IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FAKE APPLICATION DETECTION

Syed Javvad*1, Abhijeet Pawale*2, Diksha Gaikwad*3, Feroz Pathan Sir *4, Ganesh Deshpande Sir*5

*4 Assistant Professor, Department of Computer Engineering, Gramin Technical & Management Campus, Nanded, India.

*5 Professor & Head of Department of Computer Engineering, Gramin Technical & Management Campus, Nanded, India.

ABSTRACT

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. 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 necessary to check and initiate a system to make assured that the apps present are genuine or fraud. The objective is to develop a web system in detecting fraud apps before the user downloads by using sentimental analysis and support vector machine. 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. The reviews may be fake or genuine. Analyzing the reviews involving both user and admins comments, we can determine whether the app is genuine or not. Using sentimental analysis and support vector machine, the machine is able to learn and analyze the sentiments, emotions about reviews and other texts. The manipulation of review is one of the key aspects of App ranking fraud. By using sentimental analysis and support vector machine, analyzing reviews and comments can help to determine the correct application for both Android and iOS.

Keywords: Positive negative neutral reviews, Sentiment analysis, Support Vector Machine, Users reviews.

INTRODUCTION

With the broadening in technology, there is an enlarge 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 market. This 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.

Our Webpage will show the customers reviews on that particular application which the want 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 apps or malicious ones use it as a platform to spread malware throughout. 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 liftmen 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 surfing. Hence it is always important to ensure that before installing an app, the users are provided with proper and genuine comments in order to avoid certain mishaps. For this, an automated solution is required to overcome and systematically analyses the various comments and ratings that are provided for each application.

It is necessary to know the users that doubtful applications must be marked as fraud. It will be difficult for user to modify the comments of the app they see is a fraud or genuine. Thereby, we are proposing a system which will identify such fraudulent applications on Play or App store by providing a holistic view of review fraud detection system.

By considering support vector machine and sentiment analysis, we can get a higher probability of getting real reviews and hence we propose a system that intakes reviews from registered users for a single product or multiple and evaluate them as a positive or negative or neutral sentiments. This can also be useful to determine the fraud application and ensure mobile security as well.

LITERATURE SURVEY

The main focus of this project is upon the sentiment analysis and support vector machine 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 stores. Such a proposed system will contain a huge amount of data set that has to be dealt with and using support vector machine along with visual data will help in carrying out the system.

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for either of two categories, they’re able to categorize new examples. 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.

Information is gathered from different internet based life, portable applications and exchanges which contain surveys, remarks and different data identified with the individual business. The investigation of extensive informational collections is a critical however troublesome issue. Data representation procedures may help to take care of the issue.

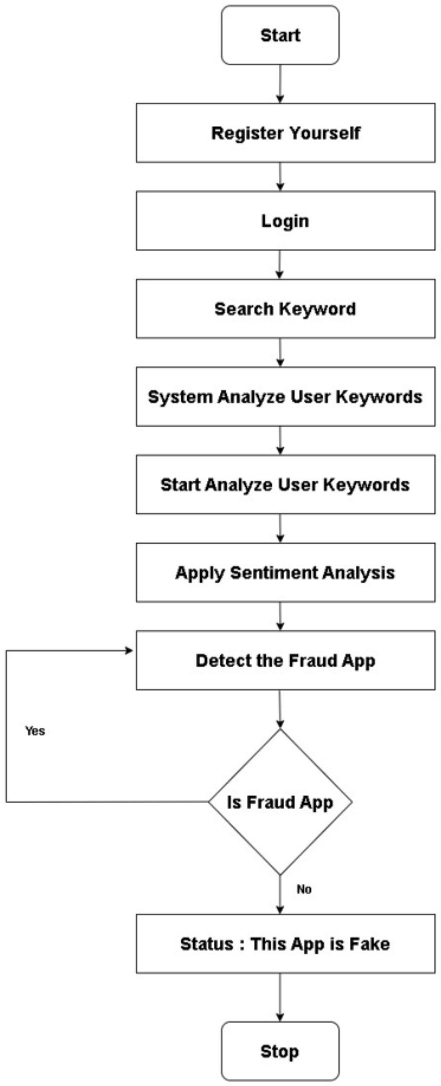
Visual information investigation has high possible and number of applications. Support Vector Machine is utilized in determining fraud efficiently and that’s what we propose and implement in this paper. By utilizing various sentiment analysis techniques and algorithms, it would become easier for us to modify our backend comeback of data. Fraud can be classified into various type which are the applications of data mining.

The reviews for every single application is not enough in determine whether the app is genuine or fraud. As described that it's not quite right to believe into star ratings as they can be manipulated by the developers themselves. It is considered into reading the reviews more than ratings. Generally, it is advised to check more reliable sources such as curated third part reviews or checking the developer's other apps. Collection of a specific app dataset for a period of time and differentiating them as positive and negative reviews. Utilizing fewer words in the reviews is more efficient for the accuracy of semantic classification. Lesser the words, it is easier to classify them according to their category as the proposed system.

PROBLEM STATEMENT

The Medical Center problem facing uses the hands-on process to provide health services, use of cards, files to track progressive patient records, and set records of new patients in the hospital. This procedure has a lot of inconsistencies such as patients losing their health cards or their files being misplaced. Therefore, the automated system helps to keep track of patient records and medical bills making it easier to determine new and ongoing status. This program improves the quality of the service in terms of the medical assistance provided by empowering physicians and administrators to be able to view the level of common diseases and their percentage. Managers can also determine the number of patients a doctor has visited over a set period of time.

Flow Diagram of Fake Application Detection



THE PROPOSED SYSTEM

The admin is allowed to add and create new applications along with the links to the actual app in the play or app store. A set of data is collected for that specific application from both the stores and saved in the database from a specific period of time. Several data pre-processing methods are used in order to clean the data which has been given by the user. As in the architecture, it can be logically visualized with the tokenization, stop word removal and stemming algorithms being used. Here the user's comments and reviews are stored in the database act as the input to the algorithm. Now the number of positive and negative words that appear in reviews are counted. If the number of positive word appearances is greater than the number of negative word appearances, the system returns a positive sentiment, and vice versa. If the numbers are even, the system will return a neutral sentiment. Now the training set will be fitted to the SVM classifier. To create the SVM classifier, we will import SVC class from Sklearn.svm library. we have used kernel='linear', as here we are creating SVM for linearly separable data. However, we can change it for non-linear data. And then we fitted the classifier to the training dataset(x_train, y_train). The model performance can be altered by changing the value of C(Regularization factor), gamma, and kernel.

Predicting the test set result Now, we will predict the output for test set. For this, we will create a new vector y_pred.

After getting the y_pred vector, we can compare the result of y_pred and y_test to check the difference between the actual value and predicted value. Now we will see the performance of the SVM classifier that how many incorrect predictions are there as compared to the Logistic regression classifier. To create the confusion matrix, we need to import the confusion_matrix function of the sklearn library. After importing the function, we will call it using a new variable cm. The function takes two parameters, mainly y_true(the actual values) and y_pred (the targeted value return by the classifier).

CONCLUSION

This paper had presented about determining fraud applications by using the concept of support vector machine and sentiment analysis. It was supported by the architecture diagram which briefed about the algorithm and processes which are implemented in the project. Data gets collected and stored in the database 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 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.

ACKNOWLEDGEMENT

We would like to thank the management of Geetanjali Institute of Technical Studies, Dabok, and Udaipur for giving the necessary infrastructure support to smoothly conduct this research work. Also, we would like to thank Mr. Vikas Misra for his motivating words and guidance at appropriate stages ofthe work. Finally we would like to thank the technical team viz. Mrs. Nikita Somani and Mr. Girish Ameta for their technical support in successful implementation

REFERANCES

[1] Daniel A. Keim, "Information Visualizing and Visual Data Mining" IEEE Trans. Visualization and Visual Data Mining, vol. 8,Jan-Mar 2002.

[2] Fuzail Misarwala, Kausar Mukadam, and Kiran Bhowmick, "Applications of Data Mining in Fraud Detection", vol. 32015.

[3] Esther Nowroji., Vanitha., "Detection Of Fraud Ranking For Mobile App Using IP AddressRecognition Technique", International Journal for Research in Applied Science & Engineering Technology, vol. 4, 2016.

[4] Ahmad FIRDAUS, Nor Badrul ANUAR, Ahmad KARIM, Mohd Faizal Ab RAZAK,"Discovering optimal features using static analysis and a genetic search based method for Android malware detection" Frontiers of Information Technology and Electronic Engineering, 2018.

[5] Javvaji Venkataramaiah, Bommavarapu Sushen, Mano. R, Dr. Gladispushpa Rathi, "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, Sakir Sezer, Igor Muttik, "Android Malware Detection Using ParallelMachine Learning Classifiers", 8th International Conference on Next Generation Mobile Applications, Services and Technologies,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 Engineering &Science,Vol. 4, October2018.

[9] Patil Rohini, Kale Pallavi, Jathade Pournima, Kudale Kucheta, Prof. Pankaj Agarkar, "MobSafe: Forensic Analysis For Android Applications And Detection Of Fraud Apps Using CloudStack And Data Mining", International Journal of Advanced Research in Computer Engineering & Technology, Vol. 4, October2015.

[10] Neha M. Puram, Kavita R. Singh, "Semantic Analysis of App Review for Fraud Detection using Fuzzy Logic", International Journal of Computer & Mathematical Sciences, Vol. 7, January2018.

[11] Vivek Pingale, Laxman Kuhile, Pratik Phapale, Pratik Sapkal, Prof. Swati Jaiswal,"Fraud Detection & Prevention of Mobile Apps using Optimal Aggregation Method", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 8, March2016.

[12] L.Azzopardi, M.Girolami, and K.V.Risjbergen, "Investigating the relationship between language model perplexity and ir precision-recall measures," in Proc. 26th Int. Conf. Res. Develop. Inform