Fake Review Detection using Machine Learning

Prof. Sujata Salunkhe, Sudip Shinde, Rushikesh Kele, Ajinkya Bhase, Saurabh Khavale ,

Abstract

In the digital era, online reviews significantly influence consumer decisions and business credibility. However, the increasing manipulation of reviews through deceptive or fake content poses a serious challenge to both customers and service providers. This project presents a machine learning-based approach to identify and filter out fake reviews by analyzing textual patterns and user behaviors. Using supervised algorithms such as Support Vector Machines (SVM) and Decision Trees (DT), along with natural language processing techniques, the system classifies reviews as genuine or fraudulent. The architecture supports user interaction via a GUI built in Python, integrated with a secure login system. The effectiveness of the proposed model is validated through multiple testing strategies, ensuring reliability and performance. By automating the detection of fake reviews, this system aims to restore trust in online platforms and aid users in making informed purchasing decisions.

I.INTRODUCTION

In today's digital economy, online reviews play a vital role in shaping consumer decisions, influencing purchasing behavior, and building brand reputation. With the rapid growth of e-commerce platforms and review-based services, customers often rely on feedback from previous buyers to assess product quality and reliability. However, the increasing presence of fake or deceptive reviews has become a serious concern. These misleading reviews are often generated to artificially boost or damage a product's image, misleading genuine buyers and affecting overall trust in online platforms.

Fake reviews may be posted for promotional gains, competitive sabotage, or even personal bias, making it difficult for both users and businesses to identify truthful information. Manually identifying such reviews is neither practical nor scalable. This has led to the emergence of automated solutions using machine learning techniques that can detect patterns and anomalies in review data.

This project proposes a machine learning-based system to detect fake reviews using textual features and behavioral indicators. By leveraging algorithms such as Support Vector Machine (SVM) and Decision Tree (DT), the system classifies reviews as either genuine or fake. The goal is to assist users in making informed decisions and help businesses maintain the credibility of their online presence. Through proper preprocessing, training, and testing, the proposed model aims to improve the reliability of review systems across digital platforms.

II.PROBLEM STATEMENT

With the growing reliance on online reviews for making purchasing decisions, the authenticity of user-generated content has become a critical issue. E-commerce platforms are increasingly affected by the presence of fake reviews—intentionally misleading feedback created to manipulate product perception. These reviews may promote low-quality products or unfairly harm the reputation of competitors. Since users often lack the means to differentiate between genuine and fake feedback, their trust in digital platforms is compromised.

Traditional methods of detecting such reviews are inefficient, subjective, and cannot scale with the massive volume of data generated online. Moreover, linguistic ambiguity and context-specific meanings make it even harder to identify deceptive content. For instance, a word like "long" could imply either a positive or negative sentiment depending on its context (e.g., long battery life vs. long loading time).

Therefore, there is a pressing need for an automated, reliable, and scalable approach that can effectively detect fake reviews. This project addresses this challenge by applying machine learning techniques to analyze review text and user behavior in order to classify reviews as genuine or fake. The aim is to support users in making informed decisions and to help businesses maintain transparency and trust.

III. LITERATURE SURVEY

Various researchers have explored the problem of fake review detection using machine learning and sentiment analysis techniques. Hassan and Islam proposed a sentiment-based model that distinguishes between positive and negative reviews to improve detection accuracy. Agarwal et al. focused on social media sentiments, highlighting how emotional tone affects perception and trust.

Krishna et al. analyzed the impact of opinion spam and compared machine learning techniques for identifying deceptive content. Bailurkar and Raul addressed bot-generated reviews on social media using sentiment analysis on Twitter data. Liu et al. introduced a temporal feature-based method using isolation forests to spot outliers in review timing.

Li et al. proposed semantic and emotion-based models to improve classification performance. Mohawesh et al. conducted a benchmark study comparing traditional ML and deep learning models, with RoBERTa showing strong results. Deng et al. applied semi-supervised learning to handle limited labeled data and detect clustered fake reviews.

These studies show that combining linguistic patterns, user behavior, and advanced ML algorithms leads to more accurate and reliable fake review detection systems**.**

IV.OBJECTIVE

The objective of this project is to design and implement a robust machine learning-based system for the detection of fake reviews across e-commerce platforms. In today's digital era, customer reviews significantly influence purchasing decisions, but the increasing prevalence of deceptive or manipulated feedback undermines trust and misguides consumers. This project aims to mitigate that risk by developing an automated solution capable of analyzing textual patterns, sentiment orientation, linguistic features, and behavioral signals to distinguish between authentic and fraudulent reviews.

Through the integration of natural language processing (NLP) techniques and supervised learning algorithms such as Support Vector Machine (SVM) and Decision Trees, the system will be trained on labeled datasets to accurately classify reviews. The goal is not only to improve the reliability of online consumer feedback but also to assist businesses in identifying malicious review behavior that may affect their brand credibility.

Furthermore, the project aspires to contribute to the broader field of opinion mining by offering a scalable and adaptable model that can be extended to multiple domains beyond e-commerce, such as hospitality, services, and mobile applications. Ultimately, this solution will support users in making better-informed decisions while promoting ethical digital practices in online platforms.

V.SYSTEM ARCHITECTURE

The system architecture for detecting fake reviews using machine learning is structured to enable seamless interaction between users and the backend classification engine. It is composed of multiple functional modules that together facilitate user management, review collection, data processing, and intelligent classification.

1. Admin Module

The admin interface serves as the central control point for managing users, products, websites, and system activity. Upon successful authentication, the administrator can:

- View and authorize registered users.
- Monitor and authorize e-commerce platforms.
- Access and review submitted product reviews.
- Track user activity including keyword searches and product interaction ratios.
- Generate reports and visualizations related to product review rankings and system usage statistics.

2. User Module

The end-user interface allows individuals to:

- Register and log in to the system.
- Search for products using keywords.
- Submit reviews for analysis.
- View the classification results (i.e., whether the review is genuine or fake).
- Review their search history and system interaction logs.

3. Review Analysis and Classification Engine

At the core of the system lies the review analysis module, which is responsible for processing and classifying textual data. This engine performs the following tasks:

- Text Preprocessing: Cleanses the input text by removing stop words, punctuation, and irrelevant content.
- Feature Extraction: Converts text into a structured format using techniques such as term frequency-inverse document frequency (TF-IDF) or other NLP-based vectorization.
- Classification: Applies supervised machine learning models like Support Vector Machine (SVM) or Decision Trees (DT) to categorize reviews into “Fake” or “Genuine.”

4. Data Management Layer

This component handles the storage and retrieval of all system-related data, including:

- User information and credentials.
- Product metadata and reviews.
- Classification results and system logs.

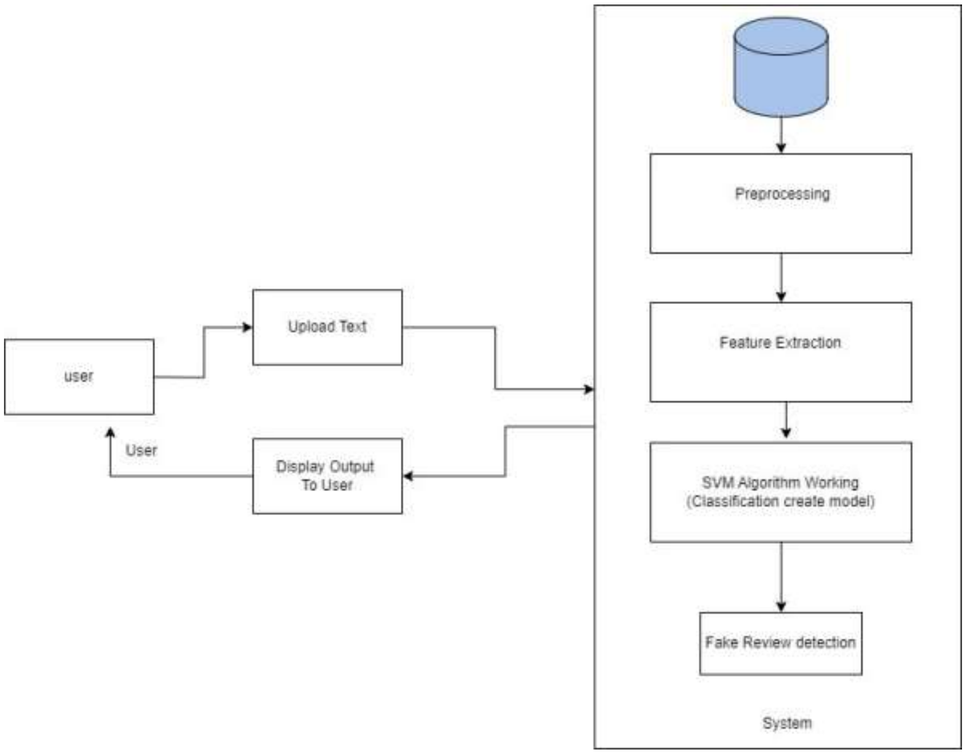
The data is stored in a secure, scalable database to ensure efficient access and management.

5. Visualization and Reporting

The system includes the capability to generate graphical insights, such as:

- Product search ratios.
- Keyword search trends.
- Review ranking outcomes.

These analytics are accessible via charts and dashboards, helping both users and administrators to understand system performance and review authenticity trends.



VI.METHODOLOGY

The methodology adopted for this project involves a systematic approach to detect and classify fake reviews using machine learning techniques. The system is structured into multiple stages, each responsible for a specific function in the detection pipeline. These stages are outlined below:

1. Data Acquisition

The initial step involves collecting review data from e-commerce websites or publicly available datasets. This data typically consists of product reviews along with metadata such as user information, timestamps, and ratings. The dataset used is labeled, containing both genuine and fake reviews, which is essential for supervised machine learning.

2. Data Preprocessing

Raw review data is often noisy and unstructured. Therefore, it undergoes a series of preprocessing steps to make it suitable for analysis. This includes:

- Text cleaning (removal of punctuation, numbers, and special characters),
- Tokenization (breaking the text into words or tokens),
- Stopword removal (eliminating commonly used words that add little value),
- Stemming or lemmatization (reducing words to their root form),
- Vectorization using techniques like TF-IDF to convert text into numerical features.

3. Feature Extraction

Meaningful features are extracted from the preprocessed reviews. These include:

- Linguistic features (word count, sentiment polarity),
- Behavioral features (frequency of reviews by a user),
- Semantic and emotional cues, which help in identifying patterns that may indicate deception.

4. Machine Learning Model Training

With the features extracted, the next step is to train machine learning models. The system primarily employs:

- Support Vector Machine (SVM) for its effectiveness in binary classification tasks, and
- Decision Tree (DT) for its ability to provide clear, interpretable decision paths. The model is trained on the labeled dataset and evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

5. Review Classification

Once trained, the model is deployed in the system to analyze new reviews. Users can input a review, which is then processed and classified as either *genuine* or *fake*. The classification result is displayed in real time and stored in the system database for future reference and analysis.

6. Visualization and Analysis

To provide insights to both users and administrators, the system includes visualization tools. These tools display product review trends, keyword-based search ratios, and other statistical outputs through charts and graphs.

7. System Integration

The entire system is built using Python and integrated with a user-friendly interface. The admin panel allows for user and review management, while the user interface supports search, submission, and classification operations.

Mathematical module :

Let S be the System

S = {I,P,O}

I - Input

P – Procerdure

O - Output

VII.RESULTS

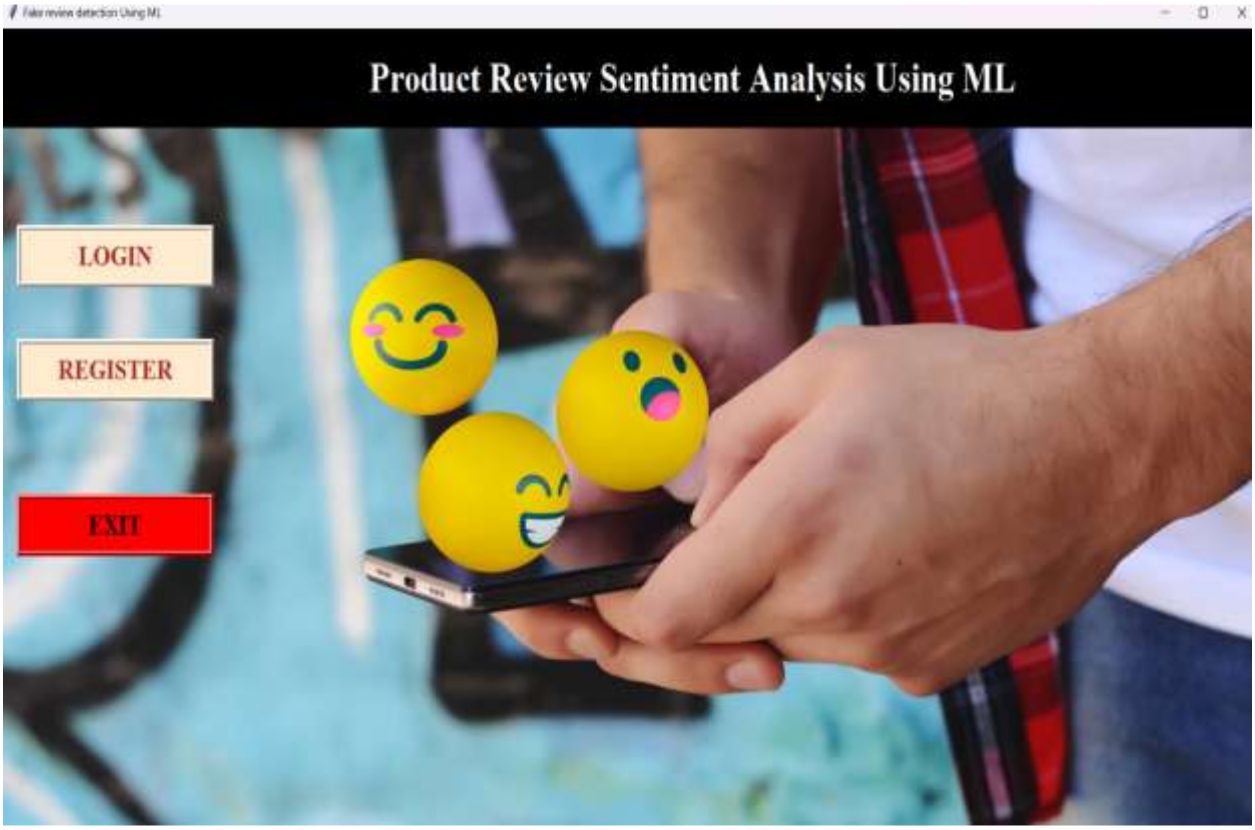
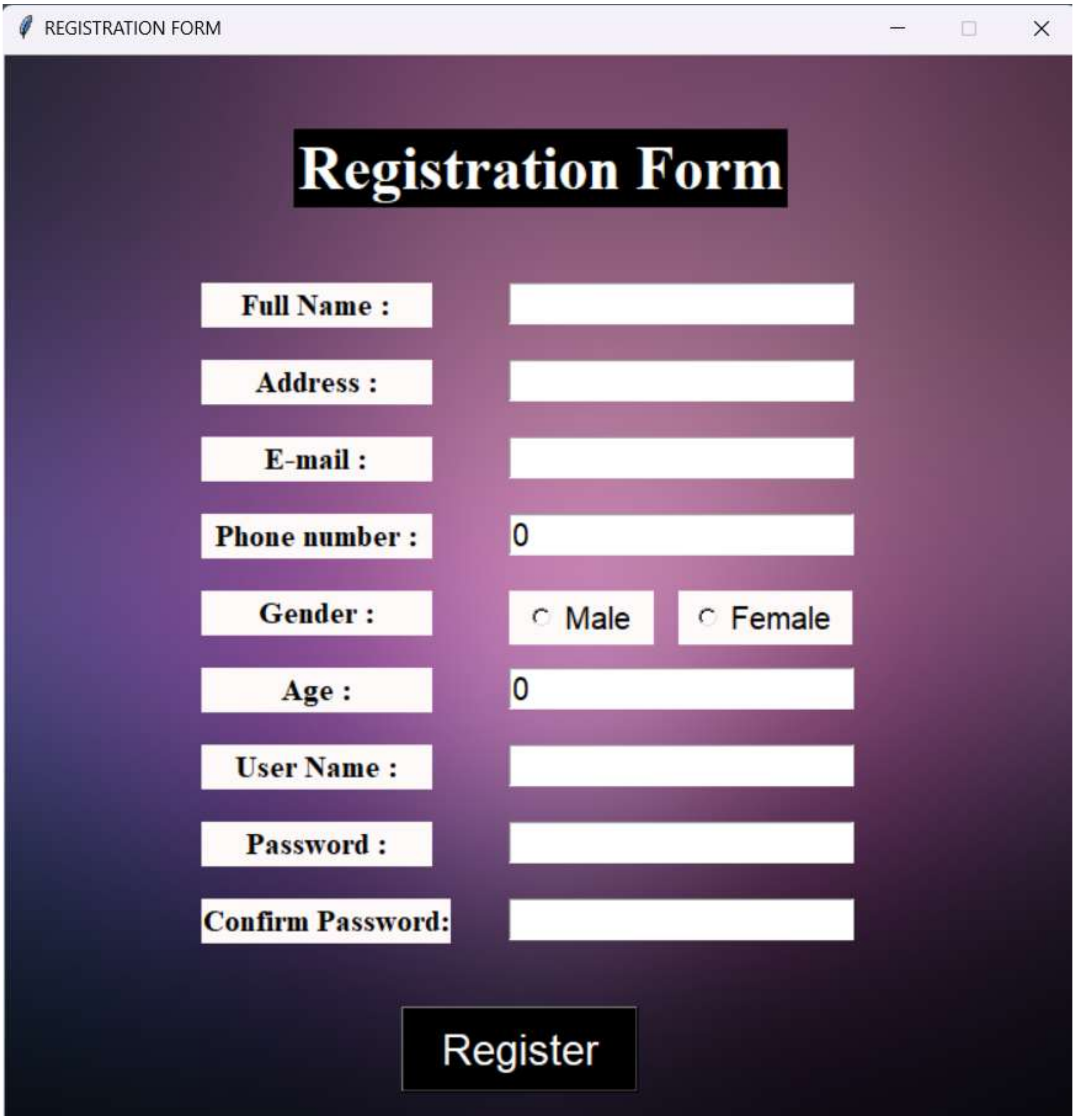

Fake review detection Using ML
Product Review Sentiment Analysis Using ML
LOGIN
REGISTER
EXIT

Fig 1 : Home Page



REGISTRATION FORM

Registration Form

Full Name :	
Address :	
E-mail :	
Phone number :	0
Gender :	○ Male ○ Female
Age :	0
User Name :	
Password :	
Confirm Password:	

Fig 2 : Registration form

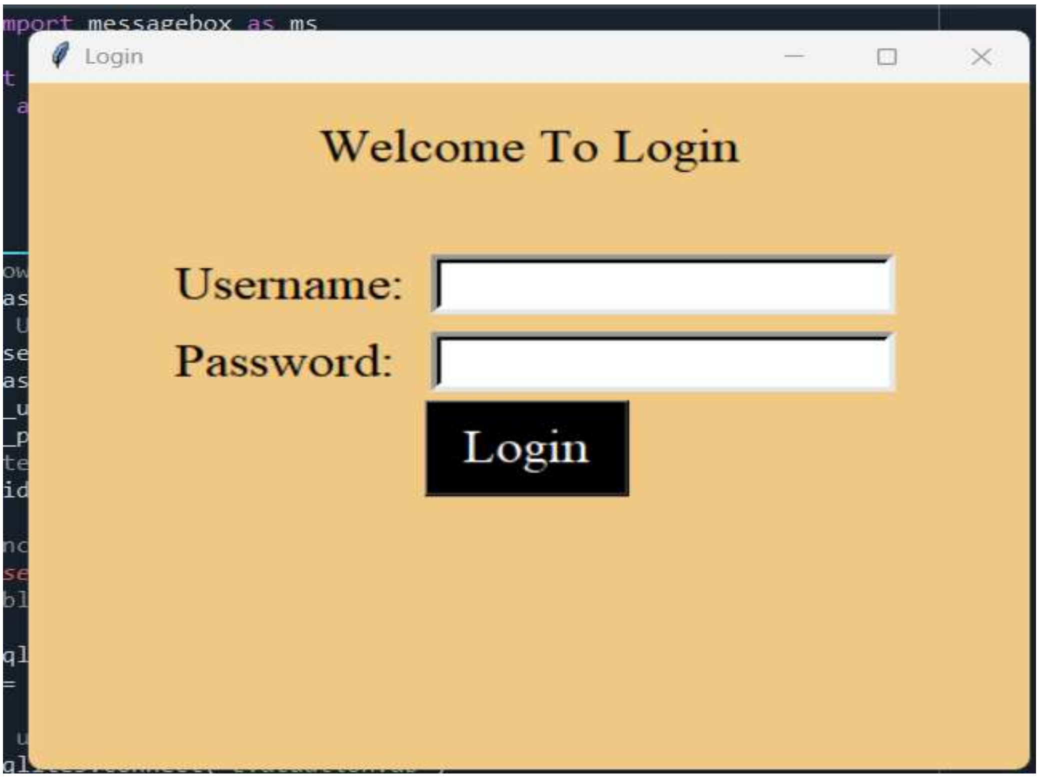


Fig 3 : Login Page

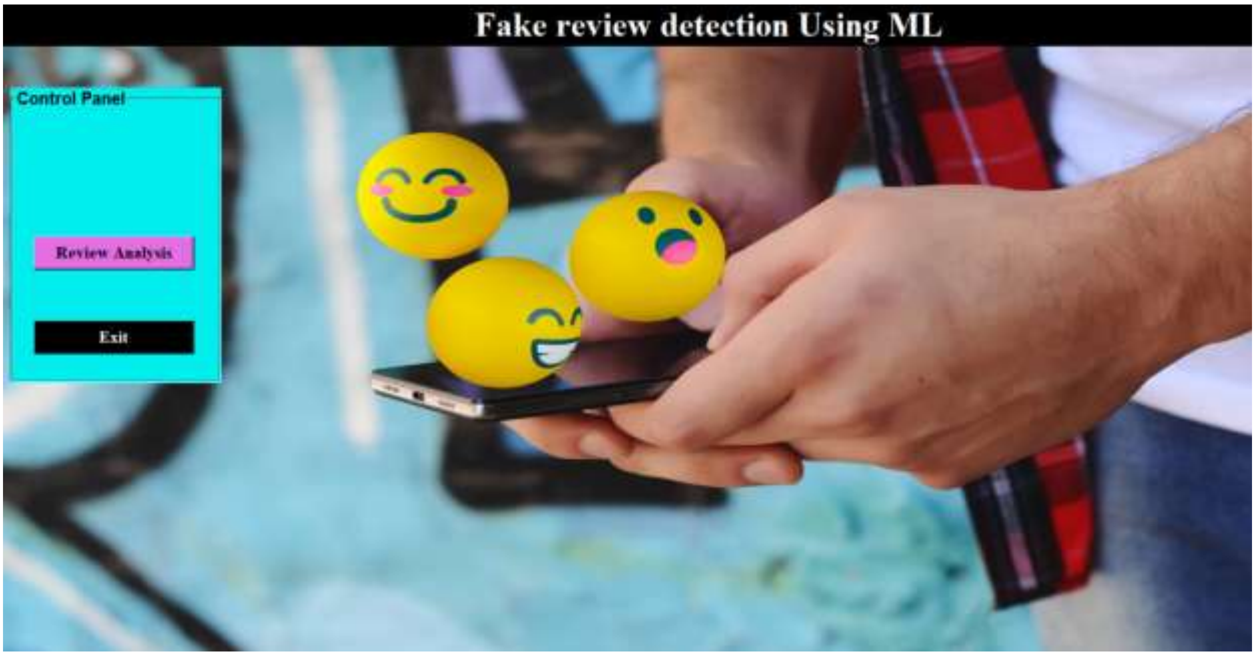
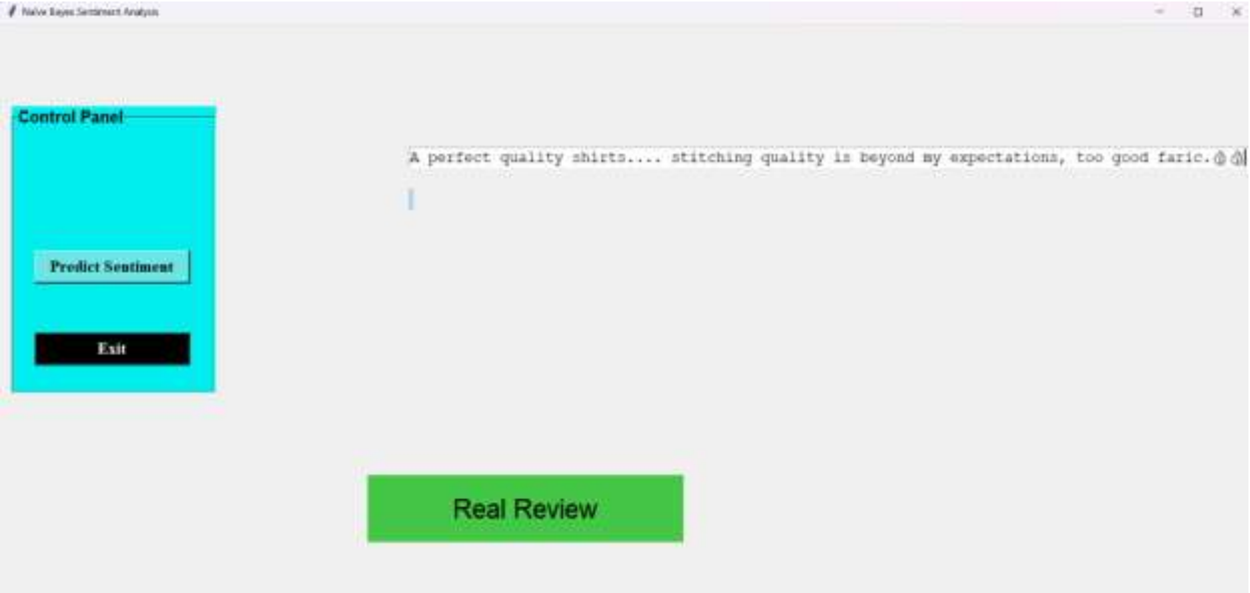
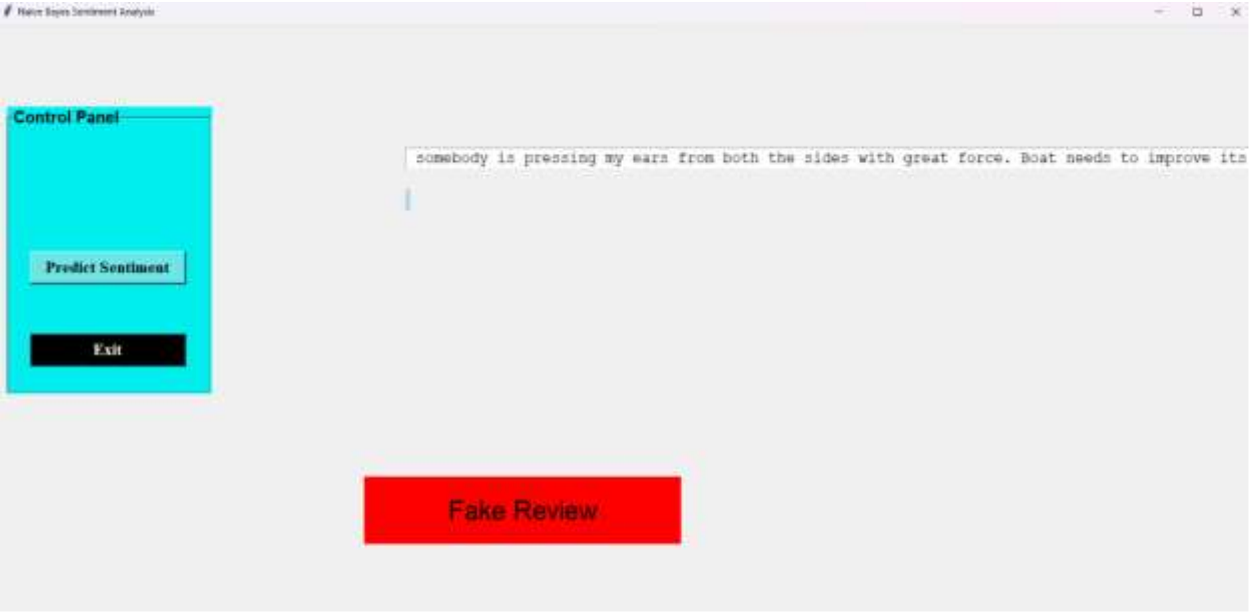


Fig 4 : Master GUI



A perfect quality shirts.... stitching quality is beyond my expectations, too good faric.

Fig 5 : Real review detected output



somebody is pressing my ears from both the sides with great force. Boat needs to improve its

Fig 6 : fake review detected output

VIII. FUTURE SCOPE

The detection of fake reviews using machine learning is a rapidly evolving domain with significant potential for advancement. As digital platforms continue to influence consumer behavior, ensuring the authenticity of user-generated content becomes increasingly critical. The following future enhancements and research directions can further improve the performance, scalability, and applicability of the proposed system:

1. Integration of Deep Learning Models
   Future work can incorporate deep learning approaches such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), or Transformer-based models like BERT. These models are capable of understanding contextual relationships within text and may yield higher accuracy compared to traditional machine learning algorithms.

2. Multilingual and Cross-Domain Review Analysis
Expanding the system to handle reviews written in different languages and from multiple domains (e.g., restaurants, mobile apps, travel services) would make the system more versatile and practical for real-world deployment.

3. Real-Time Fake Review Detection
Implementing the system for real-time use on live platforms could help detect and filter suspicious reviews before they are published, thereby improving trust among users and protecting business reputations.

4. Incorporating User Behavior and Metadata Analysis
The system can be enhanced by analyzing additional metadata such as IP address, review frequency, user history, and account creation patterns. Combining this with textual analysis could improve the robustness of fake review detection.

5. Adaptive and Continual Learning
The model can be designed to update and retrain itself periodically as it encounters new types of reviews and spam strategies. This adaptive approach will help maintain accuracy over time in dynamic online environments.

6. Integration with E-Commerce and Review Platforms
Collaborating with online platforms (such as Amazon, Yelp, or Flipkart) to integrate the detection engine directly into their infrastructure can provide widespread benefits, helping platforms maintain credibility and improve user satisfaction.

7. Improved Visualization and User Feedback System
Adding more advanced dashboards, user reporting features, and feedback loops can help refine the system and support better decision-making by both users and administrators.

IX.CONCLUSION

This project demonstrates the importance of review authenticity in the digital marketplace and presents a practical solution using machine learning to address the growing issue of fake reviews. With the increasing reliance of consumers on online reviews for making purchasing decisions, the presence of deceptive or misleading reviews poses a serious challenge to trust and transparency.

To mitigate this, a machine learning-based system was developed that analyzes both the content of reviews and behavioral patterns of reviewers to accurately classify reviews as genuine or fake. Through effective data preprocessing, feature extraction, and implementation of classification algorithms such as Support Vector Machine (SVM) and Decision Tree (DT), the system offers a reliable and scalable approach for review validation.

The results indicate that integrating machine learning into the review verification process not only enhances consumer confidence but also supports businesses in maintaining credibility. The modular design of the system ensures adaptability and allows for future enhancements such as real-time detection, multilingual support, and integration with live e-commerce platforms.

In summary, this project serves as a foundational step toward building intelligent review monitoring systems and contributes to the broader goal of ensuring data integrity and fairness in online platforms.

X.REFERENCES

1. Elmurngi, E. I., & Gherbi, A. (2018). *Detection of Unfair Reviews on Amazon Using Sentiment Analysis and Supervised Learning Techniques*. Journal of Computer Science, 14(5), 714–726.
2. O'Brien, N. (2018). *Machine Learning for Detection of Fake News*. [Online]. Available at: MIT DSpace Repository
3. Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). *Supervised Learning for Fake News Detection*. IEEE Intelligent Systems, 34(2), 76–81.
4. Liu, B., & Hu, M. (n.d.). *Opinion Mining and Sentiment Analysis*. [Online]. Available at: University of Illinois at Chicago - Sentiment Analysis Resources
5. Hill, C. (2018). *10 Ways to Spot Fake Online Reviews*. [Online]. Available at: MarketWatch
6. Sindhu, C., Vadivu, G., Singh, A., & Patel, R. (2018). *Approaches for Spam Review Detection in Sentiment Analysis*. International Journal of Pure and Applied Mathematics, 118(22), 683–690.