International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 4, June 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

# Fraud Malware Detection in Google Play Store

V. Booma, G. Gayathri (MCA M Tech M Phil)

PRIST University Thanjavur, Tamil Nadu, India

## ABSTRACT

Android mobile applications become an simple target for the attacker because of its open source background. Also user' lack of knowledge the process of installing and usage of the apps. To categorize fake and malware apps, all the earlier methods listening carefully on getting permission from the user and executing that particular mobile apps. A malware detection structure that discover fraudulent developer, to detect search rank fraud as well as malware in Google Play Store. The fraud application is detected by aggregate the three pieces of proof such as ranking based, co-review based and rating based evidence. It combine efficiently for all the evidence for fraud detection. Detect fraud ranking in daily Apps head board. Avoid ranking manipulation. In the proposed system the detecting of normal and harmful application is analyzed by the **SVM Algorithm.** This system will analyze the uploaded application that are to be classifying the status which is dangerous application or normal application. The client can view the both the normal and harmful apps in ASP.NET. They can download the application after screening the secret manner. After using the apps the client can give the review on that downloaded apps. By the known review post for any application the admin will analyze the ASP.NET application for giving the ranking. The reviews are analyzed by the **SVM Algorithm.**

***KEYWORDS:*** *Data mining, Malware Detection, Support vector Machine (SVM)*

***How to cite this paper:*** V. Booma | G. Gayathri "Fraud Malware Detection in Google Play Store" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-4, June 2020, pp.1730-1734, URL: www.ijtsrd.com/papers/ijtsrd31686.pdf

**IJTSRD31686**

## I. INTRODUCTION

Data Mining is a knowledge mining process. It is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in bulky data sets involving method at the intersection of artificial intelligence, machine learning, statistics and database systems. The tremendous growth of scientific databases put a lot of challenges before the research to extract useful information from them using traditional data base techniques. Hence effective mining methods are important to discover the implicit information from huge databases. Cluster analysis is one of the major data mining.

Google play first discharges its application in 2008.Since that it conveys applications to all the Android clients. In Google Play Store, it gives benefits that client can find the specific application, buy those applications and introduce it on their cell phones. Since Android is open source condition all the insight about the application clients can be effectively gotten to by the application engineers through Google play. In Google play 1.8 Million versatile applications are accessible and that is downloaded by more than 25 billion clients over the world. This prompts more noteworthy possibility of introducing malware to the applications that could influence clients cell phones. Google play store utilizes its own particular security framework known as Bouncer framework to expel the malignant applications from its store. Nonetheless, this technique isn't successful as testing some applications utilizing infection instruments numerous applications are found as noxious which are not identified by Bouncer framework . False designers utilize look positioning calculation to elevate their applications to the best while seeking. In the wake of downloading versatile applications from Google play clients are requested to give the appraisals and surveys about those specific downloaded applications. However deceitful engineers give counterfeit evaluations and audits about their application elevate their application to the best. There are two ordinary methodologies utilized for distinguishing malware in Google Play. In this way are Static and Dynamic. The dynamic approach needs applications to be keep running in a protected situation to identify its benevolent. The static approach isn't utilized as the need to give a wide range of assault in beginning period itself however that is unthinkable as ordinary assailants locate the better approach to infuse malware on applications.

## II. LITERATURE SURVEY

In this project [1] the creator proposed another technique to recognize malware in versatile applications by analyzing the runtime conduct of that specific application in the portable condition. The creator recommends that surprising conduct versatile application can shift from one application to different applications. Additionally, it changes from nature of that specific application running on various gadgets. Utilizing Xposed structure client can change the client and framework application without adjusting the application package (APK). Depend upon that client can set specific conditions to recognize the malware in the portable applications.

In this project [2] the creator proposes some of present day machine learning calculations to recognize malware. For that these calculations are connected to the metadata gathered from the Google Play. While the majority of the current strategies for distinguishing calculation concentrated on inborn qualities of the specific portable application this gives an immediate technique to recognize the applications. For the setup of the trials the gathered 25k information from Google Play. Designers refresh their applications specifically interim of days while counterfeit applications couldn't be refreshed since its transfer of the Google Play. These works concentrated on just straight models Future work may concentrate on non-liners models.

In this project [3] the creator proposes the static strategy to identify the malware in portable applications. In this framework utilizing figuring out idea the source code for the suspicious APK documents. After that utilizing organized mapping creator manufactures the structure for the classes. At long last utilizing information stream idea a few examples for the distinctive sort of dangers has been made and utilize them to distinguish the malware in applications. Contingent on the quantity of threading design the viability of this technique is computed.

In this project [4], creator proposed novel procedure for processing a rank conglomeration based on network fruition to maintain a strategic distance from clamor and deficient information. Proposed technique takes care of an organized framework finish issue over the space of skew-symmetric grids. The creator demonstrates a recuperation hypothesis specifying when proposed approach will work. They additionally play out a point by point assessment of proposed approach with engineered information and a recounted think about with Netflix evaluations. To discover the arrangements, they used the svp solver for grid fulfillment. Rank collection is joined with the structure of skew-symmetric networks. Creator connected for most recent advances in the hypothesis and calculations of framework fulfillment to skew-symmetric networks. Creator upgraded existing calculation for grid finishing dealing with skew symmetric information.

In this project [5] the creator plans to secure the audit spanners or spam surveys. The spammer may target just on the particular ensure. From that point forward, they gave counterfeit audits to that specific versatile application by making the distinctive record to survey that record. The creator proposes a novel based scoring strategy to recognize each and every audit of the specific item. The creator makes very suspicious as a subset. By utilizing online spammer assessment programming the phoniness of the survey is ascertained. After the fulfillment of the assessment, the outcome demonstrates the compelling to anticipate the phony audits.

## III. EXISTING SYSTEM

It existing malware detection framework system that detects Google Play fraud and malware. To detect fraud and malware, it proposes the incremental learning approach to characterize the dataset. It formulates the notion of review modeling by applying Porter stemmer algorithm. This use temporal session of review post times to identify suspicious review spikes received by apps; the application evidence such as rating, ranking and review evidence will be integrated by an unsupervised evidence-aggregation method for evaluating the credibility of leading sessions from mobile Apps. The malware detection structure is scalable and can be complete with other area generated proof for ranking fraud detection. It identified that for the detection of the rank ranking, rating, and review based evidence are considered. Moreover, it proposed an optimization based aggregation method to integrate all the evidence for evaluating the credibility of leading sessions from mobile Apps. Deceitful practices in Google Android application showcase fuel seek rank manhandle and malware expansion

**Disadvantages**

- ➢ Google Play uses the Bouncer system to note remove malware.
- ➢ Use risk signals extracted from app permissions
- ➢ A score to measure the risk of apps, based on probabilistic generative models such as Naive Bayes.
- ➢ The classification of application based on the keywords is not identified.

### 3.2. PROPOSED ALGORITHMS

In the proposed system the detecting of normal and harmful application is analyzed by the **SVM Algorithm.** This system will analyze the uploaded application that are to be classifying the status which is dangerous application or normal application. The client can view the both the normal and harmful application. They can download the application following presentation the classified approach. Following using the apps the user can give the review on that downloaded application. By the given review for any application the admin will analyze the application for giving the ranking. The reviews are analyzed by the **Collaborative Technique**.

**Advantages**

- ➢ Fraudulent and malicious behaviours leave behind telltale signs on app markets.
- ➢ Fair Play achieves over 97% accuracy in classifying fraudulent and benign apps, and over 95% accuracy in classifying malware and benign apps.
- ➢ Its also enabled us to discover a novel, coercive review campaign attack type, where app users are harassed into writing a positive review for the app, and install and review other apps

### 3.3. METHODOLOGY

**SVM**

SVM, a decently brand new sort of learning calculation, initially presented. Actually, SVM go for hyper plane incredible isolates the classes of information. SVMs has affirmed the ability not just too precisely isolate elements into right classes, additionally to recognize case whose build up arrangement is not upheld by information. In spite of the fact that SVM are nearly harsh characterize dispersion of preparing cases of every class. It is just reached out numerical figuring's. Two such expansions, the first is to stretch out SVM relapse examination, where the objective to deliver a direct capacity that can genuinely exact that objective capacity. An additional expansion is to figure out how to rank components as opposed to creating a characterization for individual components. Positioning can be decreased to looking at sets of case and delivering a +1 assess if the combine is in the right positioning request not withstanding −1 generally.

Algorithm for SVM:

```
Step1: Select candidate = {closest pair of opp class}
Step2: while there are violating points do
Step3: Find a violator
```

```
Step4: Candidate = Candidate U violator
Step5: if any z < 0 due to addition of c to s then
Step6: Candidate = candidate/p
Step7: repeat till all such points are pruned
Step8: End of if
Step9: End of while
```

## IV. EXPERIMENT AND RESULTS

The proposed system has been implemented using ASP.NET tool. This system will analyze the uploaded application that are to be classifying the status which is dangerous application or normal application. The user can view the both the normal and harmful application. They can download the application after viewing the classified manner. After using the application the user can give the review on that downloaded application. By this the user can identify the application by the review given. By the given review for any application the admin will analyze the application for giving the ranking. This proposed system will help to take the measures for the betterment of the end user. In the below table the execution second is analyzed between the existing and the proposed algorithm. The proposed SVM algorithm takes short time to execute.

### 4.1. USE CASE DIAGRAM

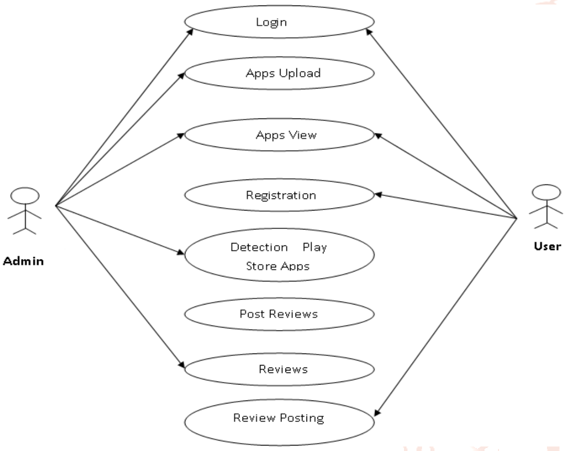
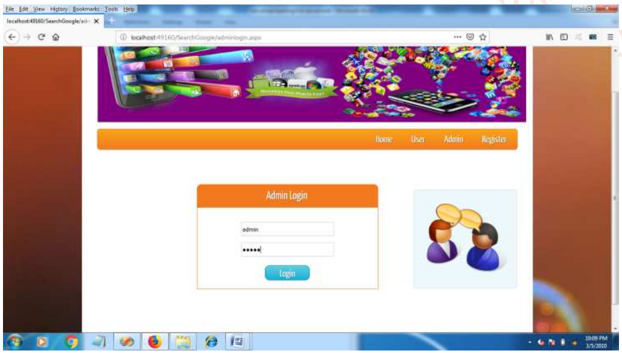


Figure 4.1 Use Case Diagram

### 4.2. RESULT

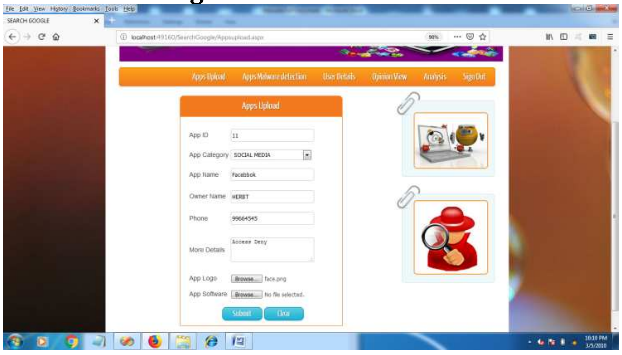

localhost:49160/SearchGoogle/adminlogin.aspx

Admin Login

admin

•••••

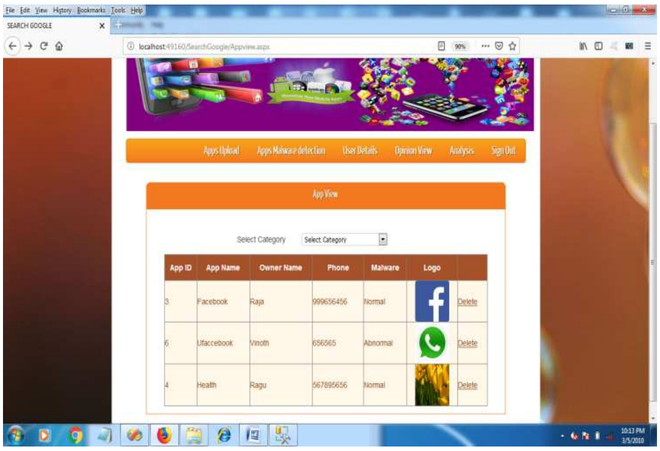
### 4.2.1. Admin Login


localhost:49160/SearchGoogle/Appsupload.aspx

Apps Upload

| | |
|---|---|
| App ID | 11 |
| App Category | SOCIAL MEDIA |
| App Name | Facebbok |
| Owner Name | HERBT |
| Phone | 99664545 |
| More Details | Access Deny |
| App Logo | Browse... face.png |
| App Software | Browse... No file selected. |

Fig: 4.2.2 App Upload

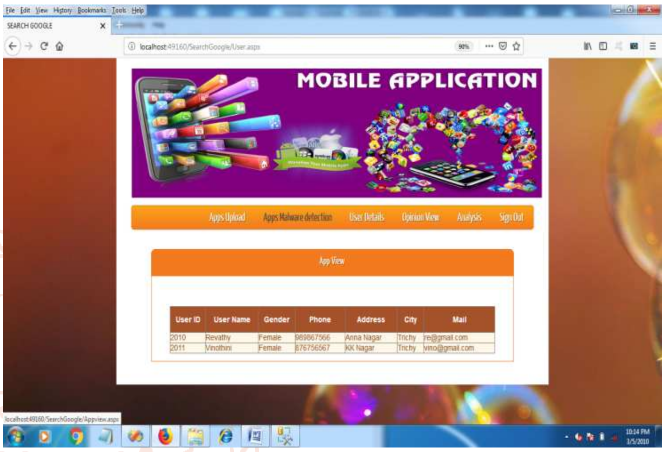

localhost:49160/SearchGoogle/Appview.aspx

App View

Select Category: Select Category

| App ID | App Name | Owner Name | Phone | Malware | Logo | |
|---|---|---|---|---|---|---|
| 3 | Facebook | Raja | 999656456 | Normal | | Delete |
| 6 | Ufaccebook | Vinoth | 656565 | Abnormal | | Delete |
| 4 | Health | Ragu | 567895656 | Normal | | Delete |

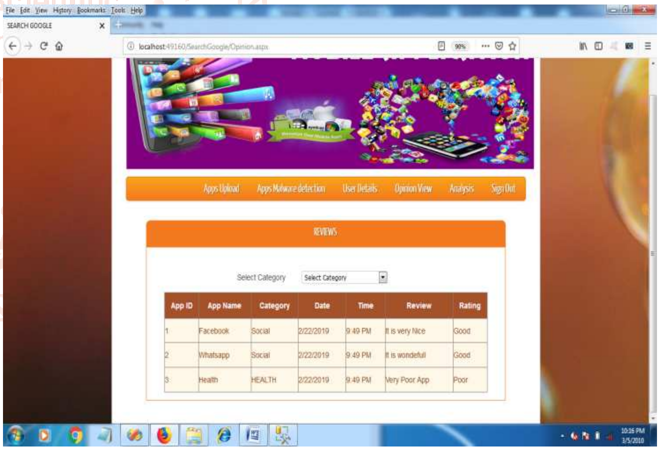
Fig: 4.2.3 App View


localhost:49160/SearchGoogle/User.aspx

App View

| User ID | User Name | Gender | Phone | Address | City | Mail |
|---|---|---|---|---|---|---|
| 2010 | Revathy | Female | 989867566 | Anna Nagar | Trichy | re@gmail.com |
| 2011 | Vinothini | Female | 876756567 | KK Nagar | Trichy | vino@gmail.com |

Fig: 4.2.4 User Details

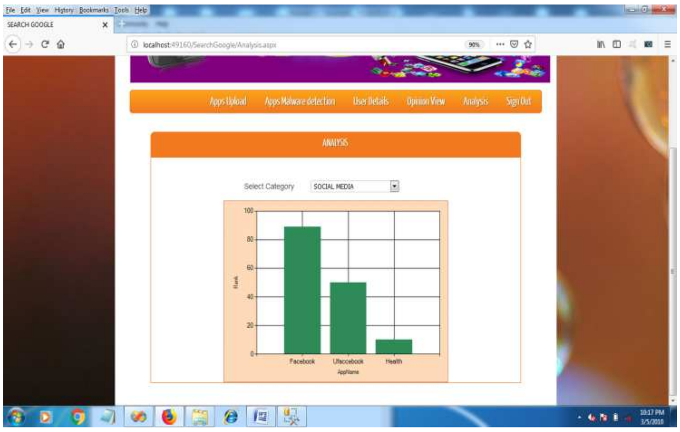

localhost:49160/SearchGoogle/Opinion.aspx

REVIEWS

Select Category: Select Category

| App ID | App Name | Category | Date | Time | Review | Rating |
|---|---|---|---|---|---|---|
| 1 | Facebook | Social | 2/22/2019 | 9:49 PM | It is very Nice | Good |
| 2 | Whatsapp | Social | 2/22/2019 | 9:49 PM | It is wondefull | Good |
| 3 | Health | HEALTH | 2/22/2019 | 9:49 PM | Very Poor App | Poor |

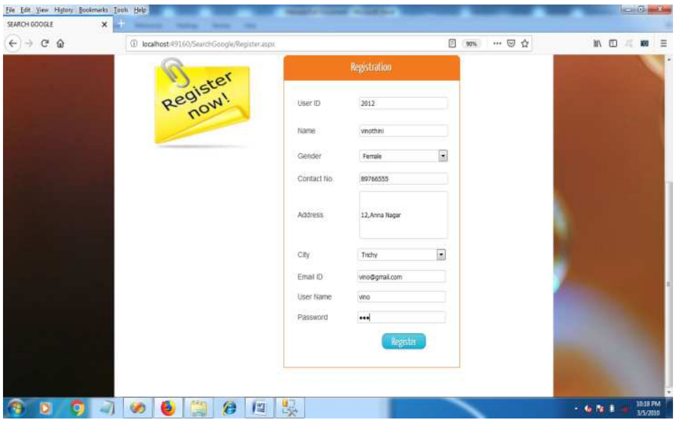
Fig: 4.2.5 Reviews


localhost:49160/SearchGoogle/Analysis.aspx

ANALYSIS

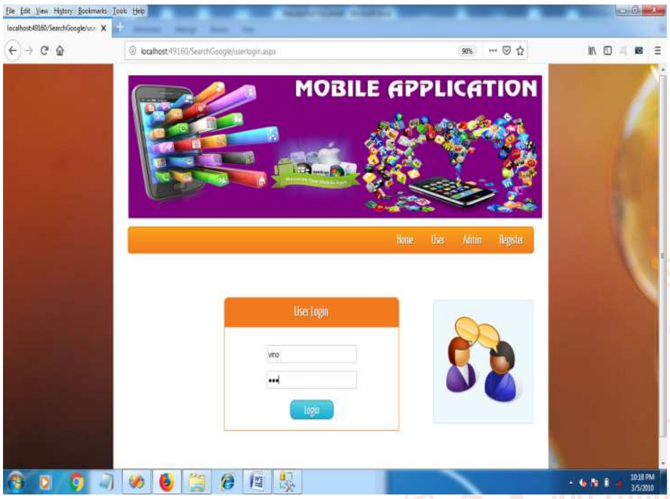
Select Category: SOCIAL MEDIA

Fig: 4.26 Apps Analysis SVM



User ID 2012
Name vinothini
Gender Female
Contact No 89766555
Address 12,Anna Nagar
City Trichy
Email ID vino@gmail.com
User Name vino
Password •••

Fig: 4.2.7 User registrations



vino
•••

Fig: 4.2.8 User Login

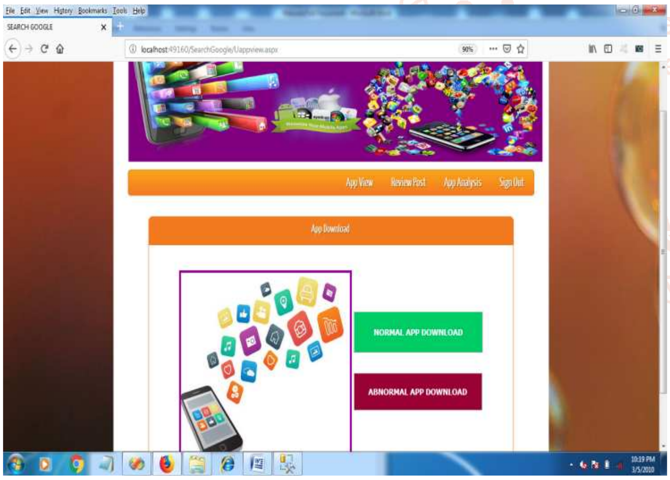
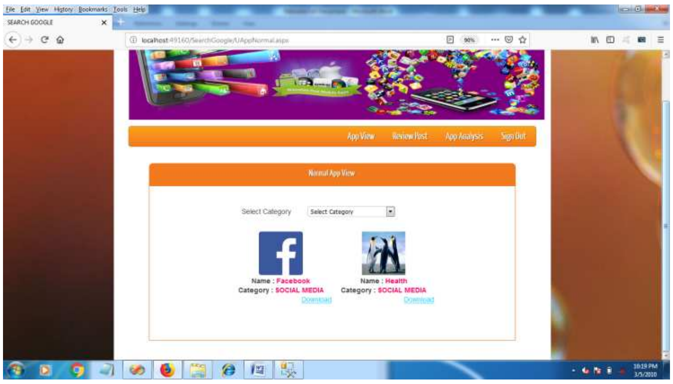


Fig: 4.2.9 Malware Detection

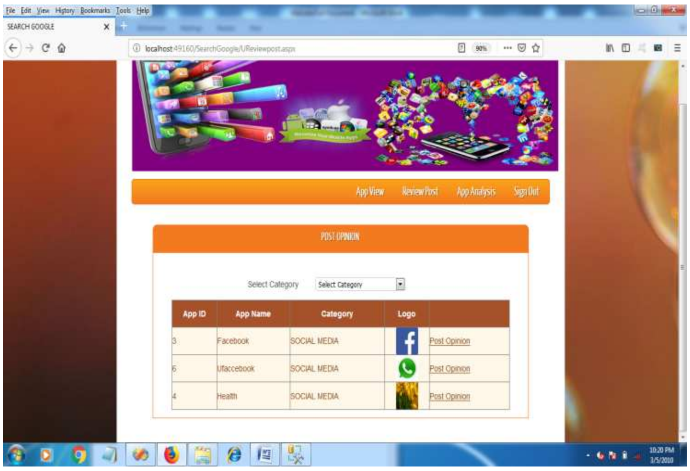


Select Category: Select Category

Name : Facebook
Category : SOCIAL MEDIA

Name : Health
Category : SOCIAL MEDIA

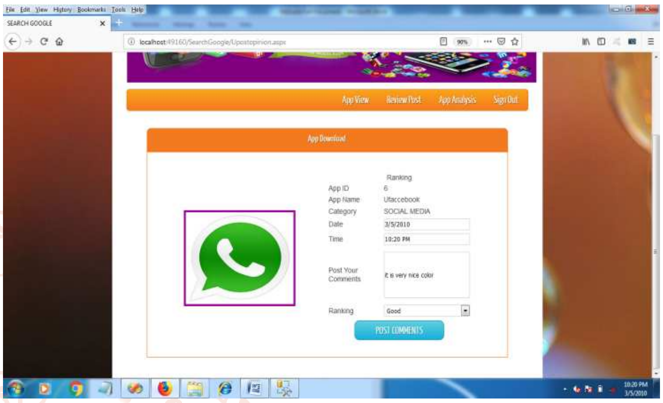
Fig: 4.2.10 App View



Select Category: Select Category

| App ID | App Name | Category | Logo | |
|---|---|---|---|---|
| 3 | Facebook | SOCIAL MEDIA | f | Post Opinion |
| 6 | Ufaccebook | SOCIAL MEDIA | ☎ | Post Opinion |
| 4 | Health | SOCIAL MEDIA | | Post Opinion |

Fig: 4.2.11 Post Opinion



App ID 6
App Name Ufaccebook
Category SOCIAL MEDIA
Date 3/5/2010
Time 10:20 PM
Post Your Comments it is very nice color
Ranking Good

Fig: 4.2.12 Post Opinion

## V. CONCLUSION

In this project developed a fraud detection system for mobile Apps. Specifically first showed that fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. An identified that for the detection of the rank ranking, rating, and review based evidence are considered. Moreover proposed an optimization based aggregation method to integrate all the evidence for evaluating the credibility of leading sessions from mobile Apps. Finally validate the proposed system with extensive experiments on real-world App data collected from the Apple's App Store. This project implemets results showed the effectiveness of the proposed approach.

## FUTURE ENHANCEMENT

In the future plan to study more effective fraud evidence and analyze the latent relationship among rating, review, and rankings. Moreover, will extend our ranking fraud detection approach with other mobile App related services, such as mobile Apps recommendation, for enhancing user experience.

## REFERENCES

[1] Alaa Salman Imad H. Elhajj Ali Chehab Ayman Kayss, IEEE Mobile Malware Exposed. International Conference on Knowledge discovery and data mining, KDD'14 pages 978- 983.

[2] Alfonso Munoz, Ignacio Mart 'ın, Antonio Guzman, Jos'e Alberto Hern ' andez, IEEE Android malware detection from Google Play meta-data: Selection of important features.2015, pages,245-251.

[3] Chia-Mei Chen, Je-Ming Lin, Gu-Hsin Lai, IEEE Detecting Mobile Application Malicious Behaviors Based on Data Flow of Source Code.2014 International Conference on Trustworthy Systems and their Applications pp 95-109.

[4] D. F. Gleich and L.-h. Lim. Rank aggregation via nuclear norm minimization. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’11, pages 60-68, 2011.Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.

[5] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM ’10, pages 939-948, 2010.

[6] N. Jindal and B. Liu, “Opinion spam and analysis,” in Proc. Int. Conf. Web Search Data Mining, 2008, pp. 219-230.

[7] J. Oberheide and C. Miller, “Dissecting the Android Bouncer,” presented at the SummerCon2012, New York, NY, USA, 2012.

[8] K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’12, pages 204-212, 2012.

[9] J. Kivinen and M. K. Warmuth, “Additive versus exponentiated gradient updates for linear prediction,” in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 2014.

[10] N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explor. Newsl., 13 (2):50-64,May2012.