iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49724

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 11 Issue III Mar 2023- Available at www.ijraset.com

Fraud Apps Detection Using Sentiment Analysis and Spam Filtering

Prof. S. P. Vanjari[1], Priyanka Rekhawar[5], Ketki Shinde[2], Sakshi Shinde[3], Prajkta Shelke[4]
Department of Information Technology, Smt. Kashibai Navale College of Engineering

Abstract: ***In the mobile app industry, ranking fraud is the practice of engaging in dishonest or deceitful behavior with the intention of artificially boosting an App's position on a popularity list. In fact, ranking fraud by app developers is becoming more and more common. These practices include inflating their apps' sales or uploading fake app reviews. Although the significance of preventing ranking fraud has long been understood, little knowledge and research have been done in this field. In order to do this, we present a comprehensive analysis of fraud app detection using sentiment analysis and spam filtering in this study and suggest a system for detecting it in mobile apps.***

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

I. INTRODUCTION

Over the past few years, the number of smartphone apps has increased at an astounding rate. For instance, the Apple App Store and Google Play each had more than 1.6 million Apps available as of the end of April 2013. Many App shops created daily App leaderboards, which show the chart rankings of the most popular Apps, to encourage the creation of mobile Apps. The App leaderboard is undoubtedly one of the most crucial tools for promoting mobile apps. As a result, in order to have their apps rated as highly as possible in such App leaderboards, App developers frequently investigate various strategies, such as advertising campaigns, to promote their apps. However, as a recent trend, unethical App developers turn to various fraudulent techniques to purposefully raise their Apps and ultimately influence the chart positions on an App store rather than depending on conventional marketing strategies. This is typically accomplished by deploying "bot farms" to quickly inflate the number of App downloads, ratings, and reviews.

II. RELATED WORK

In [1] paper, Users increasingly rely on crowd-sourced information, such as reviews on Yelp and Amazon, liked posts and ads on and compromised accounts, and collusion networks. Existing approaches to detect such behavior relies mostly on supervised (or semi-supervised) learning over known (or hypothesized) attacks. They are unable to detect attacks missed by the operator while labeling, or when the attacker changes strategy. For the detection of fraud apps, sentiment analysis is used.

In [2] paper, Spam campaigns spotted in popular product review websites (e.g., Google Play Store) 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 [3] 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. Several approaches have been proposed to deal with the problem in the past few years. This work, take a different approach, which exploits the burrstones nature of reviews to identify review spammers.

In [4] paper, Online reviews on apps 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 and focus on reviews and use more than 15 million reviews from more than 3.5 million users.

In [5] 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. Most existing supervised learning approaches 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 [6] paper, Online reviews are quickly becoming one of the most important sources of information for consumers on various apps 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 apps and services or smear 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 [7] paper, it is providing 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 [8] paper, to detect fake reviews for an app by 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 an app. The honesty value of a review will be measured. The result from the experiment shows that the proposed system has better accuracy compared with the result from the iterative computation framework (ICF) method.

In the [9] paper, Online Social Networks (OSNs), which capture the structure and dynamics of person-to-person and person-to-technology interaction, are being used for various purposes such as business, education, telemarketing, medicine, and entertainment. This technology also opens the door to unlawful activities. Detecting anomalies, in this new perspective of social life that articulates and reflects the off-line relationships, is an important factor as they could be a sign of a significant problem or carrying useful information for the analyzer.

In [10] paper, they propose a new holistic approach called SpEagle that utilizes clues from all metadata (text, timestamp, and rating) as well as relational data (network), and harness them collectively under a unified system to spot suspicious users and reviews, as well as products targeted by spam. SpEagle employs a review-network-based classification task that accepts prior knowledge on the class distribution of the nodes, estimated from metadata. It is extremely efficient.

The paper study can be summarized into a table which can be grouped into various categories for analysing whether the app is fraudulent or not.

Sr.no	Paper name	Methodology	Algorithms	Efficiency
1.	Sentiment analysis on online app review.	Machine learning NLP	K means cluster	89%
2.	Evaluation of features on sentimental analysis	Support vector machine	Porter stemming algo	82%
3.	Sentiment analysis and complex natural language,	NLP Sentiment ala	Naive Bayes	90%
4.	Ranking Fraud Detection For Mobile Apps App Reviews	Data Mining, Natural Language Processing, Sentiment Analysis	Permutation-based model, score-based, Dumpster Shafer Rules, Gaussian Approximation	Not Specified
5.	An Implementation to Detect Fraud App Using Fuzzy Logic	Fuzzy Logic	Tokenization, Fuzzy logic algorithm, Ontology	83.75%
6.	Information Extraction for Mobile Application User Review.	Filtering Content Classification Topic Modelling Sentiment Analysis	SVM Logistic Regression Non-Negative Matrix Latent Dirichlet allocation	80.5%

III. EXISTING SYSTEM

A lot of work has been done in this field because of its extensive usage and applications. In this section, some of the approaches which have been implemented to achieve the same purpose are mentioned. These works are majorly differentiated by the algorithm for fraud app detection.

IV. PROPOSED SYSTEM

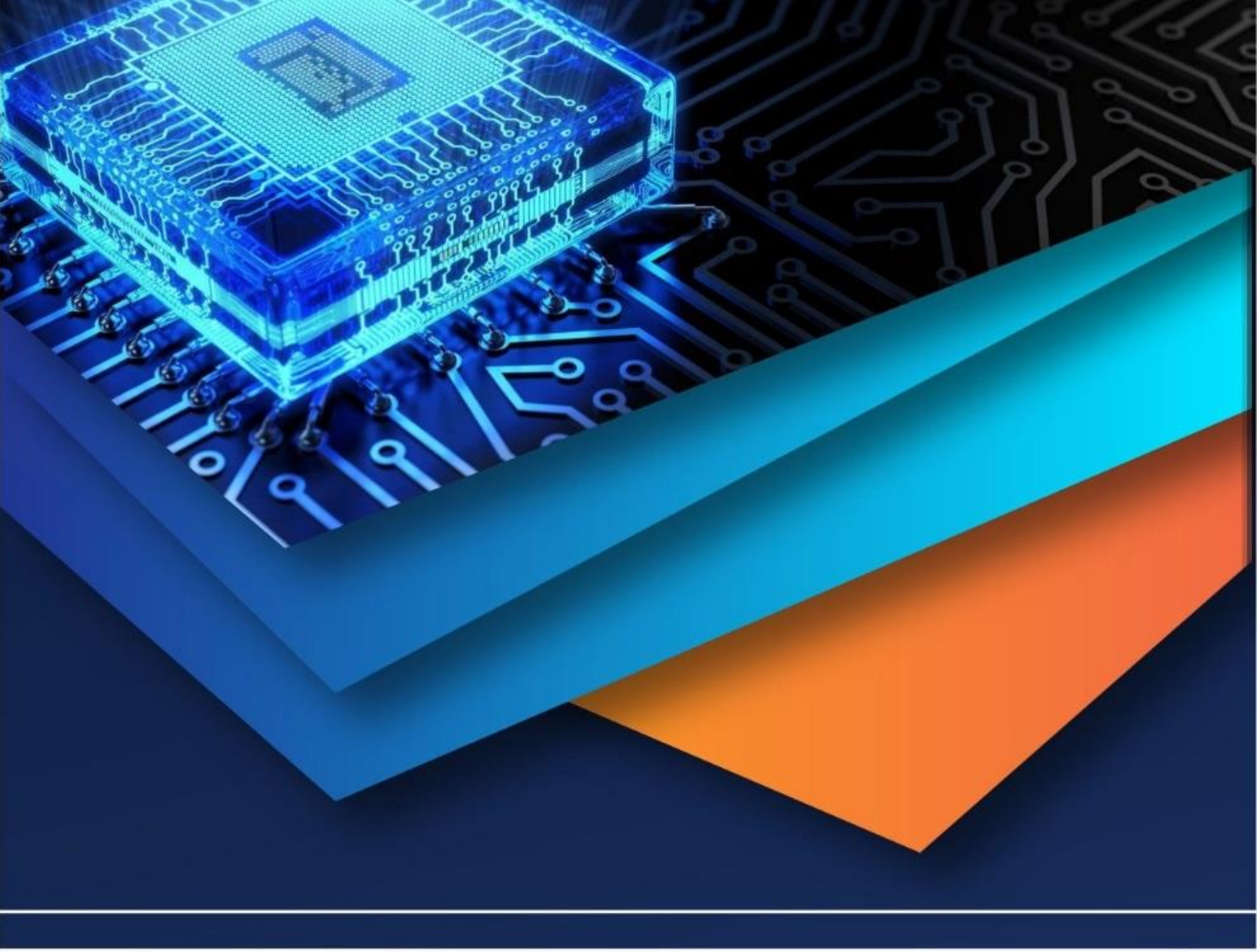
A Rating and Review Processing Method has been introduced to find the overall score of reviews for spam detection. The proposed method uses some parameters for spam detection, and these parameters show the variation of a particular review from other, thus increasing the probability of it being spam. This approach has been proposed which classifies a review as helpful or non-helpful depending on the score assigned to the review.

V. CONCLUSION

In this study, we created a fraud app detection system for mobile apps. In more detail, we first demonstrated how leading sessions were the source of ranking fraud and offered a technique for mining leading sessions from each App's historical ranking records using sentiment analysis and spam detection. Then, for detecting fraud apps, we identified evidence based on review.

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] B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks", In USENIX, 2014.
[5] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao," Spotting fake reviews via collective PU learning", In ICDM, 2014.
[6] M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa," Reducing Feature Set Explosion to Faciliate Real-World Review Sapm Detection", In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.
[7] H. Xue, F. Li, H. Seo, and R. Pluretti," Trust-Aware Review Spam Detection",IEEE Trustcom/ISPA.,2015.
[8] 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.
[9] R. Hassanzadeh," Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic", Queensland University of Technology, Nov, 2014.
[10] R. Shebuti, L. Akoglu," Collective opinion spam detection: bridging review networks and metadata", In ACM KDD, 2015.
[11] 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