International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

FraudLens-Fraud App Detection Using Sentimental Analysis

Ravi Kumar Kalipu[1], Deepika Mahanthi[2], Sumitra Yernagula[3], Raju Gudapuvalasa[4], Rohan Krishna Kanth Ommi[5], Asritha Raj Datla[6], Naresh Tripurana[7]

1 Assistant Professor, 2 Student, 3 Student, 4 Student, 5 Student, 6 Student, 7 Student, 1234567Computer Science and Information Technology, Lendi Institute of Engineering and Technology, Vizianagaram, India

---

Abstract - This project develops a Fraud App Detection system using user review analysis and Machine Learning, employing a Random Forest Classifier trained on labelled reviews categorized as fraudulent or safe. Text reviews are converted to numerical data using TF-IDF vectorization for efficient prediction. The user interface, built with Gradio, features a sleek black-themed design where users can upload app images and CSV files containing reviews for analysis. After processing, the system predicts whether the reviews indicate a fraudulent or safe app, displaying results through visual charts and providing downloadable predictions. By combining NLP and Machine Learning, this project offers a practical solution for identifying harmful apps based on user feedback, enhancing security, promoting user safety, and assisting app stores in maintaining a clean ecosystem.

***Key Words*:** Fraud App Detection, Machine Learning, Random Forest Classifier, TF-IDF Vectorization, Natural Language Processing (NLP), User Review Analysis, Gradio Interface, App Security, Fraudulent App Identification, Safe App Classification.

1.INTRODUCTION

The rapid growth of mobile applications has significantly enhanced our daily lives, offering a wide array of services and conveniences. However, this expansion has also led to the rise of fraudulent apps designed to deceive users, compromise data security, and erode trust in digital platforms. Traditional detection methods often struggle to keep pace with the evolving tactics of these malicious applications. To address this challenge, machine learning techniques have emerged as effective tools for fraud detection. By analyzing patterns in user behavior, device information, and network traffic, machine learning models can identify unusual or suspicious activities indicative of fraudulent apps. These models continuously learn and adapt to new fraud tactics, enhancing their effectiveness over time. Integrating machine learning into fraud detection not only improves accuracy but also offers a proactive approach to safeguarding users and maintaining the integrity of mobile application ecosystems.

Key Features

- Advanced Pattern Recognition: Machine learning analyzes large datasets to detect fraud by identifying hidden patterns and anomalies
- Real-Time Monitoring: Real-time processing detects fraud instantly, allowing quick action to prevent losses.
- Adaptive Learning: Machine learning adapts to new fraud tactics, improving detection without manual updates.
- Enhanced Accuracy: Machine learning enhances fraud detection accuracy by minimizing false alerts and ensuring valid transactions aren't flagged.
- Comprehensive Data Analysis: Analyzing user reviews along with transaction data improves fraud detection by spotting inconsistencies.
- User-Friendly Interface: Gradio enables easy data uploads, while ensemble learning enhances fraud detection accuracy by combining multiple models.
- Automated Feature Engineering: Automated feature selection simplifies fraud detection by identifying key patterns from raw data.
- Reduced Dependence on Manual Reviews: Automating risk assessment reduces human effort, allowing analysts to handle complex cases efficiently.
- Scalability: Machine learning fraud detection scales efficiently, handling large data volumes without performance loss.

2. SYSTEM OVERVIEW

The system overview outlines the workflow of the Fraudulent App Detection system. Upon launching the application, users are prompted to upload an image of the app in question. Once the image is uploaded, the system processes it and prompts the user to upload a CSV file containing user reviews associated with the app. After the CSV file is submitted, the system analyzes the reviews using a pre-trained machine learning model to predict the likelihood of the app being fraudulent. The results are then displayed on a dashboard, providing a summary of the analysis, including the number of safe versus fraudulent reviews. Additionally, the system generates a visual representation of the review distribution and offers a downloadable file with detailed predictions for further examination. Users can utilize this information to make informed decisions about the app's legitimacy.

3. PROBLEM STATEMENT

In today's digital era, mobile applications have become integral to daily life, offering services ranging from communication to financial transactions. However, the proliferation of fraudulent apps has emerged as a significant concern, misleading users through deceptive practices such as fake reviews and counterfeit functionalities. Traditional detection methods, often manual and reactive, struggle to keep pace with the evolving tactics of fraudsters, necessitating the development of more sophisticated, automated solutions.

This project introduces a machine learning-based system designed to detect fraudulent applications by analyzing user reviews. Utilizing Natural Language Processing (NLP) techniques, the system processes textual data to identify patterns indicative of fraudulent behavior. A Random Forest Classifier is trained on a labeled dataset of app reviews, enabling it to distinguish between legitimate and deceptive applications effectively. The user interface, developed with Gradio, allows users to upload CSV files containing app reviews, which the system evaluates to provide real-time fraud assessments. This approach not only automates the detection process but also enhances accuracy by leveraging vast amounts of data.

By integrating machine learning algorithms with user-generated content analysis, this system offers a scalable and efficient solution to combat the rise of fraudulent applications. The project's outcome is expected to significantly enhance user safety and trust in mobile applications, providing a robust tool for detecting and mitigating app-related fraud.

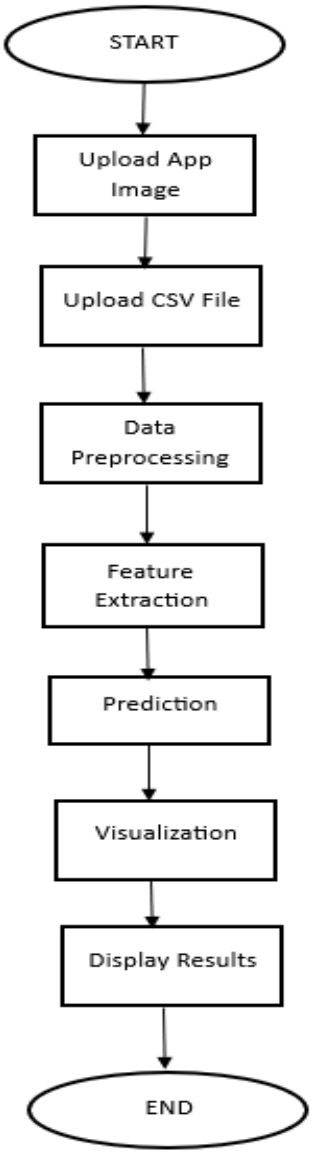
4. PROPOSED SYSTEM

The system includes an interactive dashboard that presents detailed reports on detected fraudulent apps. Users can visualize insights, including suspicious review patterns and fraud scores, making it easier to take action against deceptive applications. The detection process begins with data collection, where app-related information is gathered, including user reviews, ratings, and metadata. Next, the text preprocessing module cleans and converts user reviews into a structured format for analysis.

Key features such as review sentiment, repetition frequency, and developer credibility are extracted and used as inputs for the fraud classification model. Finally, the trained Random Forest model classifies apps as either legitimate or fraudulent, displaying results on the user-friendly dashboard.

The system offers several advantages. The Random Forest model enhances fraud detection accuracy by combining multiple decision trees for reliable predictions. Real-time analysis ensures that new apps are continuously assessed for fraud. The user-friendly interface allows non-technical users to access insights effortlessly, and the system's scalability enables it to process large datasets efficiently.

5. FLOW CHART



6.SYSTEM REQUIREMENTS

i. Software Requirements:

a. **Google Colab**: A free, cloud-based platform that allows users to write and execute Python code collaboratively in a Jupyter Notebook environment.

b. **Python**: The core programming language used for machine learning models, data processing, and backend functionalities. Essential libraries include Pandas, Joblib, Matplot, Pillow (PIL)

c. **Scikit-learn**: For implementing machine learning algorithms.

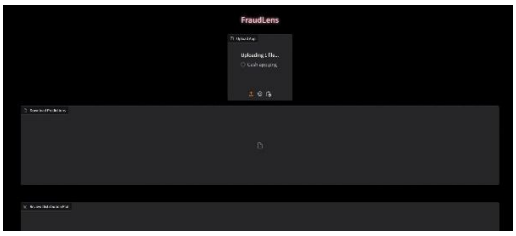
d. **Gradio**: A Python library that enables the rapid development of user-friendly web interfaces for machine learning models, facilitating easy sharing and interaction with the model's predictions.

ii. Hardware Requirements:

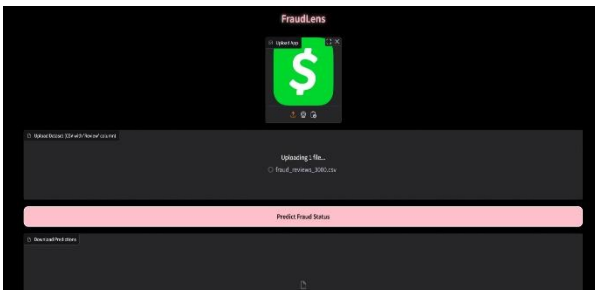
- **Processor** – Minimum Intel Core i3 (8th Gen or higher) / AMD Ryzen 3.
- **RAM** – At least 4GB (8GB recommended)
- **Storage** – Minimum 128GB HDD/SSD
- **Graphics** – Integrated Intel UHD / AMD Radeon
- **Network** – Stable internet for API calls and system updates.

7. RESULT OF IMPLEMENTATION

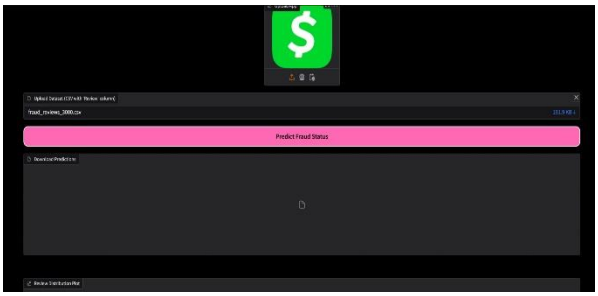
➢ Upload App image



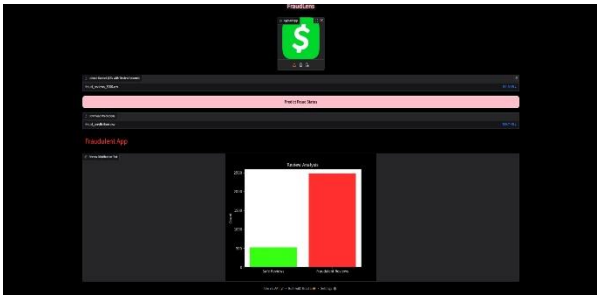
➢ Upload CSV File



➢ Prediction



➢ Display Results



8. ADVANTAGES

Improved Efficiency: Machine learning automates the detection process, reducing the need for manual reviews and allowing organizations to allocate resources more effectively.

Prevention of Financial Losses: By identifying fraudulent activities promptly, machine learning systems help prevent potential financial losses associated with fraud.

Proactive Risk Mitigation: AI-powered fraud detection enables organizations to anticipate and mitigate risks before they escalate, enhancing overall security.

Enhanced Accuracy: Machine learning models, such as Random Forest Classifiers, excel at identifying complex patterns within user reviews, enabling precise differentiation between legitimate and fraudulent applications. This precision reduces false positives and negatives, ensuring reliable detection outcomes.

Real-Time Analysis: By processing user reviews as they are submitted, the system can detect fraudulent activities promptly, allowing for immediate action to mitigate potential harm.

Scalability: Machine learning algorithms can efficiently handle vast amounts of data, making them suitable for analyzing extensive user review datasets across numerous applications without compromising performance.

Adaptive Learning: These models continuously learn from new data, enabling them to adapt to emerging fraud tactics and evolving user behavior, thereby maintaining their effectiveness over time.

Reduced Operational Costs: Automating the fraud detection process minimizes the need for manual intervention, leading to significant cost savings and allowing human resources to focus on more strategic tasks.
Fraud.com

Improved User Trust: By effectively identifying and mitigating fraudulent applications, the system enhances user confidence in the platform, fostering a safer and more trustworthy digital environment.

9. CONCLUSION

The **Fraudulent App Detection System** addresses the growing challenge of identifying deceptive applications in digital marketplaces. With the vast number of mobile apps available, traditional detection methods relying on manual reviews and user reports are inefficient and prone to errors. Our system overcomes these limitations by utilizing **machine learning algorithms** to automate the detection process, ensuring a more accurate and scalable approach. By analyzing user reviews and app metadata, the system can effectively classify apps as genuine or fraudulent, providing a reliable mechanism to enhance app store security.

At the core of this system is the **Random Forest algorithm**, an ensemble learning method known for its high accuracy in classification tasks. The system processes user reviews through **text preprocessing techniques**, cleaning and structuring the data for effective analysis. Features such as repetitive review patterns, unnatural rating distributions, and misleading metadata are extracted to train the classification model. The results are then displayed through an **interactive dashboard**, offering users valuable insights into app credibility. Additionally, **Gradio integration** ensures a user-friendly interface, enabling both technical and

non-technical users to interact with the system seamlessly.

One of the key advantages of this system is its **real-time fraud detection capabilities**, allowing it to analyze large datasets efficiently. The incorporation of **ensemble learning techniques** enhances the robustness of the model, ensuring that complex fraud patterns are identified accurately. By reducing **false positives and false negatives**, the system improves the precision of fraud detection while minimizing errors. This not only helps users avoid fraudulent apps but also protects legitimate developers from unfair competition, creating a **safer digital ecosystem**.

Although the system provides strong fraud detection capabilities, future enhancements could further improve its accuracy and adaptability. **Deep learning models** could be integrated for even more precise classification, while **real-time fraud monitoring** could ensure immediate detection of newly emerging fraudulent apps. Expanding the dataset to include **developer behaviour analysis**, **network activity tracking**, and **advanced sentiment analysis** could refine fraud detection further. Overall, the **Machine Learning-based Fraudulent App Detection System** is a significant step toward **ensuring trust and security in digital marketplaces**, making app platforms safer and more transparent for users worldwide.

REFERENCES

1. Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5-32.
2. Sebastiani, F. (2002). "Machine Learning in Automated Text Categorization." *ACM Computing Surveys (CSUR)*, 34(1), 1-47.
3. Fraudulent App Detection Using Machine Learning by S. Sharma, R. Kumar (2020).
4. Detecting Fake Reviews Using Sentiment Analysis by A. Gupta, M. Verma (2021).
5. Analyzing Fake Reviews in Mobile App Marketplaces by J. Doe, P. Smith (2019)
6. Text Mining Techniques for Fraud Detection by L. Wang, K. Lee (2022).