

Detection of Fraud apps using sentimental Analysis

¹Pushpalatha S, ²Rayadurgam Kiran Kumar, ³Salman Pasha, ⁴chandan R, ⁵Pradeep S kalmath,

^{1,2,3,4,5}UG Student,

^{1,2,3,4,5} Department of Computer Science

Presidency University, Bangalore, India

Abstract: Now a days, mobile Apps are very popular and well-known concept due to the rapid advancement in the mobile technology and mobile devices. Due to the large number of mobile Apps, rating fraud is the key challenge in front of the mobile App market. With the increase in the number of mobile applications in the day-to-day life, it is important to keep track as to which ones are safe and which ones aren't. The objective is to develop a model in detecting fraud apps by using sentimental analysis and Bag of words Algorithm. This method is useful in monitoring social media and helps to get a brief idea of the public's opinion on certain issues project aims to find the fraud intended mobile applications available in various mobile application marketplace. The mobile application reviews are analyzed using sentiment analysis and different machine learning models to see users' sentiment. Our application will help mobile users alert about the fraudulent apps installed in their mobile devices based on the users' sentiment analysis.

Keywords—Positive negative reviews, Sentimental Polarity, Bag of words Algorithm, User reviews.

I. INTRODUCTION

With the increase in technology, there is an enlarge the usage of mobiles. There has been a huge growth in the development of several mobile applications on many platforms such as the popular Android and iOS.

Due to its quick growth day by day for its everyday consumption, sales and developments, it has become a serious challenge in the world of the business market. This gives rise in the market competitiveness. Now a day's application developers are having a tough competition with one another in order to show their quality of product and spend an enormous amount of work into attracting customers to encourage their future progress.

Fraud apps are most dangerous which infects the devices. It consumes storage space in the devices with malware. Such as the level of revolutionary behind these fake apps, many are capable of obtaining the brand, model, and ID of the device upon which they are installed, even its physical location. While most dishonest apps are usually found on third-party websites, don't make the mistake of assuming an app on Google Play, iTunes Store, or the App Store is legalized. As there are many fraud applications available in app store. As we all know that mobile application market is in rise as mobile users are in large amount, rational mobile user uses those options of mobile Apps an entertaining purpose and knowledge purpose. Our webpage will display the fraud apps and non-fraud apps to user. In this project, we propose a method to detect the fraud application based on user reviews and ratings using Bag of words algorithm.

II. SYSTEM ARCHITECTURE AND DESIGN

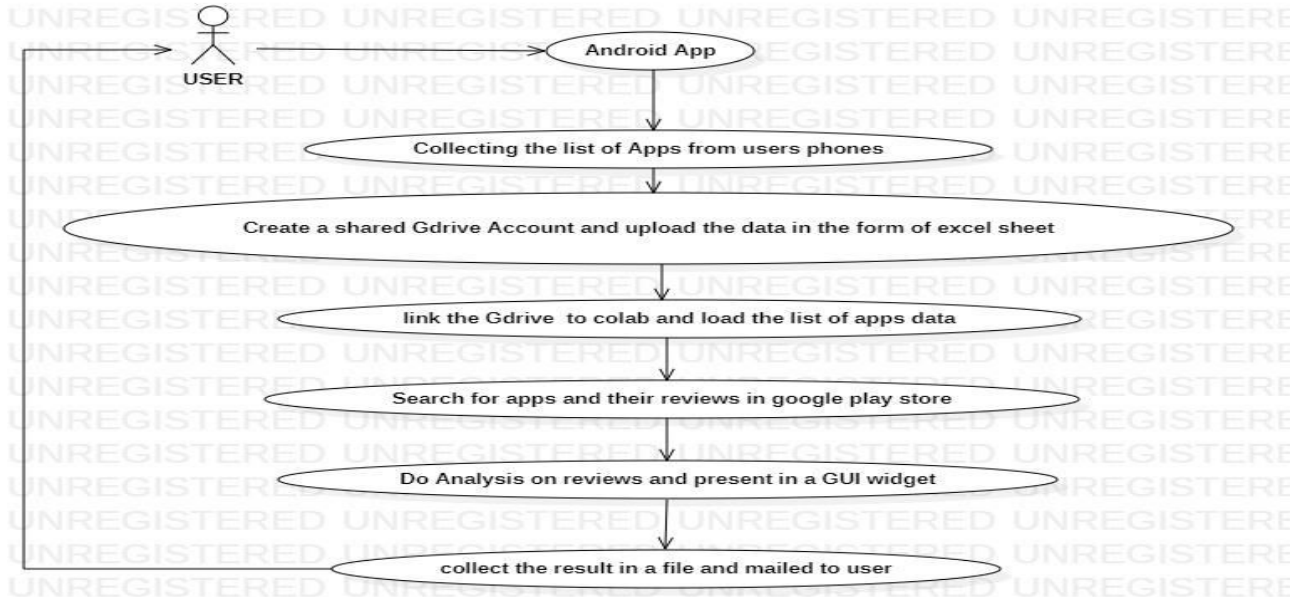


Fig 1. System Architecture

Our proposed system as in Fig 1 gives an overall flow of the process which is happening. It begins with users are going to install the android app i.e., List My Apps in the play store, this app is going to display all the apps present in our mobiles. And we can also get package names from this app. Next, we created a Google drive Account and uploaded the app data like package name in the form of excel sheet (.csv file). After that link the google drive to colab.Cola

b allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. After linking google drive to the colab, load the list of apps package names followed by”,”.

Then to get reviews for apps install python libraries did coding for extracting the data from the package names, appending app information, showcasing the app

icons, data filtering, creating csv files to store reviews and ratings, converting reviews into csv files, generating ratings for all apps, show the app titles and ratings indication using Bar plot ,tokenizing every review using bag of words algorithm, generating list of all reviews of each app in nested list and creating dictionary to store app name as a key and reviews as a value, using data visualization and sentimental polarity colab shows the positive and negative reviews. And using bar plot shows the how much percentage positive and negative apps present. And display fraud apps and non-fraud apps.

III. LITERATURE SURVEY

1. Daniel A. Keim, “Information Visualizing and Visual Data Mining” IEEE Trans. Visualization and Visual Data Mining, Visual Data

Exploration usually follows a three-step process: Overview first, zoom and filter, and then details-on demand. First, the user needs to get an overview of the data. In the overview, the user identifies interesting patterns and focuses on one or more of them.

2. FuzailMisarwala, KausarMukadam, and KiranBhowmick, "Applications of Data Mining in Fraud Detection", Data mining is the process of analysis of data from various outlooks and summarizing it into useful information to increase revenue, cut costs, or both. It allows users to analyse data from many different dimensions or angles, categorize it, and effectively summarize the relationships identified. It is the process of finding correlations or associations among a multitude of fields in large relational databases
3. Esther Now roji., Vanitha., "Detection Of Fraud Ranking For Mobile App Using IP Address Recognition Technique", Proposed system uses the rank aggregation method which promoted by ranking manipulation it will be the top in the market, other users will buy the App. the expected ranking for a required period, the manipulation will be stopped and ranking of the malicious App will be Meanwhile, cost of ranking manipulation with high ranking expectations is quite expensive due to the unclear ranking manipulation proposed system with data collected as dynamic ratings for App in daily basis.
4. Ahmad FIRDAUS, NorBadrulANUAR, Ahmad KARIM, et al. Discovering optimal features

using static analysis and a genetic search-based method for Android malware detection, Genetic algorithms simulate the process of natural selection which means those species who can adapt to changes in their environment are able to survive and reproduce and go to next generation. Genetic Algorithm is an algorithm that mimics the natural evolutionary process. It consists of a crossover process that combines multiple generations and continues to loop until the best generations are achieved.

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., Ranking fraud refers to fraudulent or vulnerable activities which have a purpose of bumping up the Apps in the popularity list. In fact, it becomes more and more frequent for App developers to use tricky means, like increasing their Apps' sales or posting fake App ratings, to commit ranking fraud.

ALGORITHM Mining Leading Sessions:

There are two main steps for mining leading sessions. First, we need to discover leading events from the Apps historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.

6. Avayaprathambiha.P, Bharathi, Sathiyavani.B, Jeyaraj's "To Detect Fraud Ranking for Mobile Apps Using SVM Classification" International Journal on Recent and Innovation Trends in Computing and Communication, Support Vector Machine (SVM) could be a machine

learning tool that's supported the concept of huge margin information classification. The tool has sturdy theoretical foundation and also the classification algorithms supported it offer smart generalization performance. commonplace implementations, although offer smart classification accuracy, square measures slow and don't scale well. thence they cannot be applied to large-scale data processing applications. They usually would like sizable amount of support vectors.

7. Suleiman Y. Yerima, Sakir Sezer, Igor Muttik, "Android Malware Detection Using Parallel Machine Learning Classifiers", 8th International Conference on Next Generation Mobile Applications, Services and Technologies, this paper proposes and investigates a parallel machine learning based classification approach for early detection of Android malware. Using real malware samples and benign applications, a composite classification model is developed from parallel combination of heterogeneous classifiers.
8. Sidharth Grover, "Malware detection: developing a system engineered fair play for enhancing the efficacy of stemming search rank fraud", International Journal of Technical Innovation in Modern Engineering & Science, they developed 28 behavioural as well as linguistic features for identification of fraud as well as malware, which is used to guide organize learning algorithms. Then they used GPad, a device to load apk's of without charge apps as well as scrutinize pro malware via Virus Total.

IV. System Implementation

1: Install the necessary python libraries: In the google colab notebook install necessary libraries like nltk, Google Play scrapper, matplotlib, seaborn, pillow etc.

2: Data Extraction: Collecting package names from my applist mount to the google drive connect to colab. Appending app information to Applist. Then showcase all apps icons

3: Data Filtering: Creating a CSV file converting into apps reviews and ratings store the values. By using word cloud, we can display all the frequent words been used in the reviews or ratings. Then generating reviews for all apps. then show app titles and ratings by using barplots.

4. Tokenizing Every Review: Using "Bag of Words" algorithm generating list of all reviews of each app in a nested list, creating dictionary to store app name as key and reviews as value.

5. Sentimental Polarity: In this step we are created a text blob and we introduced a naïve bayes analyzer to get positive and negative polarity of each app reviews. Also showing app titles and rating indication using bar plot. In this bar plot it shows the positive and negative reviews percentage and shows the which are good apps and bad apps.

V. OUTPUT

A)

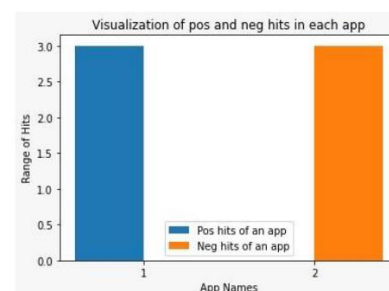


Fig 2: Data visualization graph positive and negative reviews

B)

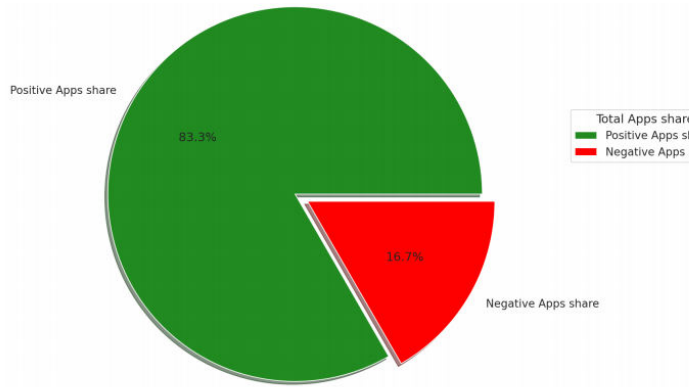


Fig 3: Fraud apps and Non fraud apps

C)



Fig 4: Positive Apps

D)



Fig 5: Negative Apps

VI. FUTURE SCOPE

Fraud is common, and fraud is rising in application cases. Importantly, the prevention of fraud until it is too late. Fraud can also be prevented. Fraud apps basically deals with fake apps. So, our system will help the user to identify which application is good and bad. In future, using the prototype model, we can build a large-scale application also. This application can be accessed anytime and anywhere in the world. Because we used colab and it is open source. In the future, we will decide to study more practical fraudevidence and analyze the inactive relationship among ratings and reviews. Moreover, we can add more services inranting fraud detection approach to enhance userexperience.

VI. ACKNOWLEDGMENT

We are greatly indebted to our guide Prof Dr. S Thiruselvan Department of Computer Science and Engineering, Presidency University for his inspirational guidance, valuable suggestions and providing us a chance to express our technical capabilities in every respect for the completion of the project work.

VII. CONCLUSION

This paper had presented about determining fraud and non-fraud applications by using the concept of Bag of words algorithm, inbuilt libraries and Sentiment analysis. Sentiment analysis is a field of study that analyzes people’s sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization processes which are implemented in the project. This is a unique approach in which the evidences are aggregated and confined into a single result. The proposed framework is scalable and can be extended to other domain generated evidences for the review fraud detection. The experimental results showed the effectiveness of the

proposed system, the scalability of detection algorithm as well as some regularity in the ranking fraud activities.

REFERENCES

- [1] Daniel A. Keim, "Information Visualizing and Visual Data Mining" IEEE Trans. Visualization and Visual Data Mining, vol. 8, Jan-Mar 2002.
- [2] FuzailMisarwala, KausarMukadam, and KiranBhowmick, "Applications of Data Mining in Fraud Detection", vol. 32015.
- [3] Esther Now roji., Vanitha., "Detection of Fraud Ranking for Mobile App Using IP Address Recognition Technique", vol.
- [4] Ahmad FIRDAUS, NorBadrulANUAR, Ahmad KARIM, et al. Discovering optimal features using static analysis and a genetic search-based method for Android malware detection[J]. Front. Inform. Technol. Electron. Eng., 2018, 19(6): 712-736.
- [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, Sathiyavani.B, Jeyaraj'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 and Technologies, Sept.2014.
- [8] Sidharth Grover, "Malware detection: developing a system engineered fair play for enhancing the efficacy of stemming search rank fraud", International Journal of Technical Innovation in Modern Engineering & Science, Vol. 4, October2018.
- [9]PatilRohini, Kale Pallavi, JathadePoornima, KudelaKucheta,Prof.PankajAgarkar, "Mob Safe:Forensic Analysis For Android Applications and Detection of Fraud Apps Using Cloud Stack And Datamining", international Journal of Advanced Research in Computer Engineering & Technology, Vol. 4, October 2015.
- [10] D.Janet, Vikrant Chole , "A Review on Ranking Based Fraud Detection in Android Market", International Journal of Science and Research. 6, January2017.