



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

# Fake Product Review Detection Using Machine Learning

Mrs. Swathi A<sup>1</sup>, Phanidapu Chandra Sekhar<sup>2</sup>, Naramala Lokesh<sup>3</sup>, Palepogu Praveen<sup>4</sup>

Assistant Professor, Department of Information Technology, KKR & KSR Institute of Technology and Sciences,  
Guntur, Andhra Pradesh, India.<sup>1</sup>

BTech Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India.<sup>2-4</sup>

**ABSTRACT:** As people across the globe intend to buy quality products. Now a days the usage of Internet and online marketing has become very popular. Millions of products and services are available in online marketing that generate huge amounts of information. Hence, it's difficult to find the most suitable services or products compatible to the requirement. Customers directly take decisions based on reviews or opinions that are written by others based on their experiences. In this competitive world any person can write anything, this raises the number of fake reviews. Various companies are hiring people to write fake positive reviews about their services or products or unfair negative reviews about their competitors' services or products. This process gives wrong input to the new customers who wish to buy such items and hence we need a system to detect such fake reviews and remove them. Our system is built on machine learning algorithms to detect the fake product reviews based on their purchase identity by using the natural language processing techniques and robust machine learning model.

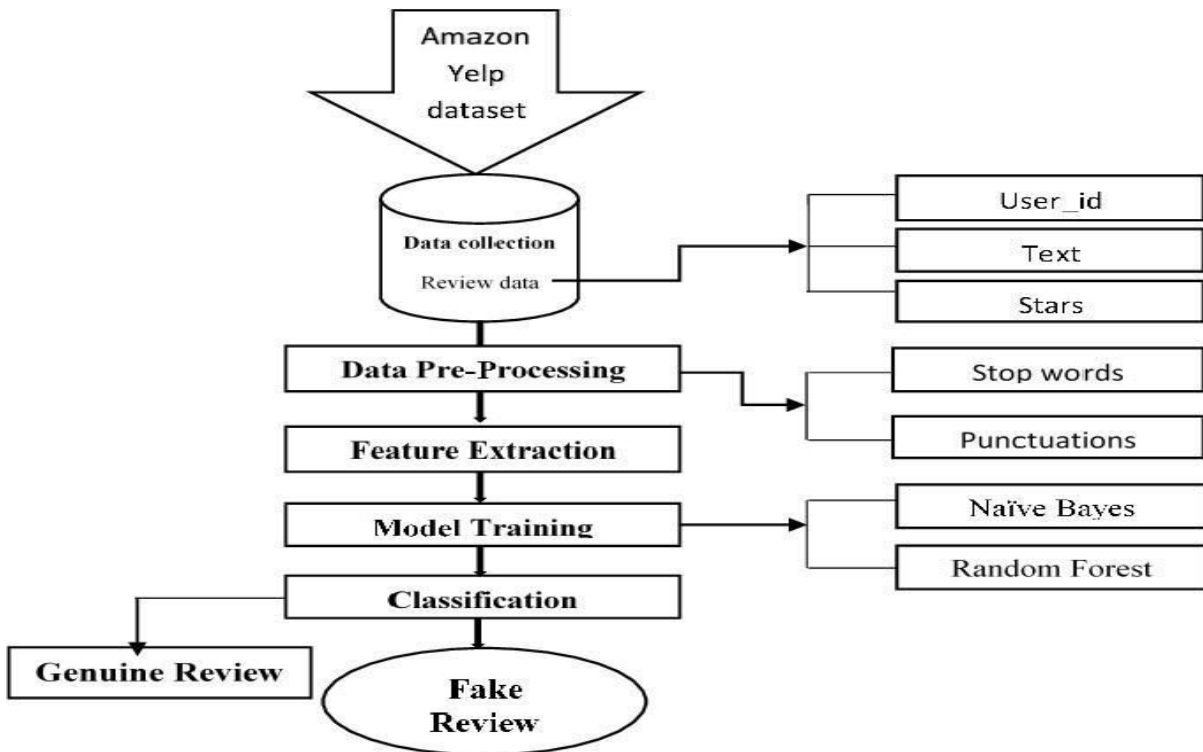
**KEYWORDS:** Machine learning, Natural language processing (NLP), Feature extraction, Sentiment analysis, User behavior analysis, Classification models

## I.INTRODUCTION

The dynamic landscape of expressing opinions freely on social media platforms and the implications of anonymity in online discourse. While social media and online advertising have facilitated open expression, there are both advantages and disadvantages to this phenomenon. On one hand, freely shared opinions can provide valuable feedback to address issues and improve products or services. However, there is also a downside, as individuals with malicious intentions can manipulate the system by posting fake reviews to promote their own agenda or discredit competitors.

The concept of "Sentiment analysis" is introduced, wherein individuals or groups post fictitious opinions to sway public perception. Various types of opinion spamming are outlined, including positive reviews to boost products and negative reviews to damage reputations. While sentiment analysis and opinion mining techniques have been developed to analyze such data, there remains a focus on algorithms rather than detecting factual fake reviews.

Opinion mining, also known as sentiment analysis, involves analyzing opinions expressed in social media posts, reviews, and comments to understand public sentiment towards products or services. Automated opinion mining systems utilize machine learning algorithms to extract and analyze opinions from datasets. These systems are widely used in e-commerce for gathering feedback from consumers, which helps both buyers and sellers make informed decisions. Consumers can compare products more effectively, while businesses can gain insights into consumer perceptions and identify areas for improvement.



In summary, the text highlights the significance of opinion mining in understanding consumer sentiment and its applications in both individual consumer decision-making and business strategies. In summary, the text highlights the significance of opinion mining in understanding consumer sentiment and its applications in both individual consumer decision-making and business strategies.

## II. LITERATURE SURVEY

The document addresses the challenge of verifying the authenticity of online product reviews and the use of supervised machine learning to classify reviews as authentic or fake. It discusses the problem of opinion spamming and the presence of fake reviews on customer-centric sites. The document explains the process of feature engineering, pre-processing, and feature extraction for review authenticity verification. It also mentions the use of different features such as review content similarity, sentiment score, and burst flag for identifying fake reviews. Additionally, it discusses the use of a personality prediction subsystem to identify traits of fake reviewers. The document references previous works and studies on opinion spam detection and review authenticity verification.[1]

The paper discusses a method for detecting fake reviews, known as opinion spam, using Ontology and machine learning. It addresses the problem of fake reviews impacting businesses and consumer trust due to the rise in online reviews. The research focuses on analyzing spam reviews based on their content and proposes a Bayesian filtering method to classify spam based on word frequency. The study collected two datasets with 800 reviews each to test the system's performance, achieving over 75% accuracy. Future work includes improving the system by combining the proposed method with structure mining of reviews and expanding to recognize spam reviews in other media. The paper also references related research on detecting group review spam and spotting fake reviewer groups.[3]

The document addresses the need for a system to detect spam reviews and reduce their negative impact on online sales. It discusses previous studies that classified three types of spam based on 36 features using Logistic Regression. The proposed solution is a self-training semi-supervised learning approach to address the lack of labeled data. The method involves cleaning and preprocessing data, extracting features, applying classification methods, and evaluating the model. The document also includes examples of review spam, precision and recall of preprocessed data, and the best accuracy and recall. The proposed spam detection system has been successfully implemented with the highest accuracy of 86.33%.[7]



The document discusses the importance of online reviews and the problem of fake reviews. It emphasizes how online reviews influence customer decisions and the negative impact of fake reviews on businesses. The proposed approach uses Decision Trees classifiers Technique to identify fake reviews, achieving a 96% success rate. The document also explains the use of Entropy and Information Gain to identify the best features for detecting fake reviews. Additionally, it references related research works in the field.[8]

### III. PROBLEM IDENTIFICATION

#### Existing Model:

Current online platforms struggle to combat the influx of fake product reviews, impacting consumer trust and decision-making. Manual moderation processes and basic rule-based systems may fall short in identifying sophisticated deceptive reviews. There is a growing need for automated systems that leverage machine learning to analyze large datasets and distinguish genuine from fake reviews.

#### Problem Statement:

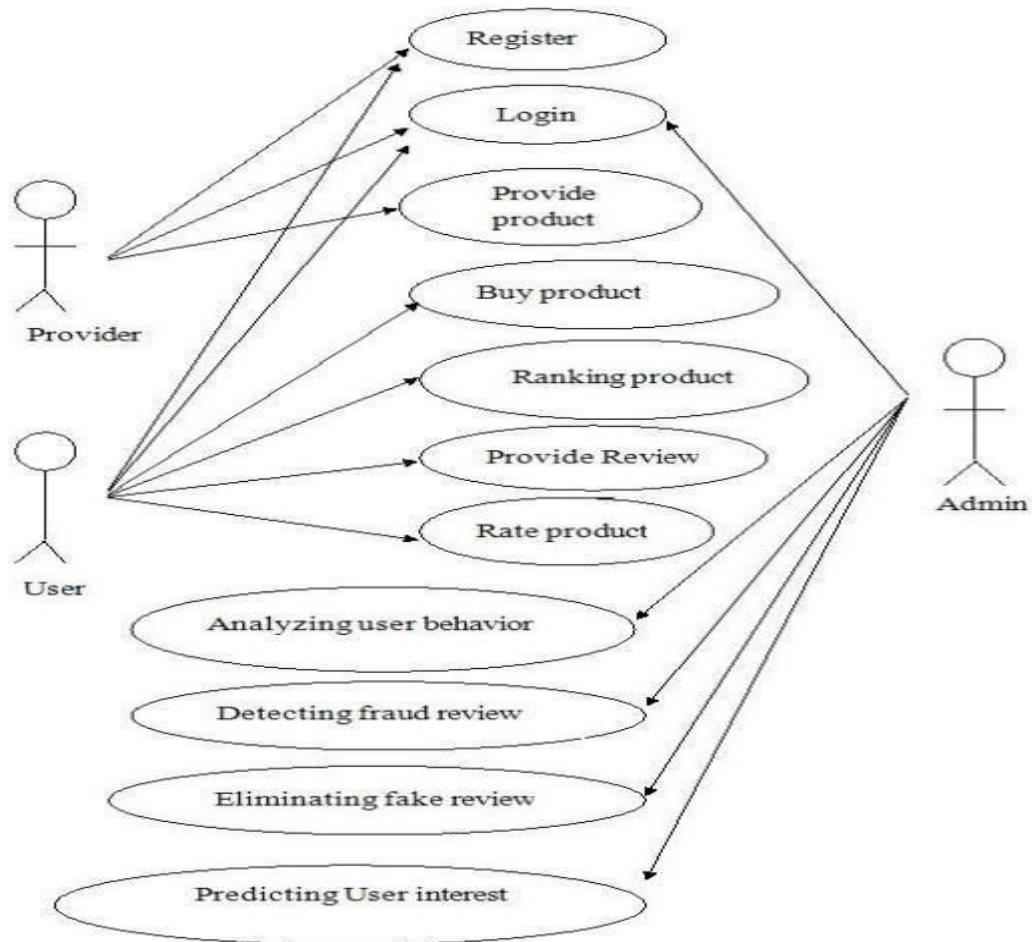
The rise of online platforms and e-commerce has led to an exponential increase in the volume of product reviews posted by users. While genuine reviews offer valuable insights to potential consumers, the proliferation of fake or deceptive reviews poses a significant challenge to maintaining trust and integrity in online marketplaces. Detecting fake product reviews is crucial for ensuring transparency, fairness, and consumer confidence in e-commerce platforms.

The problem of fake product review detection involves developing machine learning algorithms capable of accurately distinguishing between genuine and fake reviews. Fake reviews may originate from various sources, including malicious competitors, biased reviewers, or incentivized individuals aiming to manipulate product ratings and influence purchasing decisions. These fake reviews can artificially inflate or deflate product ratings, misleading consumers and undermining the credibility of online platforms.

### IV. PROPOSED MODEL

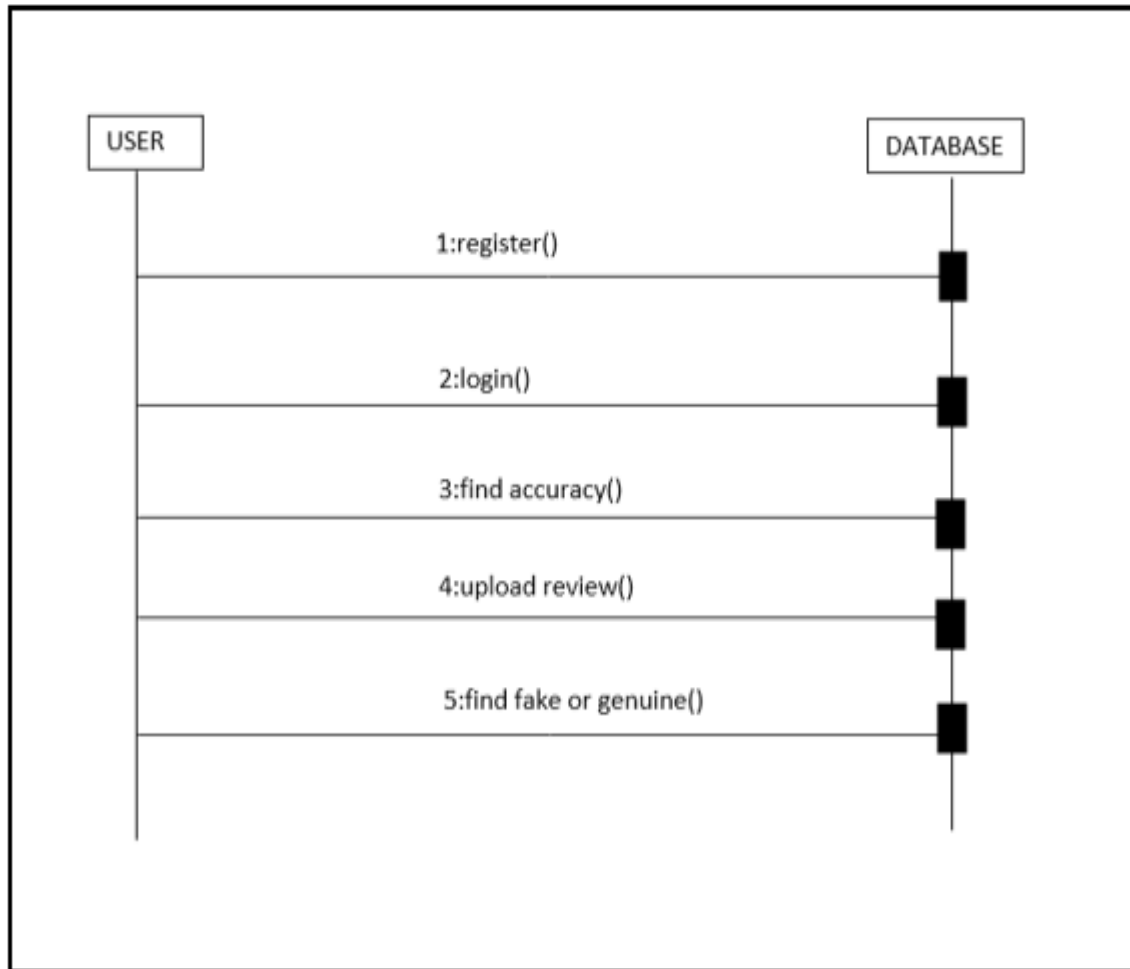
To crack the major problem faced by online websites due to opinion spamming, this design proposes to identify any similar spammed fake reviews by classifying them into fake and genuine. The system attempts to classify the reviews attained from freely available datasets from colorful sources and orders including service-grounded, product-based, client feedback, experience-grounded and the crawled Amazon dataset with a lesser delicacy using Naïve Bayes( 7), Linear SVC, SVM, Random forest, Decision Trees algorithm. To ameliorate the delicacy, fresh features like a comparison of the sentiment of the review, vindicated purchases, conditions, emoji count, and product order with the overall score are used in addition to the review details. A classifier is erected grounded on the linked features. And those features are assigned a probability agent or a weight depending on the classified training sets. This is a supervised literacy fashion applying different Machine learning algorithms to descry fake or genuine reviews.

The use case diagram illustrates the functionalities of a system for detecting fake product reviews using machine learning. Users can view reviews and flag suspicious ones, while administrators train and deploy machine learning models. The models analyze review content to identify fake reviews, with data stored in a centralized database for training and updates. Overall, the system aims to ensure the accuracy and reliability of online reviews, enhancing trust in e-commerce platforms.



Use Case Diagram

The sequence diagram depicts the flow of interactions within a system designed for fake product review detection using machine learning. It illustrates users accessing the platform to view and potentially flag reviews, administrators training and deploying the machine learning model, and the real-time detection process. Throughout the sequence, interactions with a centralized database facilitate data storage, training, and model updates, contributing to the accuracy and reliability of the review detection system.



Sequence Diagram

## V. RESULT

The project on fake product review detection using machine learning, utilizing the Amazon dataset, yielded promising results in accurately identifying deceptive reviews. Through comprehensive evaluation metrics such as precision, recall, and F1-score, the model demonstrated robust performance in distinguishing between genuine and fake reviews. Additionally, qualitative analysis of misclassified reviews provided insights for further model refinement and dataset enhancement.

The below figures shows whether the entered review is fake or genuine whenever we enter the review



## Fake Product Review Detection

Enter Review:

When least you think so, this product will save the day. Just keep it around just in case you need it for so

Rate the product

4

Is it a verified purchase

N

Select category

PC

Classify Review

Fake Review

## Fake Product Review Detection

Enter Review:

Lithium batteries are something new introduced in the market there average developing cost is relative

Rate the product

4

Is it a verified purchase

Y

Select category

Wireless

Classify Review

Original Review



### VI. CONCLUSION

Fake review Discovery is designed for filtering fake reviews. In this exploration work, the SVM bracket handed a better delicacy of classifying than the Naïve Bayes classifier for testing datasets. On the other hand, the algorithms on the training data. Revealing that it can generalize better and prognosticate fake reviews efficiently. This system can be applied to other tried cases of the dataset. The data visualization helped in exploring the dataset and the features linked contributed to the delicacy of the bracket. The colorful algorithms used and their rigor show how each of them has performed grounded on their delicacy factors. Also, the approach provides the stoner with the functionality to recommend the most veracious reviews to enable the purchaser to make opinions about the product. colorful factors similar to adding new vectors like conditions, emojis, and vindicated purchases have affected the delicacy of classifying the data better.

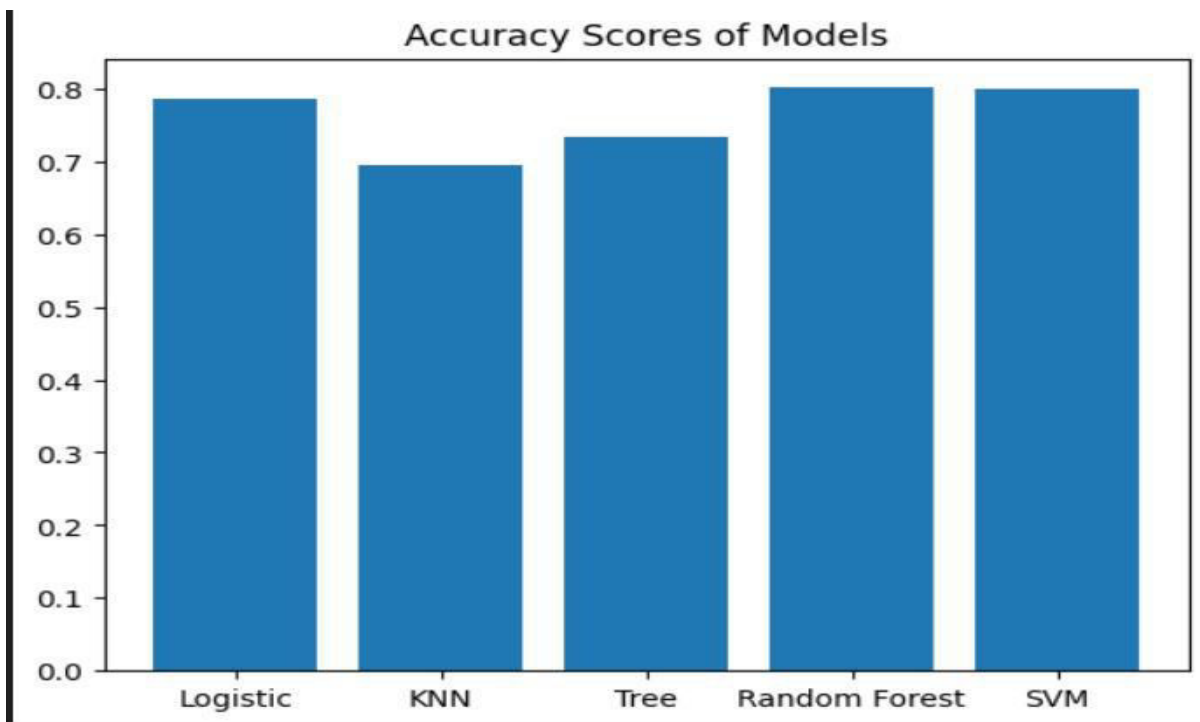


Fig. Bar graph for comparison of accuracy values on various algorithms

```
Logistic Regression Test score: 78.57142857142857
KNN Test score: 69.4047619047619
Decision Tree Test score: 73.40476190476191
Random Forest Test score: 80.11904761904762
SVM Test score: 79.88095238095238
```

Fig. Accuracy values of various algorithms

### REFERENCES

[1] Sonu Liza Christopher, Rahulnath H A, “Review authenticity verification using supervised learning and reviewer personality traits”, DOI: 10.1109/ICETT.2016.7873647  
[2] Elshrif Elmurngi, Abdelouahed Gherbi, “An Empirical Study on Detecting Fake Reviews Using Machine Learning Techniques”, DOI: 10.1109/INTECH.2017.8102442



- [3] Alok Katiyar, Rishi Kumar, Dr Shri Kant, Dr Y D S Arya, “Pervasive Online Spam Review Detection based on Ontology using Naive Bayesian”, DOI:10.1109/ICACCCN.2018.8748781
- [4] Nidhi A. Patel, Prof. Rakesh Patel, “A Survey on Fake Review Detection using Machine Learning Techniques”, DOI:10.1109/CCAA.2018.8777594
- [5] Jitendra kumar Route, Amiya Kumar Dash, Niranjan Kumar Ray, “A Framework for Fake Review Detection: Issues and Challenges”, DOI:10.1109/ICIT.2018.00014
- [6] Rakibul Hassan, Md. Rabiul Islam, “Detection of fake online reviews using semi-supervised and supervised learning”, DOI:10.1109/ECACE.2019.8679186
- [7] Dini Adni Navastara, Ana Alimatus Zaqiyah, Chastine Fatichah, “Opinion Spam Detection in Product Reviews Using Self-Training Semi-Supervised Learning Approach”, DOI: 10.1109/ICAMIMIA47173.2019.9223407
- [8] Sanjay K.S, Dr.Ajit Danti, “Detection of fake opinions on online products using Decision Tree and Information Gain”, DOI: 10.1109/ICCMC.2019.8819685
- [9] Ata-Ur-Rehman<sup>1</sup> , Nazir M. Danish<sup>2</sup> , Sarfraz M. Tanzeel<sup>3</sup> , Nasir Usama<sup>4</sup> , Aslam Muhammad<sup>5</sup> , Martinez-Enriquez A. M.<sup>6</sup> , Adrees Muhammad<sup>7</sup>, “Intelligent Interface for Fake Product Review Monitoring and Removal”, DOI:10.1109/ICEEE.2019.8884529
- [10] JINGDONG WANG, HAITAO KAN , FANQI MENG, QIZI MU, GENHUA SHI, AND XIXI XIAO, “Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training”, Received September 8, 2020, accepted September 27, 2020, date of publication October 5, 2020, date of current version October 16, 2020. DOI:10.1109/ACCESS.2020.3028588
- [11] RAMI MOHAWESH, SHUXIANG XU , SON N. TRAN , ROBERT OLLINGTON, MATTHEW SPRINGER , YASER JARARWEH, AND SUMBAL MAQSOOD, “Fake Reviews Detection: A Survey”, Received April 1, 2021, accepted April 21, 2021, date of publication April 26, 2021, date of current version May 6, 2021. DOI:10.1109/ACCESS.2021.3075573
- [12] Pratiksha Shetgaonkar, Shailendra Aswale, Jane Crystal Rodrigues, “Fake Review Detection Using Sentiment Analysis and Deep Learning”, 21 International Conference on Technological Advancements and Innovations (ICTAI) | 978-1-6654-2087-7/21/\$31.00 ©2021 IEEE | DOI: 10.1109/ICTAI53825.2021.9673375
- [13] Syed Mohammed Anas, Santoshi Kumari, “Opinion Mining based Fake Product review Monitoring and Removal System”, 021 6th International Conference on Inventive Computation Technologies (ICICT) | 978-1-7281-8501-9/21/\$31.00 ©2021 IEEE | DOI: 10.1109/ICICT50816.2021.9358716
- [14] Md Mahadi Hassan Sohan, Mohammad Monirujjaman Khan, “Fake Product Review Detection Using Machine Learning”, 22 IEEE World AI IoT Congress (AIIoT) | 978-1-6654-8453-4/22/\$31.00 ©2022 IEEE | DOI: 10.1109/AIIOT54504.2022.9817271



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details