Detection of Fraud Ranking in Mobile Apps

Dikshita R. Chaudhari[1], Renuka D. Thakur[2], Ashwini V. Patil[3], Deepti S. Sharma[4]
[1, 2, 3, 4] Department of Computer Engineering
[1, 2, 3, 4] NES GCoE's, Nagaon,Dhule(MS)

Abstract- *Fraud ranking in the smart phone app means fraudulent or manual boosting of app ranking make it more popular. The app developers and the app owner's fraud way to bring up the ranking in mobile store list. This can include dedicated agencies for boosting the ranking or with the help of social network. This means that the mobile user who wishes to download app for specific life more difficult as due to wrong ratings. The user is misguided and he/she ends up app which actually is not very helpful but yet had a good ranking in app store. This makes user waste his/her time as well as his/her data bandwidth. These also make wrong impression in users mind for that particular apps category which is due to fraud mobile ranking. Considering this a proposed system will process and predict the fraudulent nature of reviews for specific app and present user with truer app ratings and rankings.*

Keywords- Fraudulent, Mobile App, Smart Phone, Text Mining etc.

I. INTRODUCTION

Increasing use of smart phones has increase the number of mobile applications. There is smart phone based application for e-commerce, food ordering, mobile games, health and safety, entertainment etc. There is large number of applications for each and every category. So with wide range of apps each and every app is given a rank on the mobile app stores. This rating and ranking are obtained by analysing the reviews of users for that specific app. Consider these scenario the app developers and app owner's are in race to increase their apps ranking in app store. To achieve this they often use fraudulent way such as boosting their ranking by bulk reviews. They also use hiring agencies which manually boost their app ratings. Even social network can be used to accomplish this. This all is done to increase the popularity of the app. Consider this scenario a large group of friends download and reviews a specific app and increase its popularity.

People also achieve by asking their employees for to rate and boost the ranking of that particular app. All this is good from marketing point of view but this does not help the app downloader who sees the ranking of app and then downloads it. He/She gets wrong information and he is misguided totally and ends up with either downloading a not so satisfying app whereas other apps were actually more helpful. This is also wrong from the point of view that few organizations are big and they can invest more in the marketing of their apps will end up with high ranking for their app in play store. On the contrary a small organization that actually had developed a very good app which is actually more useful for users ends up with low ranking. So the proposed system aims to detect these fraud rankings in mobile apps and present user with truer picture. The system would use text mining algorithms to evaluate the opinion level of a particular review. It would initially detect reviews fraudulent level and according to it would rank the app. It would also implement ways by which a sudden increase in rating and sudden boosting of app ranking can be detected.

II. PROPOSED SYSTEM

1. System Architecture

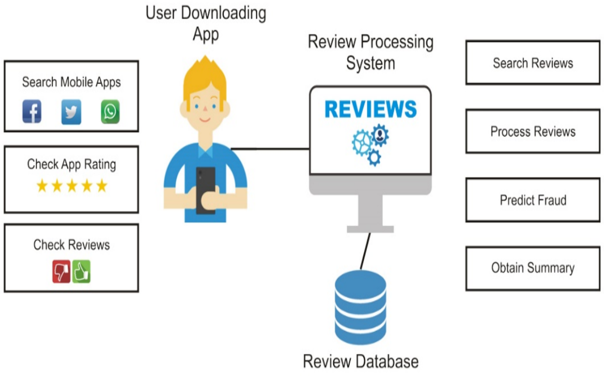


Figure 1 System Architecture

Considering the problems user face due to mobile app fraud reviews the proposed system aims to minimize the fraud review rating. The proposed system is a client server architecture where user will search for review and rating of specific mobile app from his smart phone and server will process and return the appropriate ratings and review of the application requested by user. User application will be an android application. The server will make sure that all the fraudulent reviews are analysed and user is provided with truer results. All the reviews related to a specific application will be stored in a centralized database along with reviewer information such as names, organization etc. The user can also review and rate any application from his android mobile application. The system will also make sure that only one review arrives from any mobile for that system will track users ID. The server will perform a systematic check on the reviews to find review fraud it would include the following:

1) Checking the reviewer in the list of developers other similar application. If the reviewer is common in most of the application of the developer whose application is under processing the reviewer reviews will be given a less weightage.
2) Checking for sudden rise in positive reviews for a particular application: Mostly when developers or owners use fraudulent means to boost their app rating it is found that the reviews arrive in a very short frame. Server will also process and find such pattern and these reviews will be considered as fraudulent
3) Checking for common terms in reviews: Mostly when developers or owners boost their rating it is found that reviews contains common terms as the reviews are more likely from common source. If this is the case these reviews will also be given less weightage.
4) Checking for duplicate reviews: Many times to boost the rating duplicate reviews are been posted. These reviews will be found out and filtered.

Finally once it’s made sure that all the reviews under consideration are been given right weightage according to their trueness polarity of the reviews is found out by comparing the reviews with positive and negative words corpus and the results are returned to user.

2. Algorithm Used

1) Term Frequency: Term frequency is a weighting scheme that refers to the assignment of weight to each term in the document that depends on the number of occurrences of the term in that review. The term frequency is denoted as ({tf}_{t,d}) with the subscripts denoting the term(t) and document(d),based on the weight of term(t) in document(d). Equation for Term Frequency (tf) is given as: Wt=ct log (N/ft) Where Wt is the weight of term ft is the number of times the term in the mail, ct is number of times the term in the passage, N is the total number of terms in the review.

2) Inverse Document Frequency: The inverse document frequency is a measure of how much information the word provides, that is, whether the term is common or rare across all documents. It is the logarithmically scaled inverse fraction of the documents that contain the word, obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient.

Idf (t, d) =log (N/ total number of documents where the term t appears)

N=Total number of documents in corpus.

3) Cosine Similarity: - Cosine Similarity falls under the category of distance measures. They are used to find how similarly two documents are from each other. Proposed system will use distance measures to find similar comments. The input to this equation is the tf/idf score of each comment. The output of this equation is an error value which denotes how distinct these comments are from each other.
4) Porter Stemmer: - Porter stemmer is a standard algorithm which is used to find root word of a given input word. The stemmer algorithm removes the tense from the word and makes it a normal word. Stemming the word increases the similarity evaluation accuracy. The proposed system uses this algorithm in the pre-processing phase.

III. IMPLEMENTATION

The system is client server architecture. System uses servlets technology to communicate between client and the server. The system includes following major modules:

1. Admin Module: This module is been implemented in java using Net beans editor. It is a desktop application. The admin can manage the filtering rules and manage stop words dataset from this module. The software is protected with password and username which is only provided to the legitimate user. The entire processing part and the filtering rules are managed from this end.
2. Developer Portal: This module is been implemented in java using Net beans editor. This module is for the developer who needs to upload their application. The developer can register from her and will be provided with a username and password. The developer can edit his/her profile and upload new application.
3. Web Server: The server module is responsible for maintaining the entire database and perform the analysis part. Server has the access to database and it manages the entries in the database. This module is been developed in java using Net beans editor. The server consists of servlets which are called for

appropriate purpose. The module uses JDBC library to connect to database.

4. Android Application: The application is developed in Android Studio editor. This application is used by the users to search applications and view the analysis of that application. The application gives a network call to the server when user searches for any application. The user is able to post comments and also rate a specific app via this application.

IV. RESULT AND ANALYISIS

1. Administration

The admin manage the filtering rules and stop words dataset from this mode.

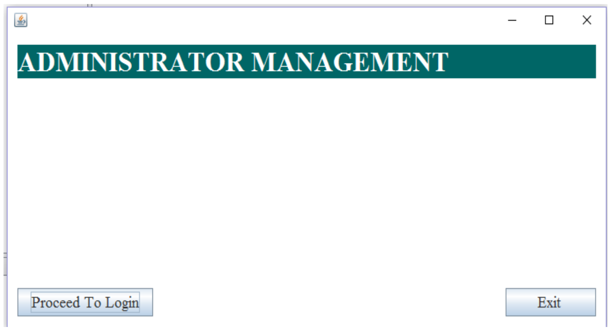


Figure 2.

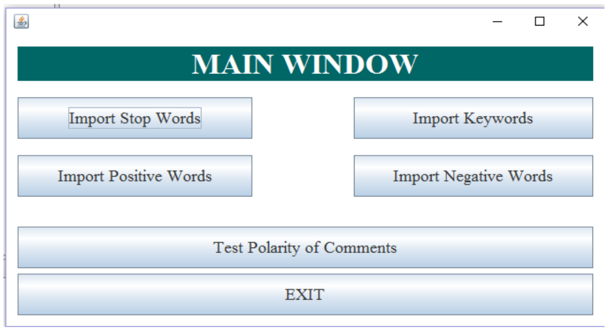
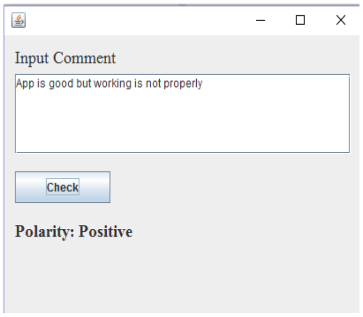


Figure 3.

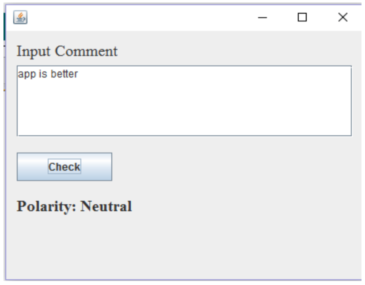
When we click on the “Proceed To Login” button then open the “Login Window” for Admin. Then admin type “Admin ID” & “Password” then click on the “Login” button. Then open the next window which is “Main Window Form”.In “Main Window” contains the words files like “Positive”, ”Negative”, ”Stop Words”, ”Keywords”. In this also check the “Polarity Of Comments”. When we click on the “Test Polarity Comments”button then the following window will be open:


Input Comment

App is good but working is not properly

Polarity: Positive

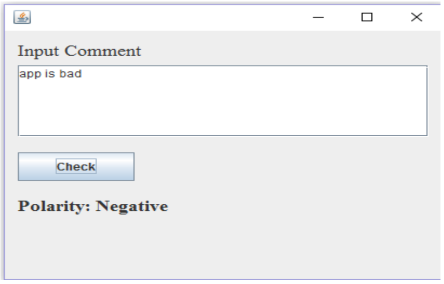
Figure 4.


Input Comment

app is better

Polarity: Neutral

Figure 5.


Input Comment

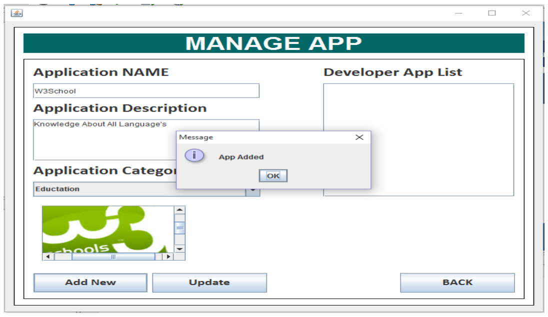
app is bad

Polarity: Negative

Figure 6.

2. Developer Portal

This module is for the developer who needs to upload their application. In this also the “Developer Portal” form in which also have the “Proceed To Login” button. When we click on this button then open the “Developer Login” window in which “Developer ID” and “Password” filed. If the developer already registered then only type his/her id and password then click on “Login” button and if the new developer then click on the “New Developer? Click To Register” button and fill the information about developer and click on the “SinUp” button.



MANAGE APP

Application NAME: W3School

Application Description: Knowledge About All Language's

Application Categor: Eductation

Developer App List

Message: App Added

Figure 7.

When authorized developer click on the “Login’ button then open the “Main Window” in which “View Profile” and “Manage App” options are provided. In “Manage App” Form contains the information about the regarding app..

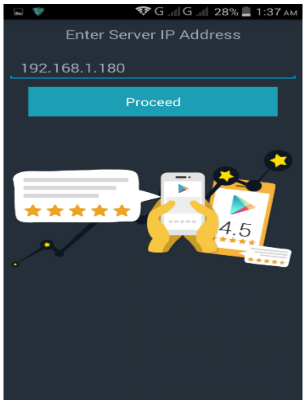
3. Android Application

This application is used by the users to search applications and view the analysis of that application.The application gives a network call to the server when user searches for any application. The user is able to post comments and also rate a specific app via this application.

Firstly user application is connected to the wireless network with server IP address then create account for searching various application. After creating account then login with “Username” & “Password” after that search the application then give the comment & rating about these application.

In our application shows the comment in 4 sessions which are “Overall Analysis”, “Leading Session Analysis”, “After Leading Session Analysis”, “Similar Positive Reviews” & “Similar Negative Reviews”.

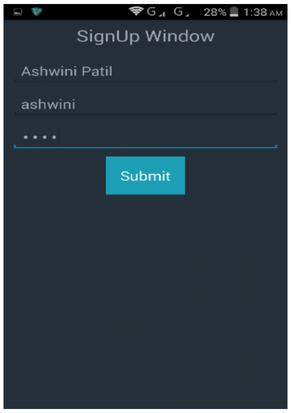
The important part of our project is that the only one user can create his/her account on mobile .This concept is managed by “IMEI Number” of mobile which is unique identity for every mobile.



Enter Server IP Address

192.168.1.180

Figure 8.



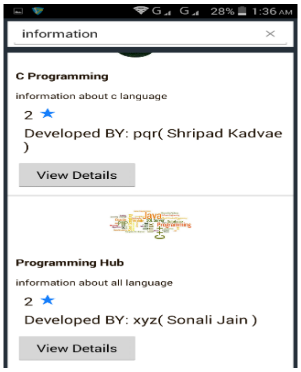
SignUp Window

Ashwini Patil

ashwini

••••

Figure 9.



information

C Programming

information about c language

2 ★

Developed BY: pqr(Shripad Kadvae)

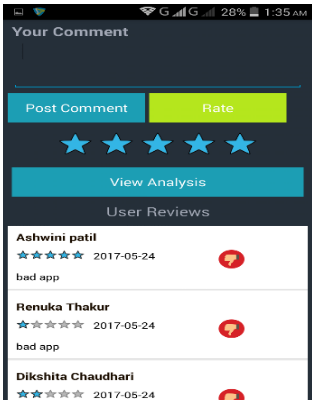
Programming Hub

information about all language

2 ★

Developed BY: xyz(Sonali Jain)

Figure 10.



Your Comment

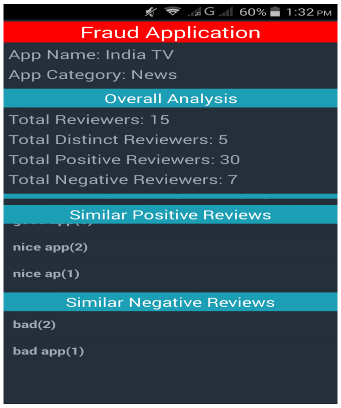
User Reviews

Ashwini patil
★★★★★ 2017-05-24
bad app

Renuka Thakur
★☆☆☆☆ 2017-05-24
bad app

Dikshita Chaudhari
★★☆☆☆ 2017-05-24

Figure 11.



Fraud Application

App Name: India TV
App Category: News

Overall Analysis

Total Reviewers: 15
Total Distinct Reviewers: 5
Total Positive Reviewers: 30
Total Negative Reviewers: 7

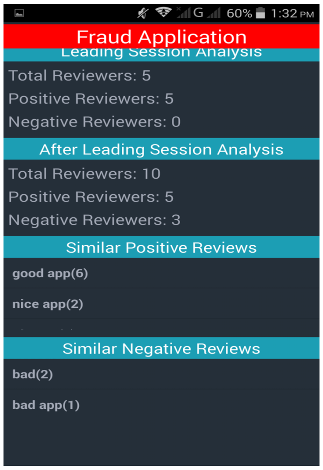
Similar Positive Reviews

nice app(2)
nice ap(1)

Similar Negative Reviews

bad(2)
bad app(1)

Figure 12.



60% 1:32 PM

Fraud Application

Leading Session Analysis

Total Reviewers: 5

Positive Reviewers: 5

Negative Reviewers: 0

After Leading Session Analysis

Total Reviewers: 10

Positive Reviewers: 5

Negative Reviewers: 3

Similar Positive Reviews

good app(6)

nice app(2)

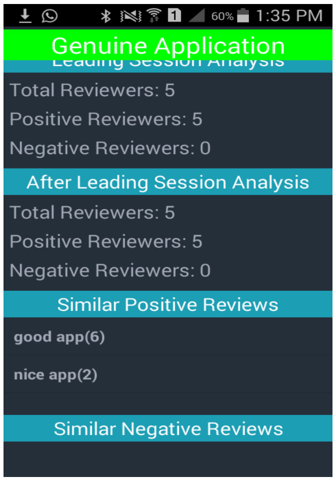
Similar Negative Reviews

bad(2)

bad app(1)

Figure 13.

Our app detect the fraud with the help of the ledding session & after ledding session and display the msessage “Fraud Application”. If the app is not fraud then display the meassage “Genuine Application”.



60% 1:35 PM

Genuine Application

Leading Session Analysis

Total Reviewers: 5

Positive Reviewers: 5

Negative Reviewers: 0

After Leading Session Analysis

Total Reviewers: 5

Positive Reviewers: 5

Negative Reviewers: 0

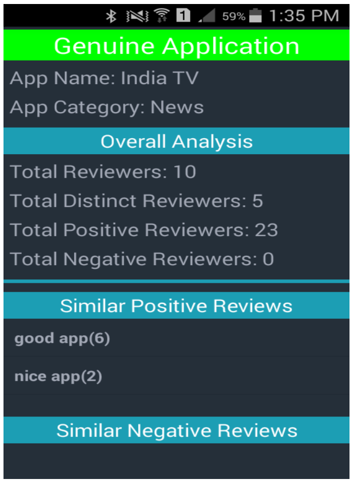
Similar Positive Reviews

good app(6)

nice app(2)

Similar Negative Reviews

Figure 14.



59% 1:35 PM

Genuine Application

App Name: India TV

App Category: News

Overall Analysis

Total Reviewers: 10

Total Distinct Reviewers: 5

Total Positive Reviewers: 23

Total Negative Reviewers: 0

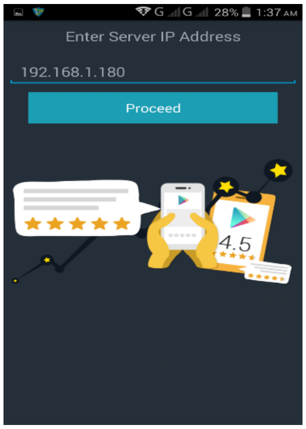
Similar Positive Reviews

good app(6)

nice app(2)

Similar Negative Reviews

Figure 15.



28% 1:37 AM

Enter Server IP Address

192.168.1.180

Figure 8.

V. CONCLUSION

Now a days, many of mobile app developers uses various fraud techniques to increase popularity of their apps . To avoid false popularity of the mobile application, this paper proposes a methodology based on data mining concept on user reviews and gives truer analysed reviews about app.

REFERENCES

[1] Marc Najork, Dennis Fetterly,”Detecting Spam Web Pages through Content Analysis”,Microsoft Research, USA

[2] Nikita Spirin, Jiawei Han,”Survey on Web Spam Detection: Principles and Algorithms”Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801, USA spirin2

[3] Bin Zhou, Jian Pei, “A Spamicity Approach to Web Spam Detection”Simon Fraser.

[4] Ee-Peng Lim, Viet-An Nguyen “Detecting Product Review Spammers using Rating Behaviors”, School of Information Systems, Singapore Management University 2010

[5] Zhiang Wu, Junjie Wu, Jie Cao, and Dacheng Tao,“HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation”Centre for Quantum Computation & Intelligent Systems, University of Technology,

[6] SihongXie, Guan Wang, Shuyang Lin, Philip S. Yu”Review Spam Detection via Temporal Pattern Discovery”KDD’12, August 12–16, 2012, Beijing, China

[7] Kent Shi,Kamal Ali,”GetJar Mobile Application

Recommendations with Very Sparse Datasets”KDD’12, August 12–16, 2012, Beijing, China.

[8] Bo Yan and Guanling Chen “AppJoy: Personalized Mobile Application Discovery” Department of Computer Science, University of Massachusetts Lowell 1 University Avenue, Lowell, MA 01854 2011

[9] ranjitha. r, mathumitha.k, meena.s"discovery of ranking of fraud apps",international journal of innovative research in engineering & management (IJIREM) ISSN:2350-0557, may-2016

[10] AlexandrosNtoulas,MarcNajork,MarkManasse, Dennis Fetterly “Detecting Spam Web Pages through Content Analysis”WWW 2006, May 23–26, 2006, Edinburgh, Scotland