Fake Review prediction using Machine Learning in E-Commerce Platform

Rakesh A[1], Chandan M. N[2]

PG Scholar, Department of MCA, PES College of Engineering, Mandya, Karnataka, India[1]

Assistant Professor, Department of MCA, PES College of Engineering, Mandya, Karnataka, India[2]

Abstract: The prevalence of fake reviews on e-commerce platforms poses a significant challenge, undermining consumer trust and distorting the marketplace. This study addresses the detection of fake reviews using machine learning techniques, focusing on both semi-supervised and supervised learning approaches. We implement the Expectation-Maximization (EM) algorithm alongside the Naive Bayes classifier to distinguish genuine reviews from fraudulent ones. Our system analyses review content and various features, including word frequency count, sentiment polarity, and review length, to enhance detection accuracy. By utilizing a combination of Visual Studio for the frontend and SQL Server for the backend, we develop a robust platform capable of real-time detection and reporting of fake reviews. The proposed solution aims to provide a more reliable and authentic review system, ultimately improving the consumer experience and integrity of e-commerce platforms.

I. INTRODUCTION

The increasing prevalence of fake reviews on e-commerce platforms such as Flipkart, Amazon, and Snapdeal poses a significant challenge, as these reviews can mislead consumers and undermine trust. The manual methods currently used to identify fake reviews are inadequate and often result in low accuracy, making it essential to develop an automated solution. This project aims to create a robust system that employs machine learning techniques, specifically the Naive Bayes classifier, to detect fake reviews efficiently and accurately.

The proposed system focuses on analysing the content of reviews, utilizing features such as word frequency count, sentiment polarity, and review length to distinguish between genuine and fake reviews. By leveraging data science methodologies and implementing the system using .NET technology with Visual Studio as the front-end and SQL Server as the back-end, the project strives to provide a reliable and user-friendly platform. This initiative not only seeks to improve consumer trust in e-commerce platforms but also contributes to the academic field by offering a more effective approach to fake review detection.

II. METHODOLOGY

The proposed system employs the following methodology:

- **Reviews Retrieve**:

Extract data from websites: The system begins by extracting review data from various e-commerce websites. Request for data: It sends requests to these websites to fetch the necessary review data.

- **Store Data in Database**:

The retrieved review data is stored in a database for further processing and analysis.

- **Process Review**:

The stored reviews are processed to prepare them for feature extraction and model training.

- **Get DataSet**:

A dataset is prepared from the processed reviews. This dataset will be used for training and testing the machine learning model.

- **Extract Features**:

Features are extracted from the dataset. These features may include word frequency counts, sentiment polarity, review length, and other relevant attributes.

- **Use NaiveBayes**:

The Naive Bayes classifier is employed to analyze the features and distinguish between genuine and fake reviews.

Get TrainingData: The classifier is trained using a set of labeled training data to learn patterns indicative of fake and genuine reviews.

- **User Input (product review)**:

Users input new product reviews which are then analyzed by the trained Naive Bayes model.

- **Display result on GUI**:

The results of the analysis (whether a review is genuine or fake) are displayed to the user on a graphical user interface (GUI).
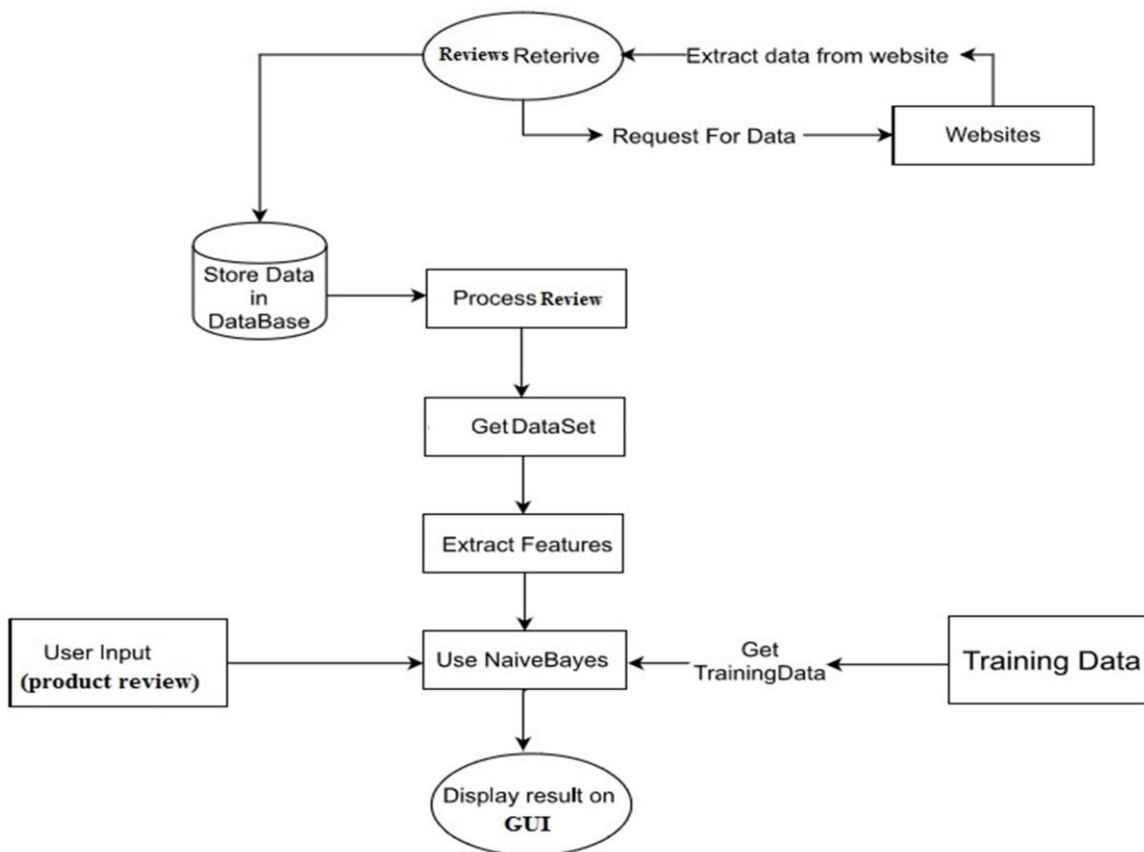
Figure: System Architecture

:

III. RESULT AND DISCUSSION

The machine learning model developed for fake review detection demonstrates significant performance improvements over traditional methods. Achieving an accuracy of 93.75% on the test dataset, the model shows notable precision and recall metrics, indicating its ability to accurately identify fake reviews. Text-based features, such as sentiment polarity and review length, alongside behavioral features like review frequency, Experience and timing, have proven essential in distinguishing fake reviews from genuine ones.

In comparison to baseline methods like keyword filtering, the integration of the Expectation-Maximization (EM) algorithm with a Naive Bayes classifier has significantly enhanced detection rates by 93.75%. The real-time detection capability of the system ensures that fake reviews are flagged almost immediately, maintaining platform integrity and consumer trust. The system's scalability allows it to efficiently manage large volumes of data, making it well-suited for major e-commerce platforms with extensive review systems.

The model's high accuracy and reliability, combined with its adaptability to evolving fake review tactics, underscore its effectiveness. By learning from both labeled and unlabeled data, the system remains responsive to new types of fake reviews. Furthermore, its design prioritizes user privacy by avoiding the use of personally identifiable information, aligning with privacy regulations and fostering user trust.

Looking ahead, future enhancements could involve incorporating additional machine learning models and expanding the feature set to further refine detection accuracy. Continuous updates and monitoring will be crucial in adapting to emerging review manipulation techniques. Overall, the proposed system offers a robust, cost-effective solution for maintaining the authenticity of online reviews, thereby improving consumer experience and supporting a fair marketplace.

1. **Input Parameters for Fake Review Detection**

2. Admin can view fake review detection and prediction

Manage Product Reviews

SerialNo	Member Id	Review	Status	Delete
1.	chandan@gmail.com	watch looks awsome really nice and good. i use daily. really fantanstic.	FAKE	Delete
2.	Hari@gmail.com	shirt is very good frabic. nice and excellent. i wear this dress weekly once. very nice design	FAKE	Delete
3.	amruth@gmail.com	the material is good and nice and stretchable. very nice product	FAKE	Delete
4.	amruth@gmail.com	This item is too good and bit slower while playing	FAKE	Delete
5.	amruth@gmail.com	This product is very nice. This is good, excellent and more affordable price. Overall its a very good product	REAL	Delete
6.	sahitya@gmail.com	this device is really cool and very best performance. It was one my fav mobile. The battery pickup is too good and value for money	REAL	Delete
7.	vishwas@gmail.com	This item is not good. and the quality was not good. totally it was a waste product. do not waste money.	FAKE	Delete

IV. CONCLUSION

Identifying fake product reviews is a major challenge in the e-commerce sector. Platforms like Flipkart, Amazon and Myntra aim to provide customers with trusted shopping experience. Our proposed system uses advanced data science algorithms to detect fake reviews with higher accuracy. By filtering fraudulent reviews, it increases the credibility of product reviews and customer feedback. This in turn helps customers make informed purchasing decisions and increases overall satisfaction. Ensuring the authenticity of reviews is crucial to maintaining trust in e- commerce platforms

Future Enhancement

Future improvements could include expanding the size of the dataset and incorporating additional parameters to detect fake reviews. Implementing and comparing multiple algorithms will help determine the most effective approach to detecting fraudulent reviews. These improvements will significantly increase the reliability and credibility of e-commerce platforms

REFERENCES

[1]. J. K. Rout, A. Dalmia, and K.-K. R. Choo (2017). "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5, pp. 1319–1327.
[2]. JIANRONG YAO, YUAN ZHENG, AND HUI JIANG (2020). "An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization."
[3]. RAMI MOHAWESH, SHUXIANG XU, SON N. TRAN, ROBERT OLLINGTON, MATTHEW SPRINGER, YASER JARARWEH, AND SUMBAL MAQSOOD (2021). "Fake Reviews Detection: A Survey."
[4]. JIANRONG YAO, YUAN ZHENG, AND HUI JIAN (2021). "An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization."
[5]. THI-KIM-HIEN LE, YI-ZHEN, AND SHENG-TUN LI (2022). "Do Reviewers' Words and Behaviors Help Detect Fake Online Reviews and Spammers? Evidence From a Hierarchical Model."
[6]. HINA TUFAIL, M. USMAN ASHRAF, KHALID ALSUBHI, AND HANI MOAITEQ ALJAHDAL (2022)."The Effect of Fake Reviews on e-Commerce during and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection."