Fraud App Detection Using Sentiment Analysis

Dr. S. Gnanapriya

Assistant professor, Department of Computer Applications , Nehru College of Management ,Coimbatore

R. Arun Kumar

II MCA, Department of Computer Applications,
Nehru College of Management, Coimbatore, Tamilnadu, India

Abstract - Since there are more and more mobile applications used in daily life, it's critical to monitor which ones are secure and which are not. Based just on the reviews listed for each program, one cannot determine how reliable and safe each one is. Therefore, it is essential to verify and start a system to ensure that the apps are authentic or fraudulent. The goal is to create a web system that uses support vector machines and sentiment analysis to identify fraudulent apps before users download them. The purpose of sentiment analysis is to assist in identifying the emotional undertones of words used in online communication. This approach is helpful for keeping an eye on social media and for quickly gauging public sentiment on particular topics. On the internet, the customer may not always find accurate or genuine product reviews. The reviews could be authentic or fraudulent. We can ascertain whether or not the app is authentic by examining evaluations that include remarks from both users and administrators. The system can learn and understand the sentiments and emotions of reviews and other materials by using support vector machines and sentimental analysis. One of the main components of app ranking fraud is the manipulation of reviews. The right app for iOS and Android can be found by examining reviews and comments using emotional analysis and support vector machines.

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

I. INTRODUCTION

The use of mobile phones has increased as a result of technological advancements. The creation of different mobile applications on multiple platforms, including the well-known Android and iOS, has increased significantly. It has become a major obstacle in the business intelligence sector because of its daily, exponential expansion in sales, development, and routine use. The market becomes more competitive as a result. Businesses and application developers are fiercely competing with one another to demonstrate the quality of their products and invest a great deal of effort in luring clients in order to maintain their future growth. Customers' reviews of the specific application they wish to download are shown on our webpage. This could help the developers identify their areas of weakness and improve upon them while keeping the needs of the public in mind. Additionally, there are instances where criminal developers utilize it as a platform to distribute malware or deceitfully guide developers about the recognition of their apps. This is typically done by using "bot ranches" or "human water armed forces" to increase the number of application downloads, audits, and evaluations in a very short amount of time.

An automated solution is needed to overcome and systematically analyze the various comments and ratings that are provided for each application. Therefore, it is always important to make sure that users receive proper and genuine comments before installing an application in order to avoid certain mishaps. Sometimes, developers hire teams of workers who commit fraud collectively and provide false comments and ratings over an application. This practice is known as crowd surfing.

Users must be aware that dubious applications need to be flagged as fraudulent. Users will find it challenging to change the app's comments, whether they are real or fraudulent. Therefore, by offering a comprehensive perspective on review fraud detection systems, we are putting out a system that will detect such fraudulent applications on the Play Store or App Store. Support vector machines and sentiment analysis can increase the likelihood of receiving genuine evaluations. For this reason, we suggest a system that collects reviews from registered users for one or more products and classifies them as either good, negative, or neutral. This can also help identify fraudulent applications and guarantee mobile security.

II. LITERATURE SURVEY

This project's primary focus is on using support vector machines and sentiment analysis to extract the generated dataset. We will be able to ascertain the actual worth of the apps that are offered in Play stores by employing this technique. A large amount of data will need to be handled by the proposed system, and support vector machines in conjunction with visual data will aid in system implementation.

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 collected via various online resources, mobile applications, and interactions that include questionnaires, comments, and other information specific to each company. An important but challenging problem is the examination of large informational sets. Procedures for data representation could assist resolve the problem.

The analysis of visual information has several potential uses. In this research, we propose and develop Support Vector Machine, which is used to efficiently determine fraud. Modifying our backend data return would be simpler if we used a variety of sentiment analysis approaches and algorithms. Fraud can be divided into many categories based on data mining applications.

Reviews of each individual app are insufficient to identify if it is a scam or a legitimate app. As stated, star ratings are not entirely reliable because the developers themselves have the ability to modify them. Reading reviews is seen to be more important than ratings. In general, it is recommended to look at more trustworthy sources, including carefully chosen third-party reviews or the developer's other programs. gathering a certain app dataset over time and classifying the reviews as either favorable or negative. For the reviews' semantic categorization accuracy, using fewer words is more effective. It is simpler to categorize words using the suggested technique when there are fewer terms.

III. PROBLEM STATEMENT

The issue facing the Medical Center is the use of cards and files to track the progression of patient records, the hands-on method for providing health services, and the establishment of records for new patients in the hospital. There are numerous irregularities in this process, such as patients misplacing their files or losing their health cards. As a result, the automated system makes it simpler to determine new and ongoing status by keeping track of patient records and medical bills. By giving doctors and administrators the ability to see the prevalence of common diseases and their percentage, this program enhances the quality of medical assistance services. Managers can also determine the number of patients a doctor has visited over a set period of time

IV. THE PROPOSED SYSTEM

New apps can be added and created by the administrator, who can also include links to the apps on the Play Store or App Store. For that particular application, a set of data is gathered from both stores and stored in the database for a predetermined amount of time. The data provided by the user is cleaned using a variety of data pre-processing techniques. The usage of tokenization, stop word removal, and stemming techniques allows for a logical visualization of the architecture. In this case, user reviews and comments are kept in the database and serve as the algorithm's input. Positive and negative words that show up in reviews are now counted. A good attitude is returned by the system if there are more positive word appearances than negative ones, and vice versa. The system will return a neutral emotion if the numbers are equal. The SVM classifier will now be fitted to the training set. We will import the SVC class from the Sklearn.svm package in order to construct the SVM classifier. Since we are building an SVM for linearly separable data, we have chosen to use kernel='linear'. For non-linear data, we can alter it. Next, we used the training dataset (x_train, y_train) to fit the classifier. By adjusting the values of C (Regularization factor), gamma, and kernel, the model's performance can be changed.

Forecasting the outcome of the test set We shall now forecast the test set's output. We will make a new vector called y_pred for this.

Once we get the y_pred vector, we can compare the y_pred and y_test results to see how much the actual value differs from the anticipated value. In contrast to the logistic regression classifier, we will now examine the SVM classifier's performance in terms of the number of wrong predictions. We must import the sklearn library's confusion_matrix function in order to generate the confusion matrix. We will use a new variable, cm, to call the function after it has been imported. The function accepts two parameters, primarily y_true (the actual values) and y_pred (the value that the classifier is trying to return).

FRONT-END

In front-end coding, we use React js for creating the web page. Material Ui is used to give proper design to page with CSS, it is used to control the layout of multiple web pages all at once and JavaScript is used for providing functioning to elements

BACK-END

Python is used to create sentiments from the comments provided by user. As our project is based on Machine learning in which Sentiment Analysis is one of its subtract,Python is very powerful and manageable language as world is moving towards Machine Learning and other aspects of it.

V. WORKING MODEL

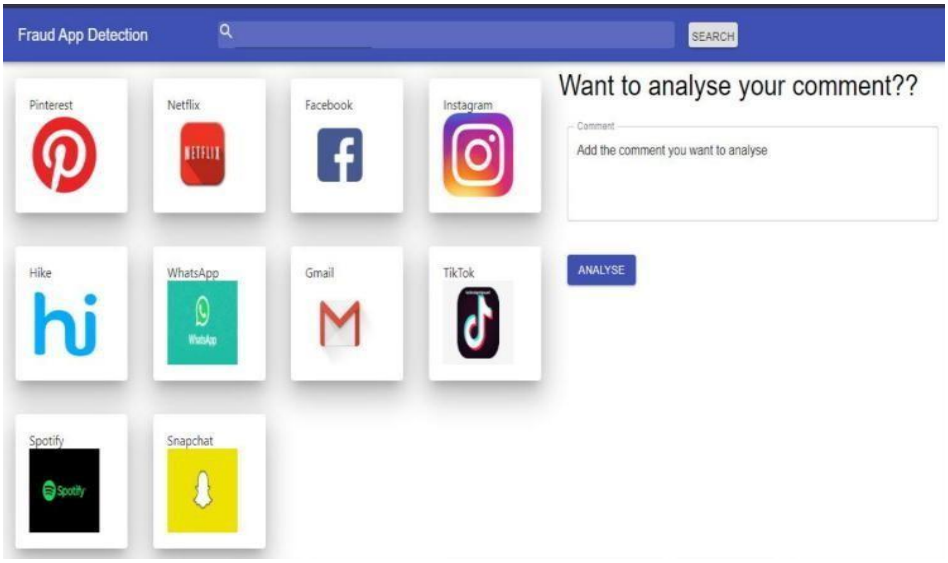

Fig 1

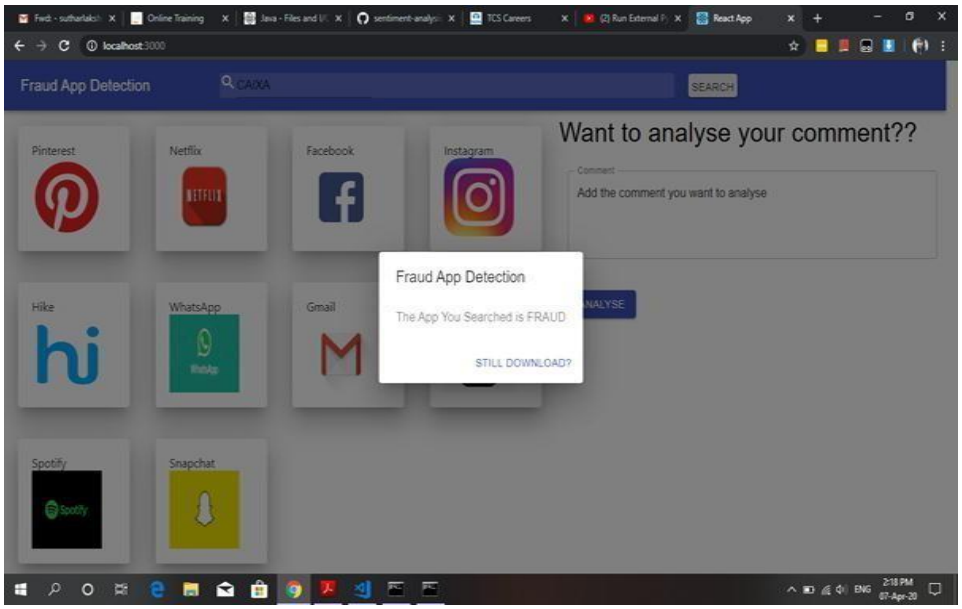

Fig 2 – Fraud App Detecting

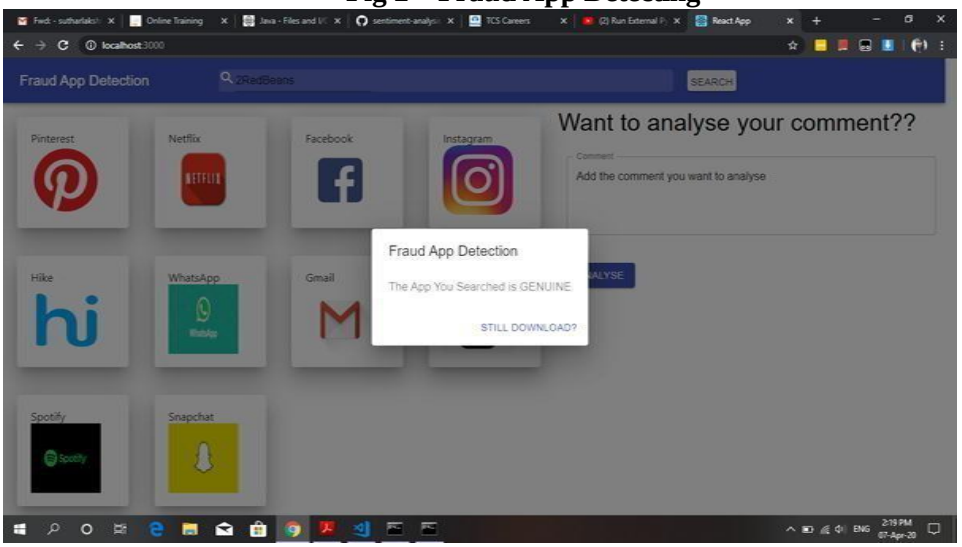

Fig 3 – Original App Detecting

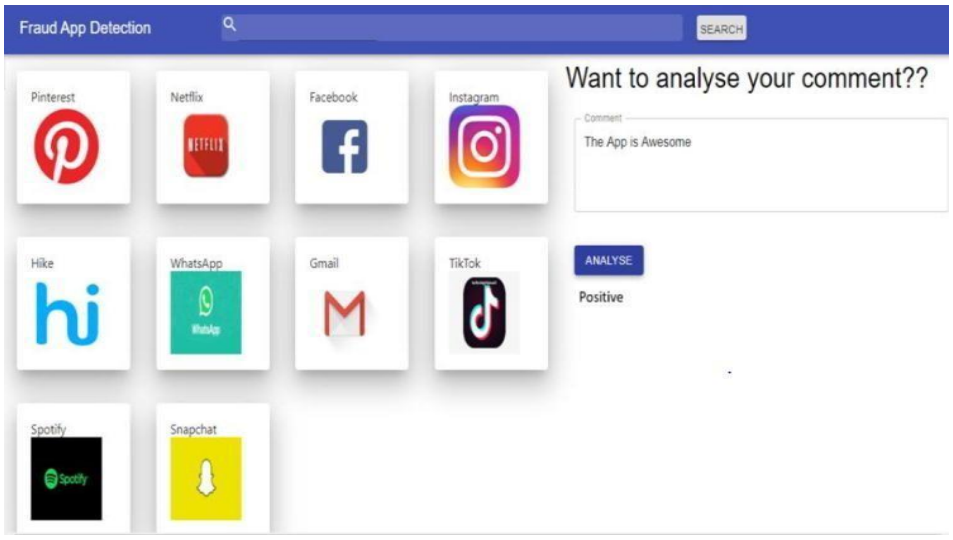

Fig-4: Review Your comment result (Additional Feature)

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.

ACKNOWLEDEMENT

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 of the 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 of the

REFERENCES

[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 Address Recognition 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 Parallel Machine 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