Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

Application of Machine Learning on Fraud App Detection

[1]Syed Abdul Moeed, Assistant Professor, Department of CSE, KITS-Warangal, TS, India
[2]G.Ashmitha, Assistant Professor, Department of CSE, KITS-Warangal, TS, India
[3]Dr. P. Niranjan, Professor, Department of CSE, KITS-Warangal, TS, India

ABSTRACT

In today's world, everyone is using smart phones and those are very important in our daily life. The extensive distribution of mobile devices and applications across society has helped create counterfeit apps amongst the major cyber security threats of today. There are so many fraud applications are available in the internet. Fake behavior is most popular in application stores like Google play store and apple's application store. The growth of mobile apps was increased to 2.86 million at Google play store and makes the users in a fuzzy state while downloading the apps. There are many apps from which any app can be fraud, so the identification of true app is needed. Fraud apps basically deals with fake apps. So, our system will help the user to identify which application is true. In this paper we propose a method to detect the fraud application based on user reviews and ratings using Naive Bayes classifier. The user reviews can be collected from Google play store and classify the reviews into positive or negative by using sentimentanalysis.

Keywords: mobile apps, reviews, ratings, Sentiment analysis, Naive Bayes Classifier, NLP.

1. INTRODUCTION

As we all know that mobile application market is in rise as mobile users are in batch, sensible phone user uses those options of mobile apps as amusing purpose, knowledge purpose then on. Android phones became very fashionable lately and also users use the play store facility frequently. According other ecent study, the quantity of application in Google playstore, grew from 1million to3.5million.Whereas, the quantity of application in Apple's play store are 2.2 million. The review section of every application on the play store could also be an honest thanks to research an application. Fraud applications may cause damage to phone and also possibly data thefts. However, once it is installed, all mobile users want highly hierarchical applications. In order to transfer sensitive applications to the user of a phone he needs to visit the play store such as the Google Play Store,

the Apples store etc. Sometimes User doesn't believe the appliance whether the applications are helpful or useless. The user often sees that the downloaded software does not work . This relates to its mobile application list fraud. Hence such applications must be marked, so as that they are getting to be avoided by the play store users.

With technological advancement, the use of mobile phones is growing. The creation of different mobile apps on various platforms such as Android and IOS has expanded enormously. Because of its rapid growth, it has become a major challenge in the world of business information every day for its daily use, sales and development. This leads to competition in the market. Companies and application developers compete hard

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

to show the quality of their products and work hard to retain buyers to proceed with their future development.

The most important role is the ratings, ratings and reviews of the customer for the particular app. This could allow developers to find their shortcomings and boost the production of a new one that takes into account the needs of the people. Not only that, certain times guile developers misleadingly the recognition of their apps or malicious ones use it as a platform to spread malware throughout.

Fig .1 Fraud detection and prevention

In comparison to any opposite place to actively promote Applications, the Software developer's option for trees is a continuous pattern and does control the outline rankings in an App store instead of relying on the normal promotional arrangements. The Software Downloads assessments and audits in an extraordinary short span of time, usually by using "bot ranches" or "human water armed forces" .

Often they prefer to employ teams of staff jointly committing fraud and offering misconceptions and scores on an application only for the lifting up of developers. This is known as turfing in the crowd. Therefore it is always necessary to ensure that users receive correct and genuine comments before downloading an application to prevent such malfunctions.For it the different comments and ratings given for each application must be overcome and systematically evaluated using an automated solution.

As a common need for mobile phones, it is important to recognise suspicious applications as fraud, so that store users can identify them. The consumer would find it hard to recognise the statement or reviews that they have seen past or are true to their advantage. We thus suggest a framework that detects certain fraudulent applications in the Play or App Store with a full description of the fraud detection system.

By taking the data mining and sentiment analysis into account, we are more likely to get real evaluations and thus we suggest a framework for evaluating registered users on a product or a number and assessing them as positive or negative. This can also be helpful for determining the application of fraud and for maintaining mobile protection.

We initiate the method by taking into account the mining lead session or the active application times. This detects local anomalies rather than the global anomaly in the application rating. In particular, we propose a simple and yet fruitful calculation in this respect to acknowledge the principal sessions of each app based on its authentic records of positioning. At this point, in the investigation of the positioning of Apps, the fake applications are found that regularly contain distinctive examples of positioning contrasted and popular apps at each driving session.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

In addition, we inspect the consolidation of the three by statistical hypotheses testing using three different kinds of evidence: ranking, rating and analysis. In any case, the positioning data may be affected by the status of the App developer and some genuine promotional practises, such as 'limited downtime.' Therefore only rating confirmations are insufficient.

In addition, it provides two forms of external proof based on the ranking of applications and survey history, reflecting some of the Apps' specific trends. A hybrid approach is also used to incorporate all the evidence required for fraud detection. To do this, we test the framework proposed by using data from Google Play Store and IOS App Store collected on the real world application for a long time.

2. PROBLEM STATEMENT

There are some connected works, for instance, we tend to positioning spam recognition, on-line survey spam identification and transportable App suggestion, and however the difficulty of distinctive positioning falsehood for mobile Apps is until under-investigated. The problem of sleuthing mobile app rankings is still understood. We tend to create a framework for putting false findings on transportable apps that is a model for sleuthing the rating fraud in mobile applications in order to address these needs. We have many problems to discover for that. To begin with, fraud happens any time the entire life cycle of the app involves a specific determination of the time of fraud. Second, it is troubling to manually mark the rating of fraud for each app because of the wide variety of mobile applications and thus it is important to notice fraud while not victimising basic information mechanically. Mobile apps do not seem to be forever high in the leader board, but fraud normally happens in leading sessions only in some of the big events.The key goal is to identify mobile app fraud at leading sessions at intervals. Initial suggested an effective rule to locate each app's historical ranking records for the leading sessions. Then examine the ranking behaviour, evaluating the dishonoured applications in every leading session normally have entirely different ranking trends as opposed to conventional applications. Thus the past rating records of Apps show some signs of fraud. Then three functions are designed to extract proof of fraud based on such rating. Subsequently, a revised Apps evaluation and audit history, representing some irregularity designs from Apps' genuine rating and survey documents, is expected in addition to two forms of error confirmation. In addition, the Partner Degree Unchecked Price Collection Process, which is used to evaluates the validity of driving sessions from flexible apps, is developed to integrate these three kinds of confirmations.

3. AIMS AND OBJECTS

The main purpose is to establish such a scheme that ranking, rating and review behaviours be noted in the case of work evaluation based evidence, rating-based evidence and ranking-based results. Thus we propose associated humanoid applications that can use the information, comments and 3 reviews of the appliance to provide results for linguistic communication. So the application of fraud is easier to assess. The key aims are to include a framework with various mappings from reviews to topics of interest and a list of reviews of each subject reflecting a user's feeling on the topic. As a developer, keep your application up-to-date with most requested features or bug fixes. It is important to "stay at the top of your game." However, for each app, most app stores offer an average rating of just 5.Therefore, why a person likes or dislikes a particular person is hard to identify. This is the dilemma we are trying to solve.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

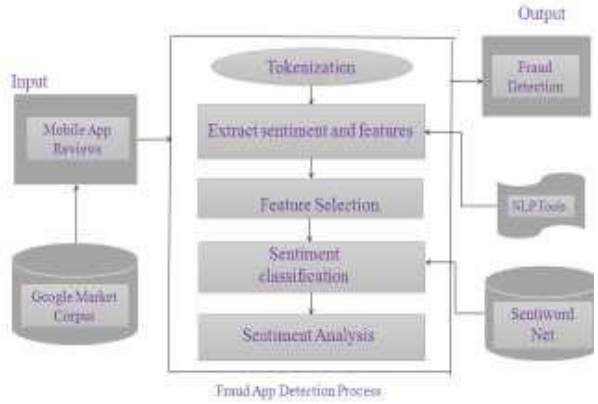

Fig 2. Fraud App Detection process

With the expansion in the quantity of web Apps, to distinguish the misrepresentation Apps, this undertaking proposes a straightforward and successful framework. Fig.1 demonstrates the Framework of Fraud positioning disclosure in portable application.

4. RESEARCH METHODOLOGY

The issue of removing fraud is still at work from the literature survey and other previously proposed systems that have been developed for this very reason. Some works include the use of spam signatures, online spam analysis and mobile device recommendations or even concentrate on malware detection in the applications before they are installed. Google uses a Fair Play method that can detect malware that is only available on certain applications but has not been sufficiently successful due to the dissimilar properties. The user can be fooled by his ratings in installing an app, even though they have viruses that may affect the mobile functioning.

Although other systems exist, the main emphasis is not only on recommending or deleting spam. Some methods may be used to search the historical evaluation and examination documents for anomalies, but for some period of time they are not fraud evidence (mining leading sessions). Which broad application growth in stores makes

it a difficult task to decide which of the lake is valid or not based solely on a rating.

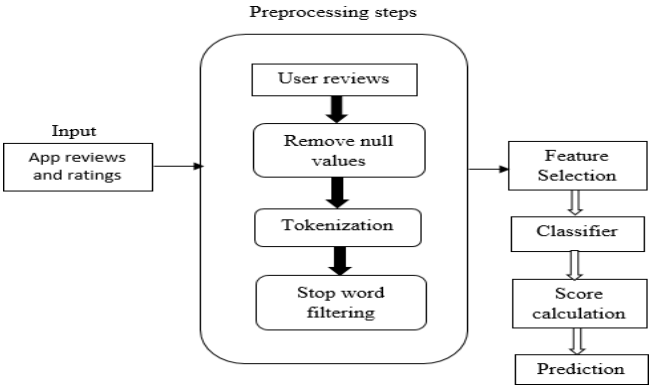

Fig 3. Demonstrates the Framework of Fraud detection in portable application

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

5. PROPOSED METHOD

Some methodologies are used for the creation of the framework. Application classification with Naive Bayes algorithm is the technique used in this project. This company provides a simple and effective platform by expanding the amount of Web Apps in order to distinguish the misrepresentation of Apps. Fig.3 displays the System for portable device fraud detection. Here we propose a framework that uses sentimental comments and data processing to detect fraud applications. We are able to identify them as constructive and negative comments by looking at these remarks. We get the highest probability with the mixture of these proofs. The method proposed includes several steps:

1. Data collection of different app reviews
2. Pre-processing of data
3. Application and score estimation of the Naive Bayes algorithm

Now we discuss about these inbriefly.

A. Collecting the various app reviewsdataset:

From open sources such as Kaggle, Google Public Data Solutions has been obtained by us. The dataset includes reviews for various social, games, training, finance, news and food classes.

B. DataPreprocessing:

In this step user feedback to delete unnecessary text are processed.

1.Tokenization:Tokenization is the process of ending the series of strings into bits, such as words, sentences, symbols or tokens. The tokens collection is applied to further processes such as text-mining and parsing.

2.Stopwordremoval:NLP calls useless words stop words. A stop words may also be used as a common word such as a, the and since, from, in and much more, it is programmed as an enquiry engine.

3.Stemming: The stemming algorithm is used to search for the basic phrase. Stemming involves the reduction of a word to its word stem and is extended to the suffixes and prefixes or roots of words known as lemmas. For finding basic phrase, Porter Stemmer Algorithm is used.

A. Algorithm:

Sentiment Analysis tests people's inclinations through NLP, linguistic and text analysis. SVM and Naive Bayes are the easiest supervised machine learning algorithms according to developers and experts in ML. Analysis of the feelings provides an insight on the positive, neutral and negative feelings of the documents.

Naive Bayes:

The Classifier Naive Bayes is the most basic and widely used. In several practical applications, the Bayesian naive classification can be a learning approach which has been considered useful. It is called naive as it introduces the simplifying assumption that, given the instance classification, the attribute values are conditionally independent. The classification model of Naive Bayes calculates a category's posterior probability. In this module, we measure and compart the typical rating of the specific app, which may have positive or negative ratings of between one and five.The rating above three is considered to be positive and below three is considered negative. Finally, the output is null and null and negative, meaning that the associated output provides null while positive outcomes have an associated output. After a User Ratings and Reviews Positive review pre-processing, adds plus one to a positive score and adds one to a negative.This

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

determines the score of all user ratings, reviews and graphs, determines positive and negative reviews and supports positive and negative reviews, whether or not the app is fraud.

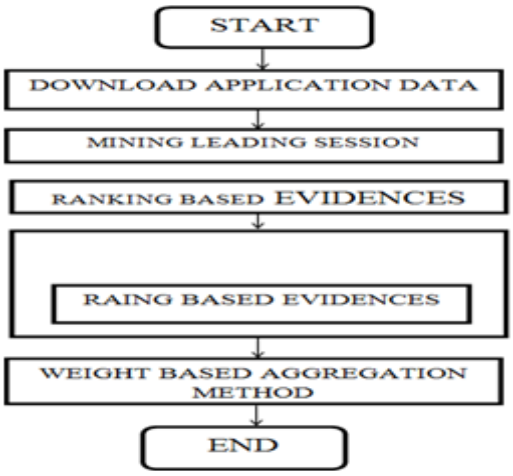

Figure 4: Flow chart of procedure

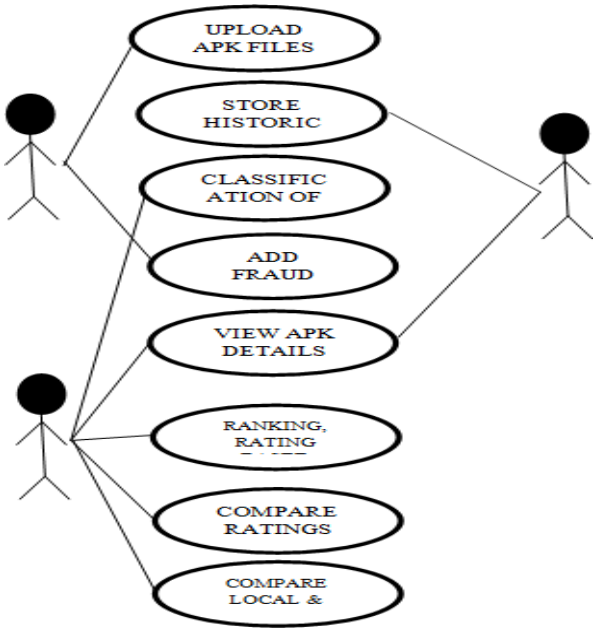

Figure 5: Flow chart of fraud rating and local analysis.

Hardware design

- Correct results should be given by the application.
- Carry out the desired function: sort applications for fraud.
- Increase the versatility and make it easier to use.
- Users should have device access for previous reports reviewed.
- Windows users can have Windows 7, Windows 8 and Windows10 operating systems (32/64 bits).
- Implementation of the framework with Android Studio (JAVA, XML).

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

• For the Android Emulator, minimum of 3 GB RAM, 8 GB of RAM is required and 1 GB is needed.
• The screen resolution of the device should be at least 1280 x 800.

Software Implementation

Sentiment Analysis

Opinion Mining Analysis is a related content mining method which recognises and removes emotional data in the source material and enables a company to gain insight into a social image, object or administration during observation on the site. The sentiment analysis is the most widely recognised content grouping device to evaluate a message and assess whether the individual estimate is certain, negative or impartial.

Stimulus analysis is actually a topic of unbelievable fascination and innovation as there are various functional applications. As data available on the Internet are constantly increasingly publicly and secretly accessible, countless communication conclusions in the areas of auditing, debate, online magazines and web-based social networking are available.

This unstructured data can therefore be converted into organised information from common assessments, with the help of opinion mining frames, concerning products, administrations, brands, governmental issues or any item that people can express feelings about. Such data can be extremely useful for business applications such as research, ads, articles surveys, net advertiser scoring and item analysis as well as client administration.

There are several forms and types of opinion mining and methods that varies from extreme positive, negative, impartial) structures that identify (irate, glad, miserable etc or differentiate feelings and feelings (for example intrigued. not intrigued) (for example intrigued not intrigued) (for example intrigued not intrigued) (for example intrigued v. not intrigued).

6. SCOPE OF THE STUDY

- Writing Scope App Store analyses involves people who carry out analysis into the selection of apps mined from an app store.
- We are particularly interested to consider consolidating specialist and un-specific traits, as the new research openings established by application stores are pioneered with these examinations.
- We will also implement checks, however, to authorise their devices for the arrangement of real applications or to use unique characteristics such as the malware search mechanism apps before they can be released in major App Stores.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

- Our overview is not a systematic analysis of literature (SLR). The App Store Review area continues to expand but has not reached a degree of growth that enables us to ask about inquiries from an extensive range of contexts.

7. SIGINIFICANCE OF THE STUDY

It plans to study the latent relationship between rating, assessment and rankings and assess more realistic proof of fraud. In addition, the fraud-detection strategy, for example mobile app suggestion, will be applied to the rating of alternative mobile app providers for improving user experience.

8. CONCLUSION

The primary aim of the work was to investigate the detection of fraud in apps and employ the methodology of sentiment analysis to differentiate individual fraud applications. Different types of apps use the proposed approach for detecting fraud applications for the experimental study. Three kinds of proof, such as ranking evidence, rating based evidence, and evaluating evidence, were detected in our system. In addition, the approach of aggregation based on optimization incorporates all three facts to identify fraud. Different class values and threshold values give different precision results of your execution time. We found through research that, compared to other algorithms, the proposed procedure provides 90% accuracy.

REFERENCE

1. Esther Nowroji., Vanitha., "Detection Of Fraud Ranking For Mobile App Using IP Address Recognition Technique", vol.4, International Journal for Research in Applied Science & Engineering Technology,2016.
2. Daniel A. Keim, "Information Visualizing and Visual Data Mining" IEEE Trans. Visualization and Visual Data Mining, vol. 8,Jan-Mar 2002.
3. FuzailMisarwala, KausarMukadam, and KiranBhowmick, "Applications of Data Mining in Fraud Detection", vol.32015.
4. Ahmad FIRDAUS, Nor BadrulANUAR, Ahmad KARIM, MohdFaizalAbRAZAK, "Discovering optimal features using static analysis and a genetic search based method for Android malware detection" Frontiers of Information Techonology and Electronic Engineering,2018.
5. JavvajiVenkataramaiah, BommavarapuSushen, Mano. R, Dr. GladispushpaRathi, "An enhanced mining leading session algorithm for fraud app detection in mobile applications" International Journal of Scientific Research in Engineering., April2017.
6. Avayaprathambiha.P, Bharathi.M, Sathiyavani.B, Jayaraj.S "To Detect Fraud Ranking For Mobile Apps Using SVM Classification" International Journal on Recent and Innovation Trends in Computing and Communication, vol. 6, February2018
7. Suleiman Y. Yerima, SakirSezer, Igor Muttik, "Android Malware Detection Using Parallel Machine Learning Classifiers", 8th International Conference on Next Generation Mobile Applications, Services andTechnologies,Sept.2014.
8. SidharthGrover,"Malware detection: developing a system engineered fair play for enhancing the efficacy of stemming search rank fraud", International Journal of Technical Innovation in Modern

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 15556 - 15564
Received 25 April 2021; Accepted 08 May 2021.

Engineering &Science,Vol. 4, October2018

9. PatilRohini,Kale Pallavi, JathadePournima, KudaleKucheta, Prof.PankajAgarkar,"MobSafe: Forensic Analysis ForAndroid Applications And Detection Of Fraud Apps Using CloudStackAnd Data Mining", International Journal of Advanced Research in Computer Engineering &Technology,Vol. 4, October2015