FRAUD DETECTION IN MOBILE RANKING WITH MACHINE LEARNING TECHNIQUES

MUKTHA RAMESH KUMAR
Research Scholar
MUIT
Lucknow

Dr. MANISH VASHNEY
Professor
MUIT
Lucknow

Abstract

Ranking fraud in the mobile App market refers to fraudulent or deceptive activities which have a purpose of bumping up the Apps in the popularity list. Indeed, it becomes more and more frequent for App develops to use shady means, such as inflating their Apps' sales or posting phony App ratings, to commit ranking fraud. While the importance of preventing ranking fraud has been widely recognized, there is limited understanding and research in this area. To this end, in this paper, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps. Specifically, we investigate two types of evidences, ranking based evidences and rating based evidences, by modeling Apps' ranking and rating behaviors through statistical hypotheses tests. In addition, we propose an optimization based aggregation method to integrate all the evidences for fraud detection. The mobile industry is growing rapidly, subsequently the number of mobile apps coming in the market is also increasing. Google Play and Apple are app stores where apps can be downloaded by either paying or they can be downloaded free of cost. Rank of an app is important, as high ranked apps get noticed by customers more easily than those ranked low. In order to get a high rank some developers are resorting to fraudulent means in order to increase their app's rank. Hence a ranking fraud detection system is required to detect rankings obtained through fraudulent means. We design and develop a ranking fraud detection system to detect fraud ranks.

Keywords: *Ranking fraud detection, Reviews, Ranking, Data mining, Mobile applications.*

INTRODUCTION

Mobile users are increasing day by day and with the invention of applications in smartphones made lives easy. Many applications make our daily lives easy such as bank payments, online shopping applications etc., The increase in the usage of such applications made many different applications. Some applications are good applications and some of them are fake. There is no harm in installing good applications where as the fraud applications may reduce the performance in the mobile or delete some important data. These kinds of applications are made by fraud application makers and create fraud ratings for those applications. So, there is a vital need of detecting such fraud ratings. The importance of preventing ranking fraud has been widely recognized, there is limited understanding and research in this area. To this end, in this system, we provide a holistic view of ranking fraud and propose a ranking fraud detection system for mobile Apps. This system conducts sentimental analysis on the reviews of the mobile app and it calculate the number of positive and negative reviews, fetch the ranking of the mobile app, do a ranking analysis and determine whether there is any ranking fraud happen or not. This system uses machine learning technique by which positive and negative new words are appended automatically to the keyword database. Furthermore, IP address blocking facility is added to the system.

LITERATURE REVIEW

Izhar Alam et al (2021) In today's era of technology, especially in the Internet commerce and banking, the transactions done by the Mastercards have been increasing rapidly. The card becomes the highly useable equipment for Internet shopping. Such demanding and inflation rate causes a considerable damage and enhancement in fraud cases also. It is very much necessary to stop the fraud transactions because it impacts on financial conditions over time the anomaly detection is having some important application to detect the fraud detection. A novel framework which integrates Spark with a deep learning approach is proposed in this work. This work also implements different machine learning techniques for detection of fraudulent like random forest, SVM, logistic regression, decision tree, and KNN. Comparative analysis is done by using various parameters. More than 96% accuracy was obtained for both training and testing datasets. The existing system like Cardwatch, web service-based fraud detection, needs labelled data for both genuine and fraudulent transactions.

V. Visvanathan, Dr. R. Gobinath., and Dr. S. Perumal (2019) worldwide around us making related with a couple of advances, thusly the utilization of Smartphone a tiny bit at a time augments. In this circumstance we utilize the application in play store sponsored by Google which are invigorated for ordinary use. By the explanation behind use, examples to make the applications predominant through android customers' .This will prompts favored point of view of misleading fashioners to make an exact of authentic. Those application will be present by customer and recognizes all the basic assents requested by google .In Existing k-suggests Clustering estimation is used to describe the far reaching instructive record. In any case, it is baffling to find malware application in uncertain condition. In this Proposed System, Avoiding Users from the harmful application and recognizing the beguiling application using PlayFair Technique, It perceives suspicious, time related co-review practices of Application. FairPlay compares overview practices and phenomenally joins distinguished study relations with semantic and social PCF (Pseudo Clique Finder) estimation that mishandles the phony or malware application.

AratiTule, Rahul Shahane (2017) As we all know every person in the world are mobile users in fact smart-phone users with android applications. So, Due to this popularity and well-known concept there will be a rapid growth in mobile technology we have seen. As well as in data mining concept mining the needed data from a particular application is very difficult and crucial task. Merging these two concepts of ranking frauds in android market and mining needed data is gone very tough for us and this is challenging situation. We are using these concepts in whole paper. As we know that the mobile Apps has grown at vast speed in some years; as for march 2017, there are nearby 2.8 million Apps at google play and 2.2 Apps at Apple Apps store. In addition, there are over 400,000 independent app developers all fighting for the attention of the same potential customers. The Apple App Store saw 128,000 new business apps alone in 2014 and the mobile gaming category alone has competition to the tune of almost 300,000 apps. Here the main

need to make fraud search in Apps is by searching the high ranked applications up to 30-40 which may be ranked high in some days or the applications which are in those high ranked lists should be verified but this is not applied when we work for thousands of applications added per day. So, we go for broad view by applying some technique to every application to judge its ranking.

Fraud Detection

Fraud Detection FRDDT (FRDDT) is a set of activities undertaken to prevent money or Using an infinite and climbing quantity of manners someone could perpetrate fraud, and detection might be hard to do. Activities like reorganization, downsizing, shifting into brand fresh information methods or falling upon a cyber-security violation may hamper a business's potential to find fraud. What this means is processes like real-time observation for frauds will be suggested. Businesses must Start Looking for fraud in fiscal trades, Spot, apparatus utilized, pioneered periods along with authentication methods Land from getting accessed by way of false pretences. FRDDT is employed to a lot of businesses like insurance or banking. In fraud could consist of things like exceeding tests or employing stolen MOBAPPs. Various other designs of fraud could demand incurring losses or inducing an injury using the only purpose for your own pay-out.

FRDDT techniques

Data fitting Regression Examination Probability models and loopholes. Calculating statistical parameters Statistical data investigation methods Incorporate the usage of: Fraud is on average a procedure that involves quite a few replicated processes; which makes hunting to get routines that an overall attention for FRDDT. By way of instance, info analysts may stop insurance fraud from making calculations to discover patterns and anomalies. FRDDT could be divided by using statistical information investigation methods or synthetic intelligence (AI)

AI techniques used to detect fraud include the use of:

Data-mining - That may categorize section and group info to hunt up to countless of trades to locate designs and discover fraud.

Neural systems - That may find suspicious appearing routines, and use these routines to find them farther.

Machine-learning which may mechanically identify faculties in fraud.

Design recognition- That may discover courses, structures and patterns of questionable behaviour

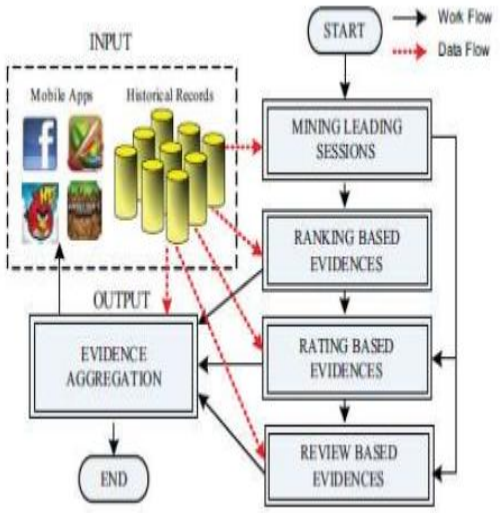

Fig -1: Existing model for ranking fraud detection for mobile apps

Types of fraud

Fraud could be committed in many of distinct methods as well as in many of distinct preferences. As an instance, fraud might be dedicated to banking, insurance, govt as well as also in healthcare industries.

One common kind of fraud on planet is purchaser accounts take over, exactly wherever by somebody gains access into your victim's banking accounts utilizing spiders. Other cases of fraud in banking range from things like the usage of malicious software, using anonymous identities, including money-laundering, MOBAPP fraud along with fraud that is mobile.

MOBAPP Fraud:

Fraud is also cyclical, based Scalar, the corporation supporting the report. Fraudsters and mobile attribution businesses are always participated in a arms race whilst the terrible guys come across new techniques to deceive the device, and also the fantastic guys locate fresh tactics to prevent them. Shifting A D pricing types additionally impression fraud prices. 1 thing is for certain: once you believe you've removed fraud, then that is when fraud is the most hazardous.

Machine learning and data mining in Accessing Fraud Techniques

The system learning and artificial intelligence answers could be categorized to two classes: 'Assessing' and also 'unsupervised' finding out. These processes hunt for reports, clients, providers, etc., act 'remarkably' in sequence to lead sensing scores, and visual or rules anomalies, based upon your own procedure. Historical statistics investigation processes were aimed toward attracting statistical and qualitative data traits. These processes ease helpful data mark-up and certainly will help get far superior insights into the procedures supporting this info. Even though conventional data investigation methods may lead us into comprehension, it's even now developed by individual analysts. To proceed over and above, a statistics analysis approach must be armed with a significant sum of history expertise, as a way to do manual jobs requiring this comprehension and also the data offered. Within an attempt to satisfy this aim, scientists have switched into notions in machine learning discipline. This really can be an all pure way to obtain thoughts, as the system learning Endeavour is clarified as spinning background wisdom and cases in to comprehension (output).

Machine Learning and Artificial Intelligence in Fraud

If completed precisely, machine-learning may definitely distinguish valid and deceptive behaviours even though adapting with the years and energy to fresh, previously concealed fraud approaches. This really may gotten very complex since there's a requirement to translate blueprints from the info and employ info science to always enhance the capacity to differentiate ordinary behavior from strange behavior. This necessitates tens of thousands of computations to be precisely achieved in milliseconds. With no suitable comprehension of the realm, in addition to fraud-specific info science methods, it's possible to readily implement machine learning algorithms which know the incorrect item, leading to an expensive mistake that's tough to relax. As People Are Able to find bad habits, so also may a badly architected Machine-learning version Due to fact organized offense strategies are really complex and easy to accommodate, protection plans predicated on almost any individual, one-size-fits-all analytical procedure will create subpar outcomes. Every use case ought to be encouraged by finely crafted embryo detection methods which can be optimal for issue available. Like a consequence,

the two supervised and unsupervised variations play essential functions in FRDDT and needs to be squeezed to thorough, future - based fraud plans. Machine-learning will help data boffins economically determine that trades are likely to become deceitful, even though reducing false positives.

RESULTS AND DISCUSSION

To acquire an even broader in sight in to the next type of fraud that is download, we now run a comparative investigation and routine together using all imitation downloads also questionable Programs. Together with all the XGBoost classifier previously, we've identified a lot greater than just one hundred questionable Programs one of all of programs while inside the company's application current market in June 2018 on December 2018.

Find indicates that the category supply of fraud Programs comparing with all an total category supply one of most of Programs around the application industry. Especially, Finance and video game Apps accounts fully for over 1 / 2 of those questionable Programs, whereas these 2 forms of Programs just simply use upto 15% one of all of Programs while inside the Program industry. We believe Sport and Finance applications possess significantly more possibility for monetizing users, which creates a large total of earnings.

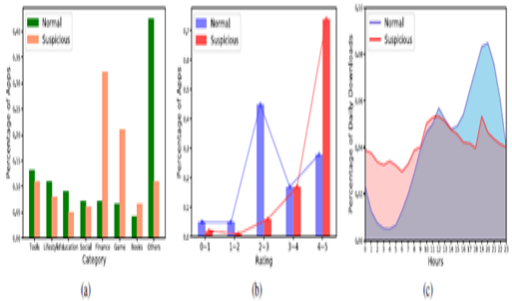

Figure 2: Download of apps and analysis

Faculties of questionable Programs and downloads comparing normal downloads and Apps. (a) Program type supply of questionable Programs versus ordinary Programs; (b) evaluation supply of questionable Programs versus ordinary Programs; (do) questionable downloads targeted visitors versus ordinary downloads targeted site visitors every day.

In addition, in our research we discovered the pre culturing effects of every coaching data collection are quite much alike, hence we place the number of recognition conditions since 13 for many five PHMMs from the Top-Free 300 information collection along with 12 for many five PHMMs from the Top-Paid 300 information group, respectively. Really, the cross endorsement may fail the consequence of temporality of version coaching. But because of this restriction of the information collection, the information loopholes in just two information collections are unbalanced, it's difficult to find appropriate time stamps for correctly devoting test and training information. Therefore, here we utilize the trusted cross analysis for assessing the total functionality of prevalence forecast. To your very best knowledge, there isn't any present job of Program Development modeling was claimed. Therefore, here we now developed two advanced baselines for assessing our PHMM, that can be static and sequential approaches, respectively.

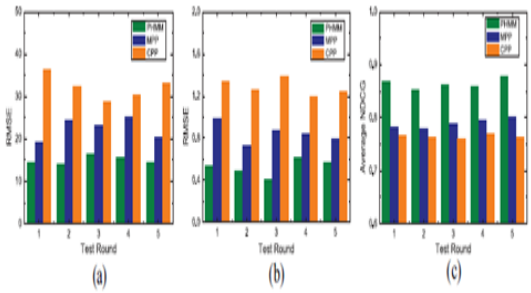

Figure 3 : sequential approaches analysis.

Performances of predicting popularity observations by each approach in the Top-Free 300 data set. (a) Ranking prediction. (b) Rating prediction. (c) Topic prediction.

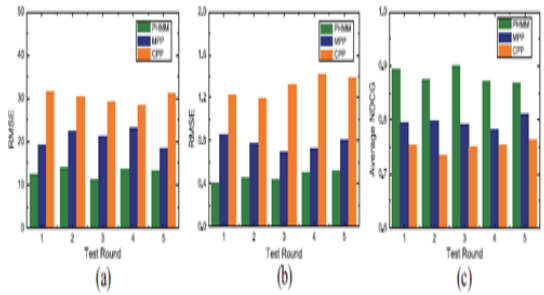

Figure 4: Ranking prediction, Rating prediction analysis.

Performances of predicting popularity observations by each approach in the Top-Paid 300 data set. (a) Ranking prediction. (b) Rating prediction. (c) Topic prediction.

CONCLUSION:

This project presented a Ranking Fraud Detection system for mobile application. A user who has an account can check for rank fraud in an application. Users can upload a list of applications for which fraud is to be detected as well as reviews of those apps. Perceptual analysis can be done for the reviews of an app, which analysis if a review is positive, negative or neutral. Based on perception analysis and the ratings of the app, the system can detect ranking fraud. Here developed a ranking fraud detection system for mobile Apps. Specifically, here first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, here identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Moreover, here proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps.

REFERENCES

1. *Izhar Alam et al, "An Enhanced Secure Deep Learning Algorithm for Fraud Detection in Wireless Communication", Wireless Communications and Mobile Computing, vol. 2021, Article ID 6079582, 14 pages, 2021. https://doi.org/10.1155/2021/6079582.*
2. *AratiTule, Rahul Shahane, Discovery of Ranking Fraud for Mobile Applications, International Journal of Engineering Science and Computing, ISSN 2319-7242 , Volume 7 Issue No.5, 2017.*
3. *V. Visvanathan, Dr. R. Gobinath., and Dr. S. Perumal, Optimized Ranking Based Search For Fraud Detection In Google Play, Journal of Applied Science and Computations, ISSN NO: 1076-5131, Volume VI, Issue II, February/2019.*
4. *Bravenboer, Martin, EelcoDolstra, and Eelco Visser, "Preventing Injection Attacks with Syntax Embeddings", 6th ACM International Conference on Generative Programming and Component Engineering, pp. 3-12, 2007.*
5. *Bandhakavi, S., Bisht, P., Madhusudan, P., and Venkatakrishnan, V. N. "CANDID: Preventing SQL Injection attacks using dynamic candidate evaluations", In Proceedings of the 14th ACM conference on Computer and communications security, ACM, pp. 12-24, 2007.*
6. *Datti, R. and Verma, B. (2010). Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis.*

7. *Ezumalai, R., and Aghila, G. "Combinatorial Approach for Preventing SQL Injection Attacks", IEEE International Advance Computing Conference, Patiala, India, pp.1212-1217, 2009.*
8. *Gould, C., Su, Z., and Devanbu, P. "JDBC Checker: A Static Analysis Tool for SQL/JDBC Applications", In Proceedings of the 26th International Conference on Software Engineering, pp.697-698, 2004.*
9. *Hallaraker, O. and Vigna, G. "Detecting malicious JavaScript code in mozilla", In Proceedings 10th IEEE International Conference on Engineering of Complex Computer Systems, IEEE, pp. 85-94, 2005.*
10. *Jovanovic, N., Kruegel, C., and Kirda, E. Precise alias analysis for static detection of web application vulnerabilities. 2006 Workshop on Programming Languages and Analysis for Security, pp. 27–36, USA, 2006.*