iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58521

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 12 Issue II Feb 2024- Available at www.ijraset.com

Survey on Fraud App Detection Using Sentimental Analysis and Machine Learning

Jaitee Bankar[1], Sakshi Sayankar[2], Kunal Veer[3], Kasturi Desai[4], Chandansingh Rajput[5]
Department of Information Technology, RMD Sinhgad School of Engineering

Abstract: ***Ranking fraud is the practice of acting dishonestly or deceitfully to artificially improve an App's position on a popularity list in the mobile app market. In actuality, app makers are increasingly using ranking fraud. These behaviors include posting phoney app reviews or exaggerating the sales of their apps. Although the importance of combating ranking fraud has long been recognized, there is little literature and research available in this area. To achieve this, we give a thorough examination of fraud app detection in this work utilizing sentiment analysis and spam filtering and propose a technique for identifying it in mobile apps. We specifically suggest mining the active times, or leading sessions, of mobile Apps to precisely locate the ranking scam in the first place.***

Keywords: Mobile Apps, Fraud Detection, Rating and Review, sentiment analysis.

I. INTRODUCTION

The number of smartphone apps has grown dramatically during the past several years. For instance, as of the end of April 2013, more than 1.6 million Apps were available on the Google Play and Apple App Store, respectively. To promote the development of mobile Apps, many App shops built daily App leaderboards, which display the chart rankings of the most popular Apps. Unquestionably one of the most important resources for promoting mobile apps is the App leader board. As a result, app developers routinely research different tactics, like advertising campaigns, to promote their apps to get them rated as highly as possible in such app leaderboards. Instead of relying on traditional marketing approaches, however, unethical app developers have recently been more and more inclined to use a variety of dishonest strategies to purposely boost their apps and ultimately affect the chart placed on an app store. To swiftly inflate the quantity of App downloads, ratings, and reviews, "bot farms" or "human water armies" are frequently used to achieve this goal.

II. LITERATURE SURVEY

Detailed The pair-wise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collusions in spam campaigns from a more fine-grained perspective. In [1] paper, Spam campaigns spotted on popular product review websites (e.g., amazon.com) have attracted mounting attention from both industry and academia, where a group of online posters is hired to collaboratively craft deceptive reviews for some target products. The goal is to manipulate the perceived reputations of the targets for their best interests.

In [2] paper, Online product reviews have become an important source of user opinions. Due to profit or fame, imposters have been writing deceptive or fake reviews to promote and/or demote some target products or services. Such imposters are called review spammers. In the past few years, several approaches have been proposed to deal with the problem. This work takes a different approach, which exploits the burrstones nature of reviews to identify review spammers.

In [3] paper, Online reviews on products and services can be very useful for customers, but they need to be protected from manipulation. So far, most studies have focused on analyzing online reviews from a single hosting site. How could one leverage information from multiple review hosting sites? This is the key question in our work. In response, develop a systematic methodology to merge, compare, and evaluate reviews from multiple hosting sites. focus on hotel reviews and use more than 15 million reviews from more than 3.5 million users spanning three prominent travel sites.

In [4] paper, Online reviews have become an increasingly important resource for decision-making and product designing. But review systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, the ground truth of large-scale datasets is still unavailable and most of the existing approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. Working with Dianping1, the largest Chinese review hosting site, present the first reported work on fake review detection in Chinese with filtered reviews from Damping's fake review detection system.

In [5] paper, Online reviews are quickly becoming one of the most important sources of information for consumers on various products and services. With their increased importance, there exists an increased opportunity for spammers or unethical business owners to create false reviews to artificially promote their goods and services or smear those of their competitors. In response to this growing problem, there have been many studies on the most effective ways of detecting review spam using various machine learning algorithms. One common thread in most of these studies is the conversion of reviews to word vectors, which can potentially result in hundreds of thousands of features.

In [6] paper, provides an efficient and effective method to identify review spammers by incorporating social relations based on two assumptions that people are more likely to consider reviews from those connected with them as trustworthy and review spammers are less likely to maintain a large relationship network with normal users. The contributions of this paper are two-fold: (1) elaborate how social relationships can be incorporated into review rating prediction and propose a trust-based rating prediction model using proximity as trust weight, and (2) design a trust-aware detection model based on rating variance which iteratively calculates user-specific overall trustworthiness scores as the indicator for spam city.

In [7] paper, to detect fake reviews for a product using the text and rating property from a review. In short, the proposed system (ICF++) will measure the honesty value of a review, the trustiness value of the reviewers, and the reliability value of a product. The honesty value of a review will be measured by utilizing text-mining and opinion-mining techniques. The result from the experiment shows that the proposed system has a better accuracy compared with the result from the iterative computation framework (ICF) method.

In [8] this paper, mangoes are graded in four types Green Mango, Yellow Mango, and Red Mango which are based on the machine learning method. This system considers RGB values size and shape of mangoes. The following analysis is used to obtain good probability. This helps to train the system to identify the appropriate maturity of mangoes. This research is conducted on two machine learning methods i.e. Naive Byes and SVM (Support Vector Machine).

III. PROPOSED SYSTEM

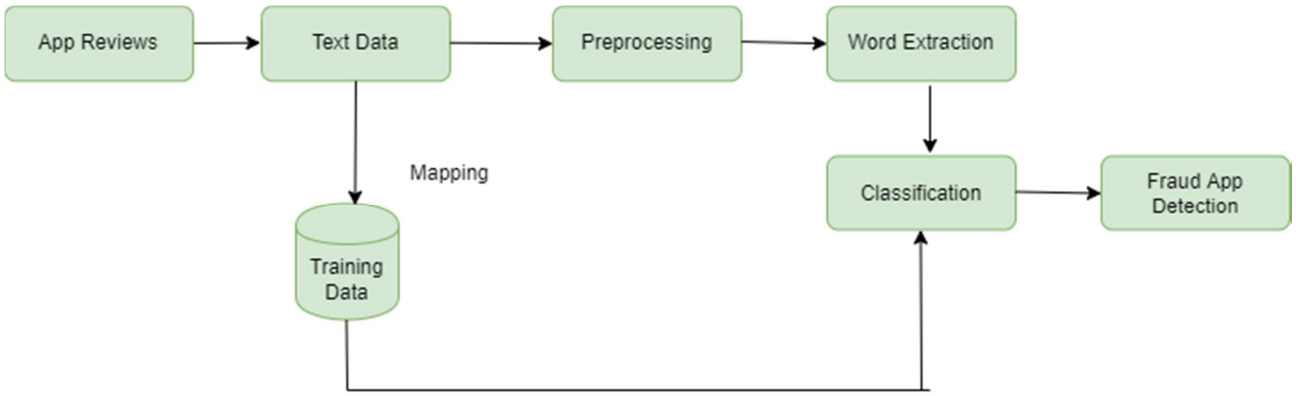


Figure 1. System Architecture

To detect fraudulent apps through sentiment analysis and machine learning:

1) *Data Collection:* Acquire a diverse dataset of app reviews and descriptions from app marketplaces.
2) *Data Preprocessing:* Clean and preprocess data by removing irrelevant content, tokenize, and standardize text.
3) *Labeling:* Annotate data to distinguish between legitimate and potentially fraudulent apps.
4) *Word Extraction:* Use text analysis to extract important words and expressions.
5) *Machine Learning Models:* Choose or develop models like logistic regression or Naive Bayes for classification.
6) *Training and Testing:* Split the dataset, train the model with training data, and evaluate with testing data.
7) *Feature Importance:* Investigate significant features to understand what contributes to fraud detection.
8) *Model Fine-Tuning:* Refine the model by adjusting parameters or exploring alternative algorithms.
9) *Deployment:* Deploy the model to a production environment for automatic app analysis.
10) *Continuous Monitoring:* Regularly update the model with new data and retrain to maintain effectiveness.

IV. CONCLUSION

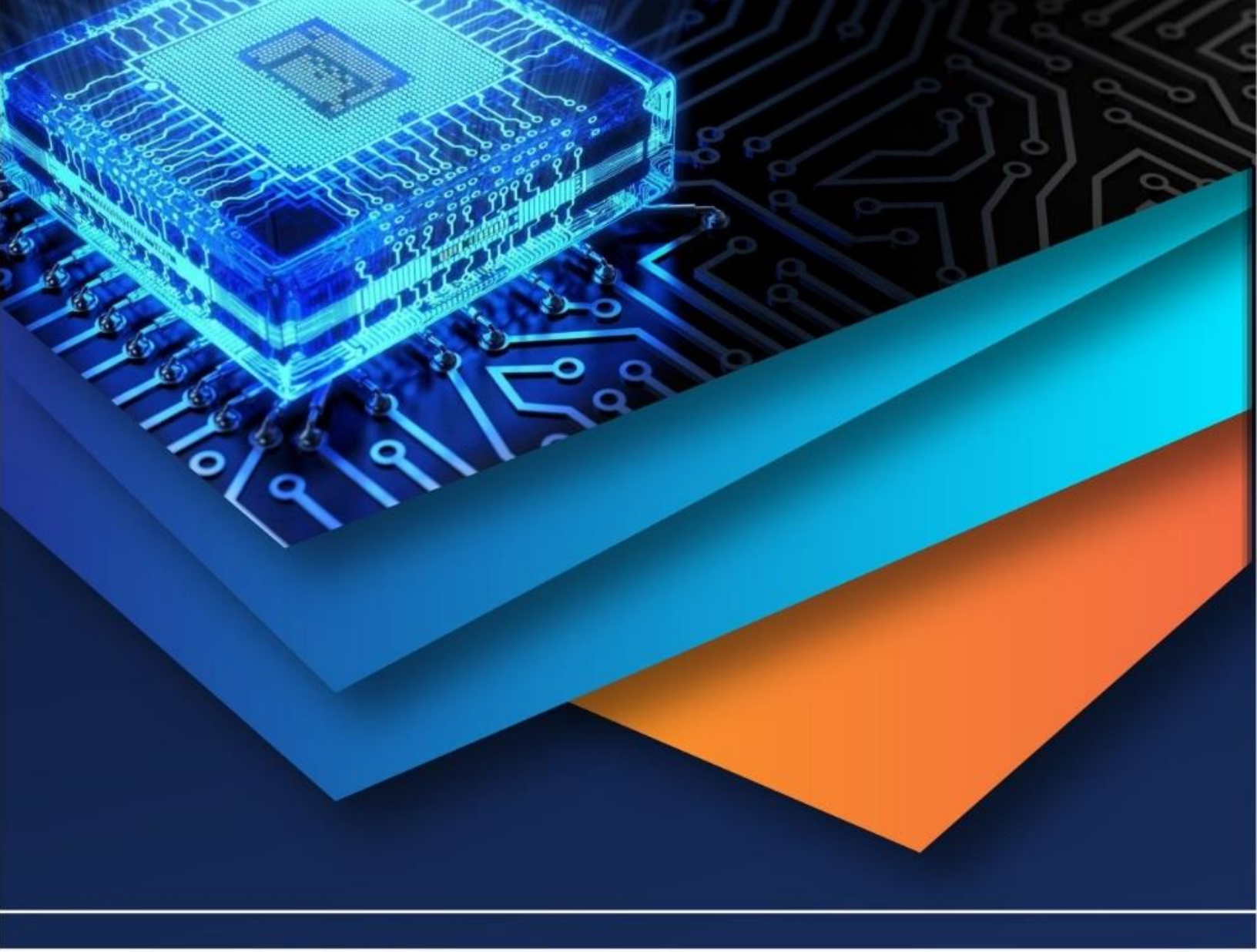
In this work, we developed a mobile app fraud detection system. To be more specific, we first showed how ranking fraud was caused by leading sessions and provided a method for mining leading sessions using sentiment analysis and spam detection from each App's historical ranking records. Then, based on a review, we found evidence for fraud-detection apps.

V. ACKNOWLEDGEMENT

We take this opportunity to thank our project guide Ms. Jaitee Bankar and Head of the Department Mr. Saurabh Parhad for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of RMD Sinhgad School of Engineering Pune for their valuable time, support, comments, suggestions, and persuasion. We would also like to thank the institute for providing the required facilities, Internet access, and important books.

REFERENCES

[1] Ch. Xu and J. Zhang," Combating product review spam campaigns via multiple heterogeneous pairwise features", In SIAM International Conference on Data Mining, 2014.
[2] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting business in reviews for review spammer detection", In ICWSM, 2013.
[3] A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, "True View: Harnessing the power of multiple review sites", In ACM WWW, 2015.
[4] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao," Spotting fake reviews via collective PU learning", In ICDM, 2014.
[5] M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa," Reducing Feature Set Explosion to Facilitate Real-World Review Spam Detection", In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.
[6] H. Xue, F. Li, H. Seo, and R. Pluretti," Trust-Aware Review Spam Detection", IEEE Trustcom/ISPA.,2015.
[7] E. D. Wahyuni, A. Djunaidy," Fake Review Detection from a ProductReview Using Modified Method of Iterative Computation Framework", In Proceeding MATEC Web of Conferences, 2016.
[8] G.D. Upadhye, D.Pise, "Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques", IEEE International conference on smart city and Emerging Technology,2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY