iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51925

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Online Fake Review Detection Using SVM

Avantika Tellawar[1], Samiksha Wattamwar[2], Bhagyashri Thorat[3]
Electronics And Telecommunication hope's Foundations International Institute Of Information Technology, Pune, India

Abstract: ***After the pandemic, our overall life is changing and challenging that's why our demand is changing now, we are focused on wellness, sustainability, technology, and the gig economy because of these trends we can observe the reflection in the changing desired need and limitation of seller and customer. Many customers can post a review on any website after making a purchase. Whether it's an online purchase or an offline retail purchase. When customers buy a product online, they check the product reviews. This is very important for today's e-commerce product decisions. There is a financial gain associated with writing fake reviews, which is why there has been a significant increase in misleading statements about certain product reviews on websites. Misleading reviews are dangerous reviews. Positive product reviews can attract customers and increase sales. Negative product reviews can reduce demand for that product and reduce sales. These misleading reviews are dangerous to your product's reputation. In this paper, we use machine learning algorithms which are SVM Support vector machines, which are one of the most popular supervised learning algorithms used for both classification and regression problems. However, it is mainly used for machine-learning classification problems.***

Keywords: SVM, Machine Learning, Classification, reviews, real, fake

I. INTRODUCTION

In today's world, more people are buying a product online and Online reviews and comments after product sales have become very important for making buying and selling decisions. Fake reviews will affect such decisions due to deceptive information, leading to financial losses for consumers and sellers. Identification of fake reviews has thus received a great deal of attention in recent years. However, most websites have only focused on dealing with problematic reviews and comments. Amazon would only remove possible fake reviews without questioning the sellers who could continue posting deceptive reviews for business purposes. In the proposed system identify whether the review is fake or truthful by using the supervised machine learning classification algorithm SVM(support vector machine). The SVM is one of the popular algorithms for classification this algorithm helps to solve the classification problem. [1] Detection of fake reviews and truthful review is done by various algorithms but it has to vary low accuracy. Random Forest gives around 50 percent accuracy followed by linear Regression gives 60 percent accuracy. Using SVM we will try to increase accuracy to nearly 80 percent efficiency[2].

A. SVM

SVM is a machine learning algorithm which use for both regression and classification both but it is mostly used for classification The purpose of the SVM algorithm is to create optimal lines or decision boundaries that can divide the n-dimensional space into classes so that new data points can be easily placed in the correct category in the future. This optimal decision boundary is called a hyperplane. SVM selects extrema or vectors to help create hyperplanes. These extreme cases are called support vectors, and the algorithm is called a support vector machine. Consider the following figure, which has two different categories classified by decision boundaries or hyperplanes.[3]

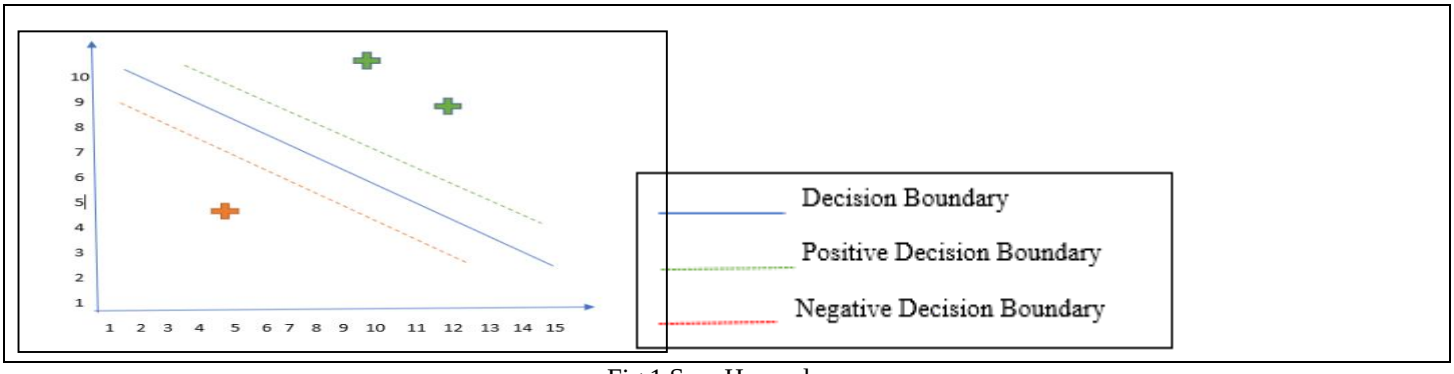


Fig.1.Svm Hyperplane

©IJRASET: All Rights are Reserved

The many characteristics or peculiarities of the text reviews might be used to help with the classification issue. E.g., Length of reviews (false reviews typically have shorter lengths and offer fewer product-specific details) and repetitious wording (fake reviews) reviews have a limited vocabulary that frequently repeats words. There are additional characteristics than the review text alone that may help to identify fraudulent reviews. Ratings confirmed purchases, and product categories are some of the important ones that were employed as additional features inclusion.[4]

An efficient approach for detecting fake reviews is designed and evaluated. It analyses aspects rather than the entire detailed review text. Spammers replicate aspects and change their sentiments in their fake reviews. As a result, aspect sentiments are computed using POS tagging and Sent WordNet. As a result, such extracted aspects are fed into a CNN/LSTM hybrid model for aspect replication and fake review detection. This method of using important aspects of reviews to training CNN and LSTM hybrid models saves computational time and provides better accuracy than peer-reviewed approaches.[5]

The remaining part of this paper is organized as follows first is an introduction in an introduction brief intro about the project such as why this system is required and the usages of the system. the second chapter include a Literature Survey. the third chapter is the Proposed Methodology in these various sub-topics are included such as Requirement analysis and impact in impact positive impact and negative impact in the fourth chapter the overall implementation of the proposed system is included in the fifth and sixth chapters Results and Discussion and Conclusions and Future Scope respectively

II. PROPOSED SYSTEM

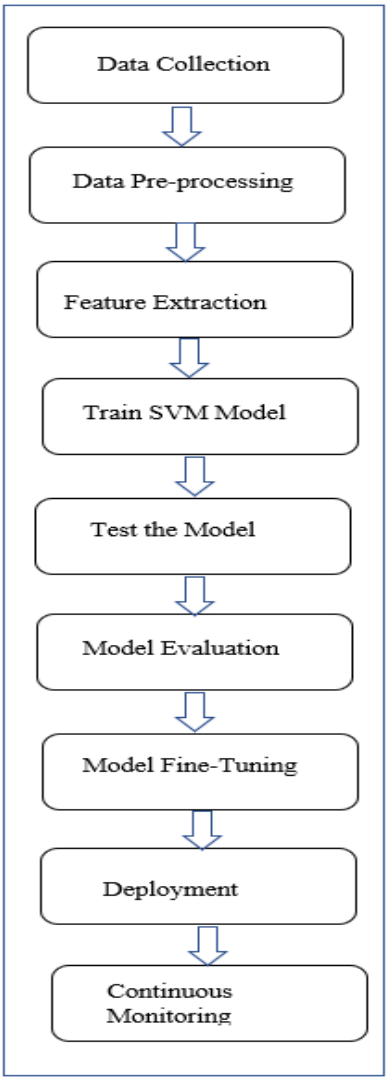


Fig .2 Proposed Methodology

In the above diagram, it shows that it involves 6 steps if we have to implement the proposed methodology

1) *Data Collection:* Collect a dataset of reviews of amazon, both fake and real.
2) Data preprocessing: Clean and prepare the reviews for analysis by removing irrelevant information, correcting spelling mistakes, and removing special characters.
3) *Feature Extraction:* Extract features from the reviews to represent the data in a form that can be used by the SVM algorithm. These features can include things like word counts, sentiment scores, and various n-grams.
4) *Train SVM Model:* Using the extracted features, train the SVM model using a labeled dataset we labeled our dataset in fake and real reviews
5) *Test the Model:* Then we Apply the model to a test dataset to evaluate its performance in detecting fake reviews
6) Model Evaluation: after training, we Evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1-score
7) *Model fine-tuning:* Based on the results of the evaluation, fine-tune the model by adjusting its parameters or using different kernels to improve its performance in detecting fake reviews
8) *Deployment:* Once the model is fine-tuned and its performance is found to be satisfactory, it can be deployed in a production environment to detect fake reviews in real-time
9) *Monitoring:* Continuously monitor and evaluate the performance of the deployed model and update it with new data

III. RESULT

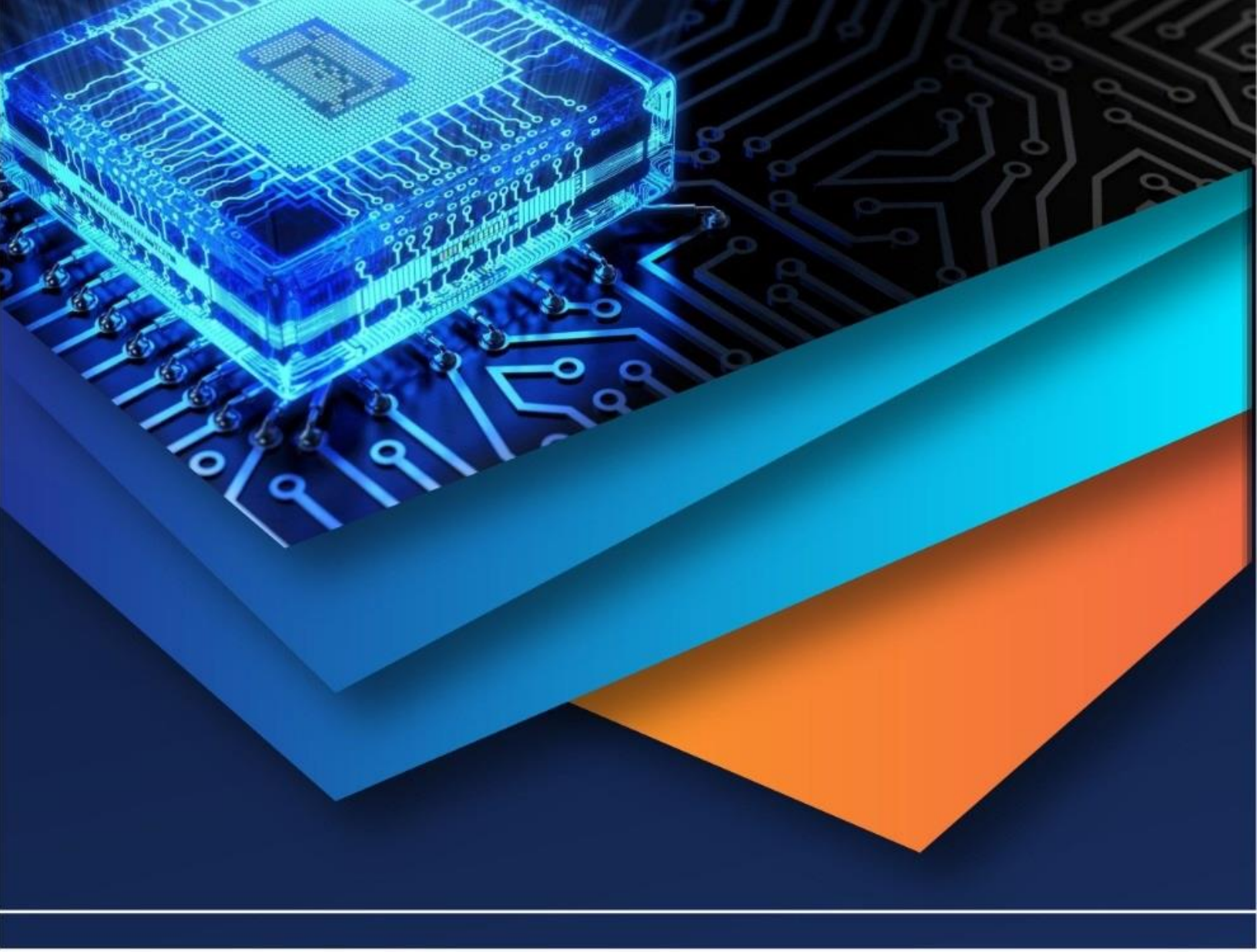
The performance of the proposed model was evaluated using a amazon dataset of 6186 reviews, which were evenly divided into a training set (4948 reviews) and a test set (1238 reviews). The model achieved an accuracy of 80.4%, a precision of 80.8%, a recall of 80%, and an F1-score of 80% on the test set. In comparison to above mention algorithm accuracy of proposed methodology is increase with 20%. In terms of error analysis, the model made more errors in detecting fake reviews that were written in a neutral or positive tone, and fake reviews that used a lot of common words. This suggests that the model may benefit from using more advanced natural language processing techniques to better understand the context of the reviews.

IV. CONCLUSION & FUTURE WORK

The research paper on fake review detection using a Support Vector Machine (SVM) algorithm has highlighted the importance of addressing the problem of fake reviews and the potential of using SVM algorithms for this task. The proposed method, which combines text-based, sentiment-based, and opinion-based features and uses a Radial Basis Function (RBF) kernel to train the SVM model, has been shown to achieve good performance in detecting fake reviews, outperforming traditional machine learning algorithms and rule-based methods. In future we can increases accuracy of the proposed methodology by using supervised and unsupervised machine learning algorithm.

REFERENCES

[1] Kaushik Daiv, Mrunal Lachake, Prathamesh Jagtap, Srishti Dhariwal, Prof. Vitthal Gutte, “An Approach to Detect Fake Reviews based on Logistic Regression using Review-Centric Features”, Volume: 07 Issue: 06 | June 2020
[2] Rami Mohawesh, Shuxiang Xu, Son N. Tran, Robert ollington Matthew Springer, Yaser Jararweh, and Sumbal Maqsood, “Fake Review Detection”, April 26, 2021
[3] Ahmed M. Elmogy1 , Usman Tariq2 , Atef Ibrahim4, “Fake Reviews Detection using Supervised Machine Learning “, Vol. 12, No. 1, 2021
[4] Aishwarya Pendyala, “Fake Consumer Review Detection”, 2019.
[5] Gourav Bathla1 • Pardeep Singh1 • Rahul Kumar Singh1 • Erik Cambria2 • Rajeev Tiwari, “ Intelligent fake reviews detection based on aspect extraction and analysis using deep learning”, 20 July 2022.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)