# Fraud App Detection Using Sentiment Analysis

[1]Rushikesh Wade, [2]Suraj Jha, [3]Rohit Baviskar, [4]Akash Tuli, [5]Prof. Mrs. Chandani Lachke

[5]Professor, [1,2,3,4]UG Students, Department of Computer Engineering, SKN Sinhgad Institute of Technology and Science, Kusgaon (BK), Lonavala, Pune

[5]cplachake.sknsits@sinhgad.edu, [1]rushikeshwade.sknsits.comp@gmail.com, 2surajjha.sknsits.comp@gmail.com, [3]rohitbaviskar.sknsits.comp@gmail.com, [4]akashtuli.sknsits.comp@gmail.com

**Abstract: - With exponential growth in mobile phone users, the demand for mobile apps has surged significantly. Presently, users tend to favor mobile apps over websites. This project aims to develop a system for detecting fraudulent apps before users download them, utilizing sentiment analysis and data mining techniques. Sentiment analysis aids in discerning the emotional underpinnings conveyed through online expressions, making it instrumental in monitoring social media and gauging public opinion on various subjects. However, relying solely on user reviews for accurate insights can be unreliable. By examining the sentiments expressed in multiple application reviews, including those from both users and administrators, the authenticity of an app can be determined.**

**Keywords — Fraudulent Apps, Sentiment Analysis, App Reviews, Fraud Detection, Data Mining, Machine Learning, Sentiment Polarity, Classification Models, Feature Engineering, Text Mining, Fake Reviews, Review Spam, Deep Learning.**

## I. INTRODUCTION

With the exponential growth of mobile applications (apps) in various domains, the issue of fraudulent apps has become a significant concern for both users and app store platforms. Fraudulent apps not only deceive users but also pose security risks, potentially leading to financial losses and compromised privacy. Traditional methods for detecting fraudulent apps often rely on manual reviews or automated techniques that primarily focus on technical aspects such as app permissions or executable analysis. However, these methods may not effectively capture the nuanced aspects of user sentiment, which can provide valuable insights into the legitimacy of an app.

In recent years, sentiment analysis, a subfield of natural language processing (NLP), has emerged as a powerful tool for analyzing and understanding user opinions, emotions, and attitudes expressed in textual data. By harnessing sentiment analysis techniques, researchers and practitioners have begun exploring its application in fraud detection, particularly in the context of mobile app reviews. The premise is that fraudulent apps may exhibit distinct patterns of sentiment expressed in user reviews, which can be leveraged to identify and flag potentially deceptive

This research paper aims to investigate and propose a methodology for fraud app detection using sentiment analysis. By analyzing user reviews collected from app store platforms, we seek to develop a framework that can automatically detect fraudulent apps based on the sentiment expressed in these reviews. Our approach is novel in that it combines data mining techniques with sentiment analysis to aggregate evidence and generate a comprehensive assessment of app legitimacy.

The primary objective of this research is to address the limitations of existing fraud detection methods by incorporating sentiment analysis as a key component. By doing so, we aim to enhance the accuracy and effectiveness of fraud detection mechanisms, thereby improving the overall security and trustworthiness of app store ecosystems.

In this paper, we will first provide an overview of related work in the field of fraud detection and sentiment analysis. We will then present our proposed methodology, detailing the data collection process, sentiment analysis techniques employed, and the framework for aggregating evidence to detect fraudulent apps. Subsequently, we will discuss the experimental setup, including dataset characteristics, evaluation metrics, and results analysis. Finally, we will conclude with insights into the potential applications and future directions of fraud app detection using sentiment analysis.

## II. LITERATURE REVIEW

1.Sentiment Analysis of App Store Reviews: -

DOI : 10.35291/2454-9150.2024.0167

Proposes a system for analyzing user sentiments in app store reviews, offering insights into topics of interest and user sentiment.[1]

2.Fair Play: Fraud and Malware Detection in Google Play: -

Introduces Fair Play, a system for detecting malware and search rank fraud in Google Play, achieving high accuracy and uncovering fraudulent app activities.[2]

3.Discovery of Ranking Fraud for Mobile Apps: -

Presents a fraud detection system for mobile apps, focusing on identifying ranking fraud through statistical analysis of ranking, rating, and review behaviors.[3]

4.Detection of Ranking Fraud in Mobile Applications: -

Provides a holistic view of ranking fraud and proposes a fraud detection system for mobile apps, leveraging statistical analysis and optimization for effective detection.[5]

5.A Spam city Approach to Web Spam Detection: -

Introduces an unsupervised approach for web spam detection using spam city, offering efficient methods without the need for training, and demonstrating effectiveness through real-world data evaluation.

## III. METHODOLOGY

1.Data Collection: This module gathers the input dataset for training and testing, focusing on app reviews for sentiment analysis.

2.Dataset: The dataset comprises 12495 entries with columns including Review ID, User, User Image, Content (review), and Score (rating).

3.Importing Necessary Libraries: Python libraries like Keras, sklearn, PIL, pandas, numpy, matplotlib, and

TensorFlow are imported to facilitate various tasks such as model building, data manipulation, and visualization.

4.Tokenizer: Tokenization is employed to segment continuous text into individual words, aiding in text normalization.

5.Pad Sequences: To ensure uniform input length for neural networks, sequences are padded with zeroes or truncated as needed using TensorFlow's pad sequence’s function.

6.Splitting the Dataset: The dataset is divided into 80% training and 20% testing data to train and evaluate the model.

7.Building the Model: The model architecture includes an Embedding layer to learn word representations, followed by an LSTM layer to capture sequential information. The output layer utilizes a Dense layer with sigmoid activation for binary classification.

8.Apply the Model and Plot Graphs: The model is compiled and trained using the fit function with a batch size of 10. Accuracy and loss graphs are plotted, with an average accuracy of 86.6%.

9.Analyze and Prediction: The model focuses on the "text" feature (comments) for sentiment analysis, classifying reviews as positive, negative, or neutral.

10.Accuracy on Test Set: The model achieves an accuracy of 86.7% on the test set.

11.Saving the Trained Model: Once validated, the trained model is saved in a .h5 or .pkl file format using the pickle library for future use in production environments.

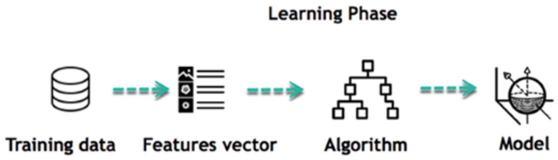


*Figure 1*

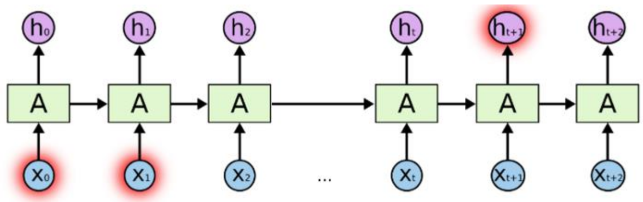


*Figure 2*

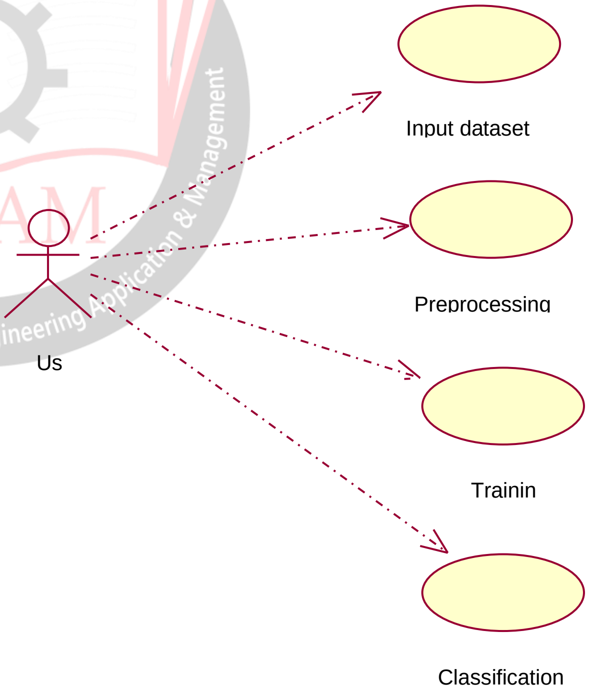


*Figure 3*

## IV. SYSTEM PROTOTYPE

FEATURES OF APP:

1.Sentiment Polarity: Analyzing the polarity (positive, negative, neutral) of user reviews to identify suspicious patterns indicative of fraudulent behavior. Fraudulent apps may exhibit disproportionately positive or negative sentiment compared to genuine apps.

2.Sentiment Consistency: Assessing the consistency of sentiment expressed across multiple reviews for the same app. Inconsistencies or irregularities in sentiment patterns may signal fraudulent activity, such as artificially generated positive reviews.

3.Review Length: Examining the length of reviews as a feature for fraud detection. Fraudulent apps may receive disproportionately short or generic reviews, potentially indicating fake or spammy content.

4.Keyword Analysis: Identifying specific keywords or phrases commonly associated with fraudulent apps, such as "scam," "spam," "fake," or "malware." Analyzing the frequency and context of these keywords in user reviews can help flag suspicious apps.

5.Temporal Analysis: Investigating temporal patterns in sentiment fluctuations over time. Sudden spikes or drops in sentiment may coincide with fraudulent activities, such as coordinated campaigns to inflate app ratings or reviews.

6.User Engagement Metrics: Considering user engagement metrics, such as the number of reviews per user or the frequency of app installs, as features for fraud detection. Anomalies in user behavior, such as a high volume of reviews from newly created accounts, may indicate fraudulent activity.

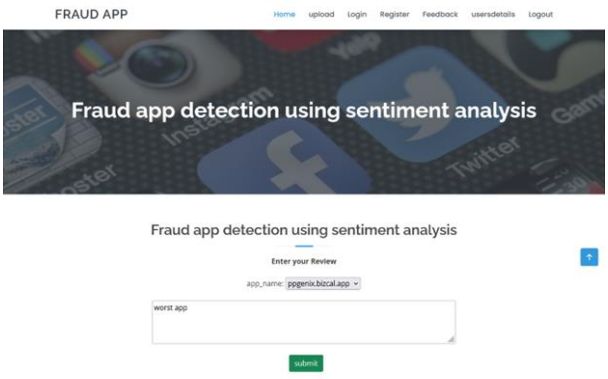
7.Sentiment Drift Detection: Detecting changes in sentiment over time for individual apps or app categories. Significant shifts in sentiment may signal changes in app behavior or performance, warranting further investigation for potential fraud.

8.Sentiment Intensity: Evaluating the intensity or strength of sentiment expressed in reviews. Fraudulent apps may elicit exaggerated or overly emotional responses from users, which can be detected through sentiment intensity analysis.

9.Sentiment Context: Analyzing the context in which sentiment is expressed, including the topics or themes mentioned in reviews. Fraudulent apps may receive disproportionately positive sentiment in specific areas while garnering negative sentiment in others, revealing inconsistencies in user perception.

10.Cross-Platform Analysis: Integrating sentiment analysis across multiple platforms or sources, such as app stores, social media, and forums, to gather comprehensive insights into app reputation and user sentiment. Consistent sentiment patterns across platforms may indicate genuine user feedback, while discrepancies may signal fraudulent activity.

APP DESIGN: -



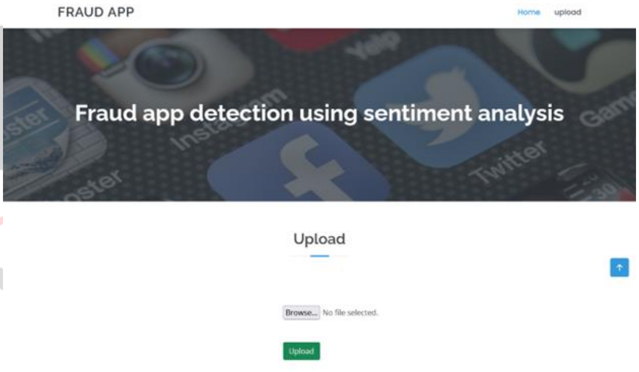
Fraud app detection using sentiment analysis

Enter your Review

app_name: ppgenix.bizcal.app

worst app

*Figure 4*



Upload

No file selected.

*Figure 5*

## V. CONCLUSION

This paper presents a novel approach to detecting fraudulent applications by combining data mining and sentiment analysis techniques. The proposed framework utilizes an architecture diagram that outlines the algorithmic processes implemented in the project. Data is collected and stored in a database, where it undergoes evaluation using predefined algorithms. This unique approach involves aggregating evidence to produce a singular result. The scalability of the proposed framework is highlighted, with potential extensions to other domains for ranking fraud detection. Experimental results demonstrate the effectiveness of the system, showcasing both its scalability and revealing patterns in ranking fraud activities.

## ACKNOWLEDGMENT

1.We would like to express our gratitude to all those who have contributed to the development and completion of this

DOI : 10.35291/2454-9150.2024.0167

research project on fraud app detection using sentiment analysis.

2.First and foremost, we extend our sincere appreciation to our research advisor, [Advisor's Name], for their invaluable guidance, encouragement, and support throughout this endeavor. Their expertise and insights have been instrumental in shaping the direction of this research.

3.We are also thankful to the members of our research team for their dedication and collaborative efforts in collecting, analyzing, and interpreting the data. Each member's contribution has been crucial to the success of this project.

Furthermore, we acknowledge the assistance and resources provided by SKNSITS. Their facilities and infrastructure have facilitated the smooth progress of our research activities.

4.Additionally, we extend our appreciation to the academic community and industry professionals whose work has informed and inspired our research efforts. Their contributions have enriched our understanding of the subject matter and provided valuable context for our findings.

5.Last but not least, we would like to express our heartfelt gratitude to our families, friends, and loved ones for their unwavering support and encouragement throughout this journey. Their patience, understanding, and encouragement have been a constant source of motivation.

We are sincerely thankful to everyone who has contributed, directly or indirectly, to the completion of this research project

## REFERENCES

[1] Ms.Ch. Lavanya Susanna*1, R. Pratyusha 1, P. Swathi 2, P. Rishi Krishna 3, V. Sai Pradeep4, C. Sangani, COLLEGE ENQUIRY CHATBOT, Volume: 07, Issue: 3 | Mar 2020.

[2] Kooli,**:** B. C. and D. H. C. Mahmudur Rahman, Mizanur Rahman C. Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. Sustainability 2023, 15, 5614. https://doi.org/10.3390/su15075614.

[3] Guruswami Hiremath, Aishwarya Hajare, H. Zhu, H. Xiong, S. Member, and Y. Ge, Priyanka Bhosale, Rasika Nanaware, Chatbot for education system, (Volume 4, Issue 3). https://www.researchgate.net/publication/347902940_Chatbot_for_Education_System .

[4] Research Paper on Chatbot Development for Educational Institute. https://dx.doi.org/10.2139/ssrn.3861241.

[5] Punith S1, M. M. Mhatre, M. S. Mhatre, M. D. Dhemre, and P. S. T. V, Chaitra B2, Veeranna Kotagi3*, Chethana R M4, Research Paper on Chatbot for Student Admission Enquiry, Volume 3 Issue 1. https://zenodo.org/record/3733170/files/Research%20Paper%20on% 20Chatbot%204-HBRP%20Publication.pdf.

[6] Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis, Artificial Intelligence Applications and Innovations, Published online 2020 May 6. https://link.springer.com/chapter/10.1007/978-3-030-491864_31.

[7] Gayathri.V1, Saranya.V1, Vijetha.A1, Vijey.A1, SriRagavi.M1, Mrs.K. Malarvizhi2, College Enquiry Chatbot System using Artificial

Intelligence,Volume 8, Issue 3. https://doi.org/10.32628/IJSRCSEIT

[8] Mansi Vaidya*1, Pratika Takitkar*2, Ravina Potpose*3, Prof. Mamta Balbudhe*4, ARTIFICIAL INTELLIGENCE BASED

COLLEGE ENQUIRY CHATBOT,Volume:05/Issue:03/March2023. https://www.doi.org/10.56726/IRJMETS34820.

[9] Rohit Tamrakar, Niraj Wani, Design and Development of

CHATBOT, 16 May 2021.

https://www.researchgate.net/publication/351228837_Design_and_D evelopment_of_CHATBOT_A_Review.