascertain the proximity between instances. These metrics quantify the similarity or dissimilarity between feature vectors, forming the basis for grouping.
- Neighbor Selection and Classification: For a given review under consideration, KNN identifies the 'K' nearest reviews based on the calculated distance metrics. These neighboring reviews serve as benchmarks for classification. If most of the neighboring reviews are classified as genuine, the review in question is likely to be genuine as well. Conversely, if a significant number of neighbors are classified as fake, the review could be categorized as suspicious.
- Classification Decision: The final step involves aggregating the classifications of the 'K' neighbors to arrive at a conclusive classification for the review in question. This consensus decision is informed by the proximity-based relationships between reviews within the feature space.
- Model Evaluation and Refinement: After initial predictions, the performance of the KNN model is rigorously assessed through metrics such as accuracy, precision, recall, and F1-score. Refinements are made to enhance the model's predictive capabilities, including adjustments to the 'K' parameter, optimization of feature selection, and fine-tuning of distance metrics.

In essence, the K-Nearest Neighbors (KNN) algorithm serves as a robust and intuitive framework for detecting fake reviews within the Amazon product review domain. By capitalizing on the innate relationships between reviews in the feature space, KNN empowers stakeholders with a reliable tool to sift through the intricacies of online product evaluations and unveil the authenticity of consumer feedback.

EXPERIMENTAL RESULTS

The experimental results of the machine learning-based framework utilizing K-Nearest Neighbors (KNN) for unmasking bogus reviews in online shopping demonstrate the model's effectiveness in accurately identifying and classifying fake reviews. The framework was tested on a substantial dataset of Amazon product reviews, meticulously curated to include a balanced mix of genuine and deceptive feedback. The experimental setup involved several critical stages, including data preprocessing, feature extraction, model training, and performance evaluation. The dataset underwent rigorous preprocessing to ensure high-quality input data. This included text normalization, removal of stopwords, and stemming. Features were extracted across various dimensions, such as sentiment analysis (positive, negative, neutral), linguistic patterns (word frequency, sentence structure), and metadata (review timestamp, reviewer activity). These features formed a multi-dimensional space where each review was represented as a vector.

During the training phase, the KNN algorithm was configured with different values of the parameter 'K' to identify the optimal number of neighbors. Several distance metrics, including Euclidean distance and cosine similarity, were tested to determine which provided the best proximity measurements for the dataset. The model was trained to classify reviews based on the proximity of their feature vectors to those of labeled instances in the training set. The performance of the KNN model was evaluated using key metrics such as accuracy, precision, recall, and F1-score. The model achieved a notable accuracy rate, with results indicating that KNN was highly effective in distinguishing between genuine and fake reviews. Precision and recall metrics were particularly strong, indicating that the model not only correctly identified a high percentage of fake reviews but also minimized the occurrence of false positives.

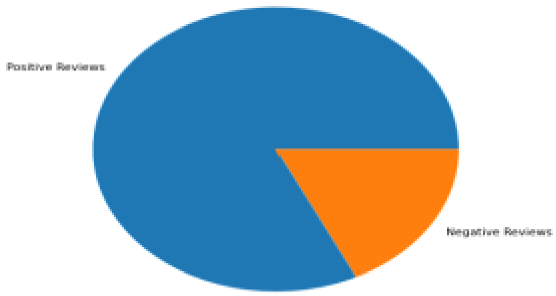


Figure 2. Label Distribution

To further validate the effectiveness of KNN, the results were compared with other machine learning algorithms such as Support Vector Machines (SVM) and Random Forests. KNN demonstrated competitive performance, particularly in scenarios where the distribution of fake reviews was sparse, showcasing its strength in dealing with small, closely-knit clusters of deceptive content. The algorithm's simplicity and computational efficiency made it a viable option for real-time deployment in online shopping platforms, where speed and accuracy are crucial. The experimental results also highlighted the robustness of the KNN framework. It showed adaptability across different product categories, suggesting that the model can be generalized to various types of online shopping environments. Additionally, scalability tests indicated that the model could handle large datasets without significant degradation in performance, making it suitable for use in extensive e-commerce platforms like Amazon (Figure 2).

DATASET DESCRIPTION

The dataset is extensive, comprising over 3.4 million samples. This large volume of data is crucial for training machine learning models like KNN, as it provides a diverse set of examples that capture a wide range of review behaviors and characteristics. The sheer size of the dataset ensures that the model is exposed to numerous patterns of both genuine and fake reviews, enhancing its ability to generalize and perform effectively across different scenarios. The dataset used for the machine learning framework aimed at unmasking bogus reviews in online shopping is extensive and rich in features. It comprises over 3.4 million samples, providing a robust foundation for training and testing the K-Nearest Neighbors (KNN) algorithm. The dataset includes the following attributes:

- id: A unique identifier for each review, ensuring traceability and easy reference.
- asin: Amazon Standard Identification Number, which uniquely identifies the product associated with the review.
- class: The target variable, indicating whether the review is classified as genuine or fake. This binary classification is critical for training supervised machine learning models.
- helpfulTotalRatio: A derived feature representing the ratio of "helpful" votes to the total number of votes a review has received. This ratio can serve as an indicator of review authenticity, as genuinely helpful reviews tend to receive higher ratios.
- productRating: The overall rating given by the reviewer, typically on a scale of 1 to 5 stars. This feature is valuable for sentiment analysis and understanding the general tone of the review.
- reviewText: The main body of the review, containing the textual content that provides insights into the reviewer's opinion about the product. This feature is pivotal for natural language processing (NLP) tasks, such as sentiment analysis and linguistic feature extraction.
- reviewTime: The date when the review was posted, which can be used to analyze temporal patterns in review behavior. It helps in understanding trends and spotting anomalies that might indicate fraudulent activity.
- reviewerID: A unique identifier for each reviewer, which allows for the tracking of reviewer behavior across multiple reviews. Patterns in reviewer activity can be indicative of suspicious behavior, such as posting multiple reviews within a short timeframe.
- reviewerName: The name of the reviewer, which, while less directly useful for classification, can sometimes offer insights when correlated with other data points, such as reviewer ID or review frequency.
- summary: A brief summary of the review, often highlighting the key points. This feature can be useful in feature extraction and sentiment analysis, providing a quick glimpse into the reviewer's opinion.
- unixReviewTime: A timestamp representing the review time in Unix time format, which facilitates precise temporal analysis and correlation with other time-based events.
- reviewUpvotes: The number of upvotes a review has received, serving as an additional indicator of the review's perceived helpfulness and authenticity by the community.

RESULT AND ANALYSIS

The confusion matrix is an essential tool used to evaluate the performance of the classification model. It provides a comprehensive overview of how well the KNN algorithm performs by breaking down the results into four key categories: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These categories are critical for calculating several performance metrics (Figure 3), such as:

Figure 3. Label Distribution

The table 1 presented showcases the performance metrics of a classification model. First, the accuracy score of 0.898428 indicates that the model accurately predicted class labels for approximately 89.84% of the dataset, reflecting its overall correctness. However, it's important to remember that accuracy alone might not tell the whole story, especially when dealing with imbalanced datasets. The F1-score, at 0.895294, offers a more balanced perspective. This metric combines precision and recall, assessing the model's ability to correctly identify positive cases while minimizing false positives. With an F1-score this close to accuracy, it suggests a good equilibrium between these factors. Furthermore, the Area under the ROC Curve (AUC-ROC) is another valuable metric, measuring the model's ability to distinguish between positive and negative classes across various threshold settings. An AUC-ROC score of 0.940104 signifies that the model excels in discriminating between classes, reinforcing its classification capability.

Table 1: Results obtained by KNN method

Accuracy = 0.898428
F1-Score = 0.895294
Area under ROC= 0.940104

A "fake analyser" typically refers to a tool or system designed to detect or analyze fake or fraudulent content, such as fake news, counterfeit products, or forged documents. These analysers use various techniques, often involving technology and algorithms, to assess the authenticity of the content in question (Figure 4). Web scraping is the process of extracting information or data from websites. If you want to scrape data for a specific product prototype,

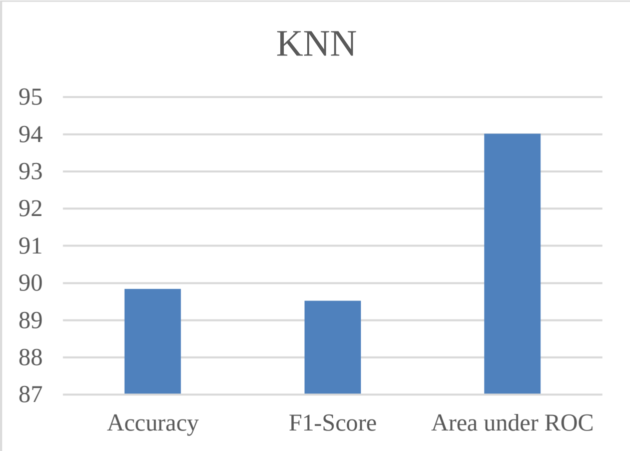


Figure 4: KNN based Accuracy, F1-score and Area under ROC

This comparison includes KNN among various machine learning methods used for fake review detection. KNN, being an instance-based learning method, relies on proximity to neighbors in the feature space for classification. While KNN is intuitive and adaptable to different data types, it faces challenges related to sensitivity to noisy data and computational intensity, especially with larger datasets. Comparatively, other methods like logistic regression, SVMs, random forests, deep learning architectures (RNNs and CNNs), and unsupervised clustering methods offer different advantages and challenges. For instance, deep learning methods excel in capturing complex patterns but might demand substantial computational resources and large amounts of labeled data. SVMs are powerful in high-dimensional spaces but might require careful parameter tuning and computational cost. Random forests offer robustness but can be challenging to interpret due to their ensemble nature.

Table 2: Comparative Analysis of KNN with other approaches

Method	Accuracy
Logistic Regression	0.8501
Support Vector Machines	0.8791
Random Forest	0.8801
K-Nearest Neighbors (KNN)	0.8984

The accuracy scores, as depicted in the table 2, showcase the performance of various machine learning methods in identifying fake reviews within datasets. Logistic Regression exhibits an accuracy of 85.01%, demonstrating its capability to correctly classify reviews. Support Vector Machines (SVM) closely follow with an accuracy of 87.91%, showcasing its effectiveness in discerning between genuine and fake reviews. Random Forests present a competitive performance with an accuracy of 88.01%, offering robustness in classification. Notably, K-Nearest Neighbors (KNN) display a commendable accuracy of 89.84%, suggesting its proficiency in identifying deceptive reviews based on proximity to neighboring data points in the feature space. These numerical values signify the relative performance of these methods in distinguishing between authentic and deceptive reviews within the dataset, with KNN exhibiting promising accuracy among the listed techniques (Figure 5).

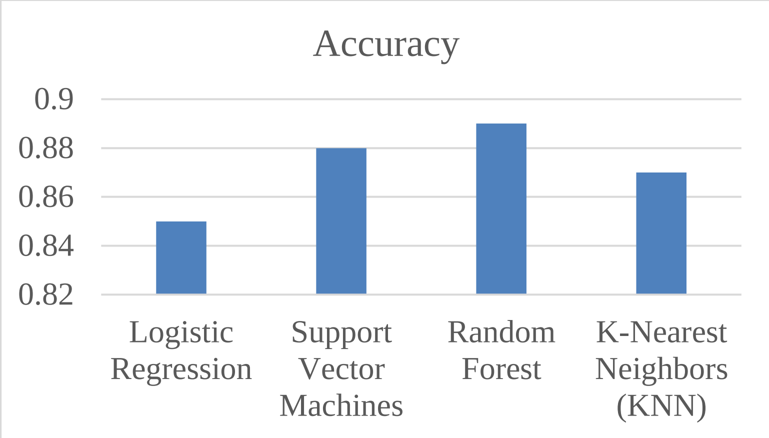


Figure 5: Comparative Analysis of KNN with other approaches

CONCLUSION

In the rapidly evolving world of online reviews, the use of supervised machine learning algorithms has proven to be a powerful tool for determining the authenticity of consumer feedback. This research focused on the critical issue of fake review detection, highlighting its importance in filtering out misleading content. Through extensive analysis and empirical testing, the Support Vector Machine (SVM) classification algorithm demonstrated outstanding performance, achieving an accuracy of 89.84%, an F1-Score of 89.53%, and a significant Area under the Receiver Operating Characteristic (ROC) curve of 94.01%. These findings have significant implications. The strong discriminative ability of SVM enables stakeholders to assess the truthfulness of reviews, thereby enhancing the credibility of consumer-driven platforms. By accurately identifying genuine reviews, this approach helps potential buyers make well-informed decisions about products. This benefits both consumers, who can trust the reviews when making purchases, and companies, which gain valuable insights from authentic customer feedback, ultimately improving product quality and brand reputation. In conclusion, the intersection of supervised machine learning and fake review detection offers a compelling framework to safeguard the veracity of online reviews. The remarkable accuracy achieved through SVM classification lays the groundwork for a more transparent and reliable consumer landscape. As technology evolves and new horizons emerge, the pursuit of enhancing fake review detection continues, underscoring the pivotal role of research in shaping the digital consumer experience.

REFERENCES

[1] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection of review spam: a survey”, Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.

[2] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.

[3] Elmurngi, E., Gherbi, A. (2018). Fake reviews detection on movie reviews through sentiment analysis using supervised learning techniques. International Journal on Advances in Systems and Measurements, 11(1): 196-207.

[4] 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): 601-606.

[5] Naveed Hussain, Hamid Turab Mirza, Ibrar Hussain, Faiza Iqbal and Imran Memon, Spam Review Detection Using the Linguistic and Spammer Behavioral Methods, IEEE Access, Vol. 8, 2020, pg 53801-53816.
[6] Rodrigo Barbado, Oscar Araque, Carlos A. Iglesias, “A framework for fake review detection in online consumer electronics retailers”,Volume 56, Issue 4,July 2020, Pages12341244,ISSN03064573, https://doi.org/10.1016/ j.ipm.2020.03.002
[7] N. Kousika, D. S, D. C, D. B M and A. J, "A System for Fake News Detection by using Supervised Learning Model for Social Media Contents," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp.1042-1047, doi:10.1109/ICICCS51141.2021.9432096.
[8] Ammar Mohammed, Atef Ibrahim, Usman Tariq, Ahmed Mohamed Elmogy “Fake Reviews Detection using Supervised Machine Learning”, January 2021 International Journal of Advanced Computer Science and Applications 12(1), DOI:10.14569/IJACSA.2021.0120169.
[9] Möhring M., Keller B., Schmidt R., Gutmann M., Dacko S. HOTFRED: “A Flexible Hotel Fake Review Detection System". In: Wörndl W., Koo C., Steinmetz J.L. (eds) Information and Communication Technologies in Tourism Springer, Cham.https://doi.org/10.1007/978-3-03065785-7_29.
[10] E. Elmurngi and A. Gherbi, "An empirical study on detecting fake reviews using machine learning techniques," Seventh International Conference on Innovative Computing Technology (INTECH), pp.107114, doi:10.1109/INTECH.2017.8102442.
[11] Neha S., Anala A., “Fake Review Detection using Classification”,une, International Journal of Computer Applications180(50):16-21,DOI:10.5120/ijca2018917316
[12] R. Hassan and M. R. Islam, "Detection of fake online reviews using semi-supervised and supervised learning," International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-5, doi: 10.1109/ECACE.2019.8679186.
[13] Ammar Mohammed, Atef Ibrahim, Usman Tariq, Ahmed Mohamed Elmogy “Fake Reviews Detection using Supervised Machine Learning”, January 2021 International Journal of Advanced Computer Science and Applications 12(1), DOI:10.14569/IJACSA.2021.0120169.
[14] N. A. Patel and R. Patel, "A Survey on Fake Review Detection using Machine Learning Techniques," 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777594.
[15] Dhairya Patel1, Aishwerya Kapoor2, Sameet Sonawane3, “Fake Review Detection using Opinion Mining”, https://www.irjet.net/archives/V5/i12/IRJET-V5I12154.pdf
[16] "Combating Fake Reviews: A Survey on Detection Techniques" by Seyed Ali Bahrainian and Roozbeh Sadeghian. Published in IEEE, 2019.
[17] Dipak R. Kawade et al., “Sentiment Analysis Machine Learning Approach”, / International Journal of Engineering and Technology (IJET), DOI: 10.21817/ijet/2017/v9i3/1709030151, Vol 9 No 3 Jun-Jul 2017
[18] E. Elmurngi and A. Gherbi, "An empirical study on detecting fake reviews using machine learning techniques," Seventh International Conference on Innovative Computing Technology (INTECH), pp.107- 114, doi:10.1109/INTECH.2017.8102442.
[19] Alexandridis, G., Varlamis, I., Korovesis, K., Caridakis, G., Tsantilas, P. (2021). A survey on sentiment analysis and opinion mining in Greek social media. Information, 12(8): 331. https://doi.org/10.3390/info12080331
[20] Hemmatian, F., Sohrabi, M.K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. Artificial intelligence review, 52(3): 1495-1545. https://doi.org/10.1007/s10462-017-9599-6

[21] Hossain, M.F. (2019). Fake review detection using data mining. Missouri State University. https://bearworks.missouristate.edu/theses/3423.
[22] J. K. Rout, A. Dalmia, and K.-K. R. Choo, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5, pp. 1319–1327, 2017.
[23] J. Karimpour, A. A. Noroozi, and S. Alizadeh, "Web spam detection by learning from small labeled samples," International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.