MobSafe: Forensic Analysis For Android Applications Using Cloud Stack And Data Mining.

Rohini Patil, Sucheta Kudale,Pornima Jathade,Pallavi Kale,Prof.Pankaj Agarkar

Abstract— **We do use many android application that are available on the internet .But the user doesn't get a true feedback about applications. The user can comment a wrong or fake feedback so, we are making an application so that we can get a review or ranking from different two websites. For that product ,we will analyze them with the help of NLP for positive and negative testing. Considering Web Services as an interface, an apk file gets interfaced with the dynamic web application. By applying NLP, it takes sentiments into consideration and do analysis for that. By applying rule for detection of fraud application, it generate the graph result. If the rating count is greater than 3 then it is considered as a positive result. And if the rating count is less than 3 then it is considered as a negative result. We propose the system to develop a android app which help out to detect fraud apps using cloud stack and data mining.**

Keywords- **Mobile Apps, ranking fraud detection, historical ranking records, rating and review,, natural language processing (NLP).**

I. INTRODUCTION TO WORKING SYSTEM

In the working system, when we want to download application from Google Play store that time we are taking APK of that application. And extracting details of the APK so that we can apply our rule to identify the fraud. When we are extracting APK details, applying NLP algorithm to classify sentiments from which we will get positive and negative result.In extraction of data we are extracting rating and comments, if their result is >3 then result will be positive so that application is fraud free and if the result is <3 then result is negative and application is fraud.

## II. COMPARISON BETWEEN EXISTING AND PROPOSED SYSTEM

EXISTING SYSTEM:-

We use various mobile applications in day to day life and using various application. During using various mobile application if someone is trying to steal our mobile data we can never know if any of those application is stealing data

from your phone or tampering your phone's it is important to know about such activity and take action on this .

PROPOSED SYSTEM:-

Hence we propose a way through forensic analysis for keeping track of such unauthorized activity. By keeping the watch on each and every action of the application which is possible by extracting APK details of uploaded application and then applying rule on that application so that we can detect the fraud application by using NLP technique for sentiment analysis.

III. ARCHITECTURE DIAGRAM
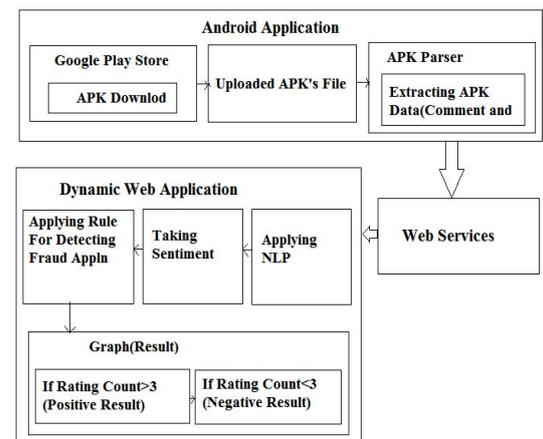
In this Architecture, whenever user wants to install any application, user downloads it from Google Play Store. It is in the apk file format. After downloading it, we upload the apk file. With the help of APK parser, the data gets extracted in

the form of comments, reviews, ratings and rankings. Considering Web Services as an interface, an apk file gets interfaced with the dynamic web application. In Dynamic Web Application, by applying NLP, it takes sentiments into consideration and do analysis for that. By applying rule for detection of fraud application, it generate the graph result. If the rating count is greater than 3 then it is considered as a positive result. And if the rating count is less than 3 then it is considered as a negative result.

IV. ALGORITHM

MAJOR TASKS IN NLP:-

The following is a list of some of the most commonly researched tasks in NLP.

A. AUTOMATIC SUMMARIZATION:-

It is used to provide a summary of a text for the articles.So that an article will be in a readable format.e.g financial section of a newspaper.

B. COREFERENCE RESOLUTION:-

If a sentence is larger chunk of text, so determine which words to the same objects.

C. DISCOURSE ANALYSIS:-

It is used to identify the discourse of the connected text. It can also recognize and classify the speech acts in a chunk of text.

D. MACHINE TRANSLATION:-

It translates the given text from one language to another.

E. MORPHOLOGICAL SEGMENTATION:-

It Separate the given word into single morphemes. The level of the task depends the morphology complexity.

f. Named entity recognition (NER):-
If we give a proper chunk of text then it determines which words map to the other proper names but in this only same words are capitalized.

G. NATURAL LANGUAGE GENERATION:-

It Convert from general computer databases into considerable human language.

H. NATURAL LANGUAGE UNDERSTANDING:-

It prepares the first order logic words that are understandable for computer programs.
.

I. OPTICAL CHARACTER RECOGNITION (OCR):-

It includes on image that has a printed text that refers to the mapping text.

J. PART-OF-SPEECH TAGGING:-

It gives the part of speech for each word. If can serve multiple words as well.

.

K. PARSING:-

It checks whether the language of the given grammer is ambiguous and analyze it.
.

L. QUESTION ANSWERING:-

If asked any question, it determine its answer .Sometimes more complex questions are taken into consideration.

M. RELATIONSHIP EXTRACTION:-

If identifies the relationship among named articles.

N. SENTENCE BREAKING (ALSO KNOWN AS SENTENCE BOUNDARY DISAMBIGUATION):-

It search for the sentence boundaries, mark by periods and characters are served.

O. SENTIMENT ANALYSIS:-

If Extracts the important information from a set of documents. For the marketing purpose, if analyzes it.

P. SPEECH RECOGNITION:-

It recognizes the speaking of a person and determine it.
.

Q. SPEECH SEGMENTATION:-

The whole sound clip separates it into words.

R. TOPIC SEGMENTATION AND RECOGNITION:-

It separate a word into segments and it is trivial in make language like english and separete by spaces.

S. WORD SENSE DISAMBIGUATION:-

A single word can have more than one meaning. In a list of words is given, the sets of related tasks are grouped into subfields of NLP.

T. INFORMATION RETRIEVAL (IR):-

The information retrieval is storing, searching and retrieving information. It is a separate field in computer science.

U. INFORMATION EXTRACTION (IE):-

If extracts the information from the given text which covers a task and recognizes the given entity.

V.   MODULES

1..COMMUNICATION :

In this module, we do select chatting applications or social apps that helps to decide whether that app should be installed or not.

2.SHOPPING:

In this, if user wants to get any details regarding shopping then it can be checked by comparing it on this application.
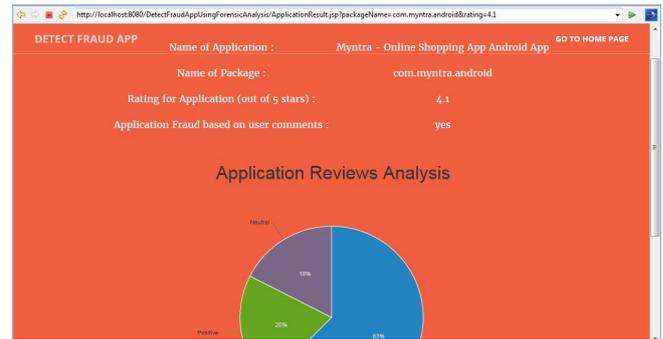
3.GAMING:

If any user wants to install any kind of game then by comparing the reviews, he/she can install it.

4.GENERAL:

This is the general category of the module in which all over data that is required is available. We can check and compare them accordingly.

VI.   RESULT ANALYSIS

Step 1:- Select Application Category.

Step 2:- Select Application Name.

Step 3:- More Shopping Application are Available.

Step 4:- Application Reviews Analysis.

Step 5:- Comparing Two Shopping Application.Select Two application.Then Decide Which is better.

Step 6:-Analysis in Two shopping Application.

VII.   CONCLUSION

We propose a system to evaluate the safety measures of Android mobile apps based on cloud computing and data mining. We propose the system to identify fraud apps for android mobile. This system we explain a android app which help to identify fraud apps. This system is based on cloud stack and data mining. To develop propose system we use natural language processing (NLP), and analyze them with

NLP for positive negative rating. We propose the system to expand a android app which help out to identify fraud apps using cloud stack and data mining.

REFERENCES

[1]Jaykumar Karnewa Snehal Umratkar. K-means algorithm for selective _ltration of malicious android mobile applications. International Journal of Advent Research in Computer , Electronics, 2014.

[2] R.C. Shivamurthy S.V. Nagendra Prasad Yadav. Faamac: Forensic analysis of android mobile applications using cloud computing. International Journal on Recent and Innovation Trends in Computing and Communication, 2014.

[3]Cloudstack project, http://cloudstack.apache.org, June, 2013

[4]CloudStack, A. Understanding Apache CloudStack. 2012 20/11/2013]; Available from: http://cloudstack.apache.org /software.html

[5]A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 632–640.

[6] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in Proc. 15th Int. Conf. World Wide Web, 2006, pp. 83–92.

[7] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[8] K. Shi and K. Ali, "Getjar mobile application recommendations with very sparse datasets," in Proc. 18th ACM SIGKDD Int. Conf.

[9]Yuan Zhang, Min Yang, Zhemin Yang, GuofeiGu, PengNing and BinyuZang "Permission Use Analysis for Vetting Undesirable Behaviors in Android Apps" in IEEE transactions on information forensics and security, vol. 9,no. november 2014.

[10] Yajin Zhou Zhi Wang Wu Zhou Xuxian Jiang"Detecting Malicious Apps in Official and Alternative Android Markets"in Department of ComputerScienceNorth Carolina State University 2011.