Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 140 (2020) 54182-54185

ISSN: 2229-712X

Detection of Fraud Application Using Machine Learning

Kanchan Mahajan, Pushpak Mutha, Mahima Naik, Shubham Deore and Prapti Patil
Department of Computer Engineering Sandip Foundation's, Sandip Institute of Technology & Research Centre, Mahiravani, Nashik – 422213

ARTICLE INFO

Article history:
Received: 28 February 2020;
Received in revised form:
6 March 2020;
Accepted: 16 March 2020;

Keywords
NLP
Machine Learning
Web Application
Python

ABSTRACT

With the increase in the number of mobile applications in the day to day life, essential thing is to keep track as to which ones are safe and which ones aren't. One can't judge how safe and true each application is based only on the reviews that are mentioned for each application. Hence it is a need to keep track and develop a system to make sure the applications present are genuine or not. The objective is to develop a system in detecting fraud applications before the user downloads by using machine learning. Sentimental analysis is to help in determining the emotional tones behind words which are expressed in online. This method is useful in monitoring social media and helps to get a brief idea of the public's opinion on certain issues. The user cannot always get correct or true reviews about the product on the internet. We can check for user's sentimental comments on multiple application. The reviews may be fake or genuine. Analyzing the rating and reviews together involving both user and admins comments, we can determine whether the application is genuine or not. Using machine learning, the machine is able to learn and analyze the sentiments, emotions about reviews and other texts. The manipulation of review is the key aspects of Application ranking fraud. By using Machine Learning, analyzing reviews and comments can help to determine the correct application for both Android and iOS platforms.

Introduction

With the growth in technology, there is an increase in the usage of mobiles. There has been a vast growth in the development of various mobile applications on numerous platforms such as the popular Android and iOS. Due to its rapid growth day by day for its everyday usage, sales and developments, it has become a significant challenge in the world of the business intelligence . Which can gives rise in the market competition. The companies and application developers are having a tough competition with one another in order to prove their quality of product and spend an immense amount of work into attracting customers to sustain their future progress. The most important role that plays is the customers ranking, ratings and reviews on that specific application which they happen to download. This could be a way for the developers to find their weakness and enhance into the development of a new one keeping in mind the peoples need. Not only that ,certain times guile developers misleadingly the recognition of their applications or malicious ones use it as a platform to spread malware throughout. As an ongoing pattern, rather than depending on customary promoting arrangements, under the trees Application developer's option in contrast to some false way to intentionally support their Applications and in the long run controls the outline rankings on an App store. This is generally executed by utilizing so-called "bot ranches" or "human water armed forces" to expand the Application downloads evaluations and audits in an exceptionally brief time. Certain times, just for the up liftment of the developers, they tend to hire teams of workers who commit to fraud collectively and provide false comments and ratings over an application. This is known to be termed as crowd turfing. Hence it is important that before installing an application, the users are provided with proper and genuine comments to avoid certain mishaps. For this, an automated solution is required to overcome and systematically analyse the various comments and ratings that are provided for each application. With mobile phones being a quite popular need, it is essential that suspicious applications must be marked as fraud in order to be identified by the store users. It will be difficult for the user to determine the comments that they scroll past or the ratings they see is a scam or a genuine one for their benefit. Thereby, we are proposing a system which will identify such fraudulent applications on Play or App store by providing a entire view of rank base fraud detection system. By considering sentiment analysis, we can get a real reviews has highest probability and hence we propose a system that intakes reviews from registered users for a single productor multiple and evaluate them as a positive or negative rating . It is also useful to determine the fraud application and ensure mobile security as well. We initiate the system by considering the mining leading session or also the active periods of the applications. This influences in detecting local anomaly than the global anomaly of the application ranking. Initially we propose a basic yet fruitful calculation to recognize each leading sessions of Application which is dependent on its authentic positioning records. At this point, of the investigation of Applications positioning practices, it finds the fake Applications

Tele:
E-mail address:pushpakmutha@gmail.com

that regularly have distinctive positioning examples in each driving session contrasted and ordinary Applications. Furthermore, we inspect through different types of evidences namely ranking based, rating based, and review based by modeling the consolidation of the three through statistical hypotheses tests. Regardless, the positioning-based evidences can be influenced by the Application developer's status and some genuine advertising efforts such as the "constrained-time markdown". Thus, it is inadequate only consider the rank-based confirmations. Along with this, the proposed system introduces two sorts of extortion evidences dependent on Applications rating and survey history which mirror some unique patterns from Applications. Also, an aggregated method is utilized for the collection of all the evidences that are necessary for detecting fraud. In order to do so, we evaluate the proposed system by using real-world application data collected from Google play store and iOS app store which is present from a large duration of time.

Related work

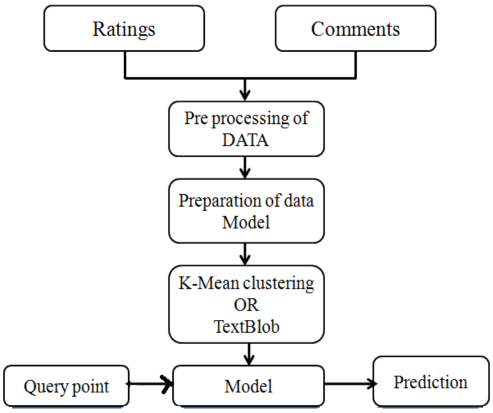
HuiXiong[1] found framework for the location of positioning extortion for versatile Apps however it is still under examination look into. To sick this pivotal need, we propose to build up a positioning misrepresentation identification framework for portable Apps. We additionally decide a few significant challenges. First Challenge, within the entire life cycle of an App, the positioning misrepresentation doesn't generally occur, so we'd prefer to recognize when extortion occurs. At long last, because of the dynamic idea of graph rankings, it's hard to search out and check the confirmations identified with positioning misrepresentation, which spurs us to get some understood extortion examples of portable Apps as confirmations. Y. Ge, H. Xiong has proposed taxi driving misrepresentation location framework for Advances in GPS following development have en-abled us to present GPS reference points in city taxicabs to accumulate a lot of GPS follows under operational time necessities. These GPS follows give unmatched opportunities to us to uncover taxi driving blackmail works out. In this paper, increment a taxi driving distortion distinguishing proof structure, which may effectively investigate taxi driving extortion. In this system, first offer abilities to find two pieces of affirmations: travel course evidence and driving division verification. Also, a third limit is expected to joint the two pieces of evidences in perspective on Dempster-Shafer theory. They propose effective online connection spam and term spam identification strategies utilizing spamicity. This techniques needn't bother with preparing and furthermore financially savvy. A genuine informational collection is utilized to assess the adequacy and the productivity [2]. Numerous tricky practices happen in understood Android application showcase i.e., Google Play Store. Along these lines, to identify malware, already the work just centered around authorization examination and application executable. Despite the fact that, Mahmudur Rahmanet. al., presented FairPlay, a framework which recognized and utilized follow desert by false to detect the malware and furthermore the applications exposed to look rank misrepresentation [3]. FairPlay associated audit movement and distinguished their connection with phonetic and social signals that are accumulated from Google Play application information to recognize the suspicious applications. Step by step utilization of versatile has expanded. Likewise to get to a wide range of versatile application, the portable clients like better to utilize cell phones. Clients are the important thing who download versatile applications relying on what rate clients do as of now have downloaded that application?, what are its appraisals and audits? , what are the remarks? and so on. Extortion positioning inside the application showcase demonstrates bogus or wrong deeds which may have motivation to push up applications on the acknowledgment list. Most application engineers use misrepresentation intends to build their application's deals by advising bogus evaluations of the applications, and doing positioning misrepresentation. Just as, Varsha A. Patilet. al., exhibited take a shot at assessment of research on emojis which is a collection of images speaking to various faces in content based correspondence [4]. Be that as it may, in paper Josh Jia Ching Ying et. al., proposed a very successful misrepresentation telephone call recognition approach called parallelized diagram mining, to be specific Fraud Detector. It consequently marked misleading telephone numbers by the tag named ―fraud‖, in order to separate the phony telephone call numbers from the certified ones. It additionally utilized Hyperlink-Induced Topic Search (HITS) calculation and a novel conglomeration approach. In paper [6] Navdeep Singh et. al., introduced a streamlining based conglomeration way to deal with fuse the confirmation in order to break down the limit of primary specific time range from portable application. In spite of the fact that, Hengshu Zhu et. al., exhibited a far reaching approach for positioning extortion and furthermore for distinguishing positioning misrepresentation in the portable applications. Right off the bat, a functioning period has been mined precisely to find the positioning fraud.[6]Hengshu Zhu, HuiXiong proposed a positioning misrepresentation discovery framework for android versatile applications. In this positioning misrepresentation occurred in essential sessions for each application from its past positioning records. At that point, they recognized positioning based, rating based and audit based confirmations for discovering positioning extortion. Moreover they proposed a streamlining based collection strategy to consolidate each of the confirmations for assessing the responsibility of driving sessions from versatile applications. PranjaliDeshmukh, PankajAgarkar ―Mobile Application For Malware Detection In this paper creator proposed techniques for assessment of examination and configuration example of android applications dependent on distributed computing and information mining. In this paper creators created instrument ASEF and SAAF for android applications to achive security. In this creators depict a technique that performs applications security and give easy to use interface on a cell phone Anuja A. Kadam ,Pushpanjali M. Chouragade ―A Review Paper on: Malicious Application Detection in Android System creators give a precise report on the various strategies of pernicious application location in android mobiles. The examination of consent prompted chance in Android applications on an enormous scale in three levels .First upon rank all the individual authorizations regarding their plausible hazard with various strategies. At that point, arrange subsets of hazard consents. At that point utilizing a few calculations recognize the malapps supported the distinguished subsets of dangerous permissions. Muneer Ahmad Dar &Javed Parvez, In this paper creators portrays the confinements in the present Android security model in detail. In this paper creators gives the detail depiction of security necessities which require to be consider at the hour of planning security component for cell phones.

Problem Statement

The mobile industry is growing rapidly, subsequently the number of mobile applications coming in the market is also increasing. As there are many applications available in market users are confused while downloading the applications for their use. There are many fraud applications available in market So detect such fraud applications we develop a system based on Review i.e. Fraud Application Detection Using Machine Learning.

Proposed Work

The way we marching for our study follows the traditional data analysis steps, is shown in the Fig.



We have Dataset of application audits as appraisals and remarks which is given by various users. Ratings implies an arrangement as indicated by request or evaluation and Comments implies an announcement of certainty or supposition, particularly a comment that communicates an individual response .Dataset is in unstructured organization. The dataset incorporates insights regarding the diverse sort of remarks like positive ,negative likewise it incorporates evaluations like fortunate or unfortunate or moderate which is given by various clients who download the application and shared there understanding about the application.. Our point is to investigate the unstructured and unlabeled informational index utilizing AI procedure.

After assortment of dataset we perform ace handling of information on gathered information by utilizing Natural Language Processing. NLP is the computational strategies to investigation and combination of common language and discourse.

In the period of readiness of information model we prepare the information for tasks which should be finished with information accessible. Content Blob is a Python library for preparing printed information. It gives a basic API to plunging into normal common language preparing (NLP) undertakings, for example, grammatical feature labeling, thing phrase extraction, slant investigation, grouping, interpretation and so on. It is utilized in our product to get the real likelihood of the application which is Fraud or Genuine for the clients.

Model comprise of the consequence of specific tasks made by our product in which entire activities are finished to judge. In the Query point we give the comparative kind of use for the examination and their insights which are performed before. For getting aftereffect of that question point we give this as a contribution to the model. Prediction is the last stage where we give the judgment to client whether the application is misrepresentation or genuine. To get this outcome we think of some as limit proportion implies on the off chance that we get above 80% positive remarks or rating, at that point it is consider as an authentic application. If we get result underneath 80% then it is think about an extortion application.

Outcome

It determining fraud applications by using the concept of machine learning and sentiment analysis. In the database we store collected data which is then evaluated with the supporting algorithms. This is a unique approach in which the evidences are aggregated and confined into a single result. The proposed structure is scalable and may be extended to another domain generated proofs for the ranking fraud detection.

Conclusion & Future Enhancement

This system will present about determining fraud applications by using the concept of machine learning and sentiment analysis. It was supported by the architecture diagram which briefed about the algorithm and processes which are implemented in the project. In the database we store collected data which is then evaluated with the supporting algorithms defined. This is a unique approach in which the evidences are aggregated and confined into a single result. The structure is scalable and can be expanded to another domain generated proofs for the ranking fraud detection. The experimental results in the effectiveness of the proposed system, the scalability of detection algorithm also as some regularity within the fraud actions. In the upcoming time , we decide to study simpler fraud proofs and consider the latent connection among rating, review and rankings. Moreover, we'll expand our ranking fraud detection approach with other mobile App related services, like different Apps recommendation, for enlarge user experience.

References

[1] H. Zhu, H. Xiong, S. Member, and Y. Ge, “For Mobile Apps discover the Ranking Fraud” no. March, 2015.

[2] Ranjitha.R, Mathumitha.K, Meena.S, S.Hariharan, “Discovery of Ranking of Fraud for Mobile Apps”, IJIREM ISSN: 23500557, Volume-3, Issue-3, May-2016.

[3] H. Zhu, H. Cao, E. Chen, H. Xiong, and J. Tian, “For the mobile app classification exploiting enriched contextual information,” in Proc. 21stACMInt. Conf. Inform. Knowl.Manage., 2012, pp. 1617–1621.

[4] Monali Zende, Aruna Gupta, “Survey of the Fraud Ranking in Mobile Apps”, IJSR Volume 5 Issue 2, February 2016.

[5] N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explor. Newsl., 13(2):50–64, May 2012.

[6] Sabbineni Poojitha, Balineni Venkata Sai Mrudula and Vemuri Sindhura, “A Novel Method To Identify False Apps Through Data Mining”, IJETCSE ISSN: 0976-1353 Volume 23 Issue 5 –SEPTEMBER 2016.

[7] A. Klementiev, D. Roth, and K. Small. Unsupervised rank aggregation with distance-based models. In Proceedings of the 25th international conference on Machine learning, ICML ’08, pages 472– 479, 2008.

[8] Mahmudur Rahman, Mizanur Rahman, Bogdan Carbunar and Duen Horng Chau,“FairPlay: Fraud and Malware Detection in Google Play”.

[9] Esteves, K.M.; Rong, C. Using Mahout for clustering Wikipedia’s. In the cloud campare between fuzzy c-means and k-means. Which is proceed in 2011 Third IEEE International Conference on Science, Cloud Computing

technology and IEEE Computer Society, Washington, DC, USA, 29 November– 1 December 2011; pp. 565–569.

[10] Huang, Z. Expandes the clustering algorithm like k-means algorithm for large data sets which has categorical values. Data Min. Knowl. Discov. 1998, 24, 283–304.

[11] Kanik ,Neha , Kirti, “Comparative Analysis of k-means and enhanced k-means clustering algorithm for data mining”, IJSER, Aug 2012, Vol-3,Issuse3.

[12] https://www.kaggle.com/mlgulb/creditcardfraud