Detection of Fraud Apps Using Sentiment Analysis

Sadhana Chougule [1], Husnara Samdole [2], Yogita vankore[3], Poonam mashalkar[4], Minakshi Yetanale[5].

[1]*Dept of Information Technology, Dr. j. j. Magdum College of Engineering, Jaysingpur, Maharashtra, India*

[2]*Dept of Information Technology, Dr. j. j. Magdum College of Engineering, Jaysingpur, Maharashtra, India*

[3]*Dept of Information Technology, Dr. j. j. Magdum College of Engineering, Jaysingpur, Maharashtra, India*

[4]*Dept of Information Technology, Dr. j. j. Magdum College of Engineering, Jaysingpur, Maharashtra, Ind*

[5]*Dept of Information Technology, Dr. j. j. Magdum College of Engineering, Jaysingpur, Maharashtra, India*

Abstract— **There are several mobile applications in the day to day life so it's very important to keep track of safe and dangerous mobile applications. There are some safe apps and there are some fake apps or fraud apps. So one needs to judge which is fraud and which is real. One can not judge the app is how safe and true. One can not judge how safe and true each application is based only on the reviews. Such reviews are mentioned for every application. So it is very important to keep the track of reviews or comments and to develop such a system to make sure apps present are genuine or not. The main objective of this paper is that to develop a system for the detection of fraud apps before downloading the application. Such a system is based on sentiment analysis and data mining.**

The sentiment analysis helps in the determination of the emotional tones behind words that are expressed in the online mode. This method is useful to monitor social media and which in turn helps to get a better idea of public opinion. These opinions are on specific issues. The user may not get the correct or true review always. We can check for user' s sentimental comments on multiple applications. The reviews can be genuine or fake. To analyze the rating and review altogether involves both user and admin comments. Based on it one can determine whether the app is genuine or not. The machine can learn, analyze the sentiment of the people, emotions about reviews, and for the other text also by using sentiment analysis and data mining. The manipulation of reviews is one of the key aspects of App ranking fraud. With the use of sentimental analysis and data mining, analysis of the reviews and comments will help to determine the correct application for both Android and ios platforms.

Keywords— Sentimental Analysis, Data mining, Review based evidence, positive negative ratings, Rate evidence, Users review, Leading session

I. INTRODUCTION

There is an increase in the use of mobile and its

applications because there is growth in technology. There are numerous platforms like Android and iOS where mobile applications are developed. Because of the rapid development in technology, usage, sales, and developments, it has become a significant challenge in the world of the business intelligence market. This is turn rise in the market competition. So there is tough competition among the companies and application developers. This is to prove their quality of product and spend an immense amount of work into attracting customers to sustain their future progress.

The customer rankings, ratings, reviews on the specific application are the most important that play a vital role when the user downloads the application. This could be a way for the developers to find their weaknesses and enhance the development of a new one keeping in mind the people's needs. Certain times the developers misguiding 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 App developer' s option in contrast to some false way to intentionally support their Apps and in the long run controls the outline rankings on an App store. This is executed by utilizing so-called "bot ranches" or "human water armed forces". This expands the Application downloads evaluations and audits in an exceptionally brief time. Sometimes, the upliftment of the developers tends to hire teams of workers that commit fraud collectively and provide false comments and ratings over an application.

This is called crowd turfing. so it is important to ensure that before installing an app, the users are provided with proper and genuine comments to avoid certain misguiding. so an automated solution is required to overcome. It systematically analyses the various comments and ratings. This is provided for each application[19,20].

With mobile phones being a quite popular need, suspicious applications must be marked as fraud 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 are a scam or a genuine one for their benefit. Thereby, we are proposing a system that will identify such fraudulent applications on Play or App store by providing a holistic view of ranking fraud detection system. We can initiate the system by the consideration of the mining leading session

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 app ranking. In particular, in this, we first propose a basic yet fruitful calculation to recognize the leading sessions of each App dependent on its authentic positioning records. At this point, of the investigation of Apps' positioning practices, it finds the fake Apps that regularly have distinctive positioning examples in each driving session contrasted and ordinary Apps.

II. LITERATURE SURVEY

The main focus of this project is upon sentiment analysis and data mining to extract the dataset produced. By using this method, we will be able to determine the true value of the applications which are provided in Play and App stores. Such a proposed system will contain a huge amount of

data set that has to be dealt with and using data mining along with visual data will help in carrying out the system.

Information or data mining is the way toward extricating required information from substantial informational collections and changes it into a justifiable arrangement for sometime later, essentially utilized for some, business-based reason. Sentiment Analysis is pitched into this procedure as a piece of it. Since it is the way toward examining explanations and acquiring abstract data from them. At an exceptionally fundamental dimension, it is discovering the extremity of the announcements.

Information is gathered from different internet-based life, portable applications, and exchanges which contain surveys, remarks, and different data identified with the individual business. Further here feeling examination is utilized for breaking down the information for future upgrades dependent on the measurements acquired by estimation investigation.

The investigation of extensive informational collections is a critical however troublesome issue. Data representation procedures may help to take care of the issue. Visual information investigation has high potential and numerous applications, for example, misrepresentation discovery also, information mining will utilize data representation innovation for an improved information examination [1].

Data mining is utilized in determining fraud efficiently and that' s what we propose and implement in this paper. By utilizing various data mining techniques and algorithms, it would become easier for us to determine our backend retrieval of data. Fraud can be classified into various types [2] which are the applications of data mining. With the end goal of grouping, extortion has been separated into four general classifications budgetary misrepresentation, media communications extortion, PC interruption, and protection

III. LIMITATIONS OF THE EXISTING SYSTEM

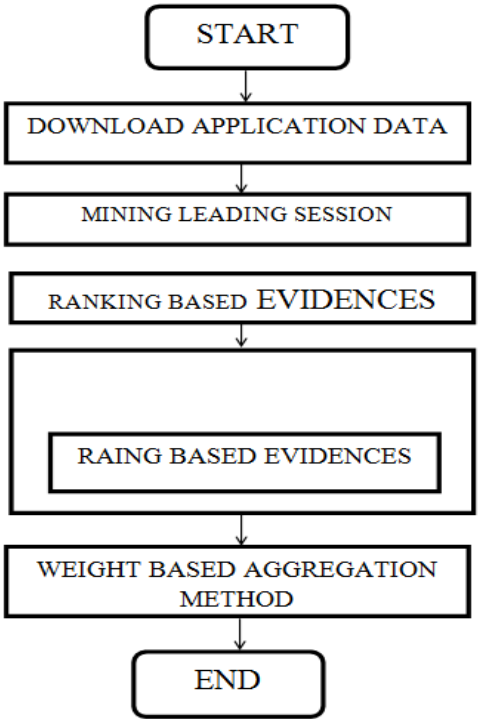
With the help of Literature survey and also other past proposed systems has been studied. These system were developed for this very purpose. But the problem in eradicating the fraud application is still under work.There is some work involve the usage of web ranking spam detection, online review spam, and mobile application recommendation or even focuses on the detection of malware in the apps before downloading them. The advanced systems in Google uses the FairPlay system. This system can detect the malware that is present in certain apps only. But this system hasn' t been efficient enough to do so due to the concealing properties. The user can be trapped into downloading an application by its ratings even when it does contain certain viruses that can affect the functioning of the mobile.

Though there have been other existing systems, our main focus is not only on recommendation or spam removal.There are some of the approaches can be used for anomaly detection This detection is by the use of the historical rating and review records but they aren' t fraud pieces of evidence for a certain period (mining leading sessions). Which the wide growth of apps in stores, it becomes a cumbersome task to determine which of them are genuine or not based on the ranking alone.

Here we propose a system that involves detecting the fraud apps using sentiment of the users. These sentiments are in the form of comments and we are also using data

mining process. We can check the user' s sentimental comment on multiple applications by comparing their views of the admin and the user. By looking into these comments, we can distinguish them as positive or negative comments. With the aggregations of three pieces of evidence: rank based, rating based, and review based we can get a higher probability of the result. The data is extracted and processed by the data mining technique and machine learning techniques. The data is then evaluated on the three mentioned pieces of evidence and are concatenated before the result. It is important to brief about sentiment analysis and data mining before continuing further into the proposed system and algorithms.

IV. PROPOSED SYSTEM MODEL/ARCHITECTURE



METHODOLOGY

Data Collection:

We have collected data from different social media sources such as Twitter. The data may have annotations or fine-grained annotations. We have collected datasets or tweets of the Twitter website which is a social media source. We have collected tweets that have bullying words.

Data Cleaning:

Once the data is collected, data is preprocessed before using it. The empty rows and unnecessary data are removed from the dataset. There is some library in python like pandas or numpy which are used for data cleaning. Once the data is cleaned, we need to preprocess the data. There are some processing libraries in python like Scikit-learn.

Training model:

The data is split into training and testing. Generally, we keep a 70: 30 ratio or 80:20 ratio of training and testing respectively. The data is generally divided using the Scikit-learn library. Once the data is trained using the Naïve Byes algorithm, it is tested on the remaining data.

Test Model:

The model is tested for accuracy with the real-time dataset. Once the dataset is trained and tested successfully, the model is created and is used in the application. We have designed an application on the web and it's deployed. The model is tested on the new data and the accuracy is checked.

V. IMPLEMENTATION DETAILS

A. SentimentAnalysis

Sentiment Analysis is relevant mining of content. This recognizes and extricates emotional data in the source material. This helps businesses to comprehend the social slant of their image, item, or administration. This implies observing on the web discussions. Sentiment analysis is the most widely identified content grouping device. These investigations an approaching message. This tells the basic estimation is sure, negative or unbiased.

The sentiment analysis is a theme of incredible intrigue and advancement. It has numerous handy applications. Since freely and secretly accessible data over the Internet is continually growing, countless communicating conclusions are accessible in audit locales, discussions, online journals, and web-based social networking.

With the assistance of opinion mining frameworks, this unstructured data could be consequently changed into organized information of popular assessments about items, administrations, brands, governmental issues, or any point that individuals can express feelings about. This information is exceptionally valuable for business applications like showcasing examination, advertising, item surveys, net advertiser scoring, item criticism, and client administration.
There are numerous sorts and kinds of opinion mining and tools run from frameworks that attention on the extremity (positive, negative, unbiased) to frameworks that recognize sentiments and feelings (irate, glad, miserable, and so forth) or distinguish aims (for example intrigued v. not intrigued).

B. Data Mining

There are various features for information to be accessible in the Information Industry. This information is not useful until it is changed over into helpful data. It is important to examine this gigantic measure of information and concentrate helpful data from it. The extraction of data is not the main procedure of the project. we have to perform; information mining that includes different procedures. These include Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation, and Data Presentation. When every one of these procedures is finished, It would likely utilize this data in various applications, significantly lower with an F-score of 0.12 % and 7 % respectively. These various fields are Fraud Detection, Market Analysis, Production Control, Science Exploration, etc. The data mining is utilized here to look into the review data by the apps. This data is then filtered and processed before it can go through the process of sentiment analysis. The reviews are extracted and distinguished based on various datasets that are in the database. Accordingly,thetextisevaluated.To be particular, we are using text data mining which is also referred to as text mining. From the texts which are extracted(reviews), it is easier to analyze words or a cluster of words that are used.

IV . CONCLUSION

This paper describes the determination of fraud mobile applications. This is happened by using the concept of data mining and sentiment analysis. The architecture diagram explains the algorithm. The processes that are implemented in the project. The data is collected and stored in the database.

This is then evaluated with the supporting algorithms that are defined in the paper. This is a unique approach in which the pieces of evidence are aggregated and confined into a single result. The proposed paper is scalable and can be extended to other domain areas. The pieces of evidence for the ranking fraud detection. The experimental results showed the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity in the ranking fraud activities.

V. FUTURE WORK

In the future, it's planned to add more datasets and to improve the accuracy of the model. Also to review more practical fraud pieces of evidence. The analysis of the latent relationship among rating, review, and rankings should be done in more detail. It will be extended to ranking fraud detection approach with alternative mobile App related services, like mobile Apps recommendations, for enhancing user experience.

REFERENCES

[1]. Esther Nowroji, Vanitha, " Detection Of Fraud Ranking For Mobile App Using IP Address Recognition Technique" , vol. 4.

[2]. Javvaji Venkataramaiah, Bommavarapu Sushen, Mano. R, Dr. GladispushpaRathi, " An enhanced mining leading session algorithm for fraud app detection in mobile applications"

[3]. S.R.Srividhya, S.Sangeetha – " A Methodology to Detect Fraud Apps Using Sentiment Analysis"

[4]. Keerthana. B, Sivashankari.K and Shaistha Tabasum.S, " Detecting Malwares and Search Rank Fraud in Google Search Using Rabin Karp Algorithm" , IJARSE, 7(02), 2018, pp.504-527.

[5]. Shashank Bajaj, Nikhil Nigam, Priya Vandana, Srishti Singh, " Detection of fraud apps using sentiment analysis" , International Journal of Innovative Science and Research Technology.

[6]. Harpreet Kaur, Veenu Mangat and Nidhi, — "A Survey of Sentiment Analysis techniques", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017, pp. 921