https://doi.org/10.46335/IJIES.2021.6.3.5 **e-ISSN: 2456-3463**
Vol. 6, No. 3, 2021, PP. 21 -25

International Journal of Innovations in Engineering and Science, www.ijies.net

Detecting Fraud Apps Using Sentiment Analysis

Rashmi Jain[1], Rutuja Zatale [2], Shamli Kene[3], Himanshu Pote[4], Shruti Duske[5], Utkarsha Zamare[6]

[1]Assistant Prof., Department of CSE, Rajiv Gandhi college of Engineering Research, Nagpur, India
[2,3,4,5,6]UG Students, Department of CSE, Rajiv Gandhi college of Engineering Research, Nagpur, India

***Received on**: 28 April, 2021, **Revised on**: 24 May, 2021, **Published on**: 26 May, 2021*

***Abstract* –** *Nowadays, mobile apps are considered common place because of how popular and mainstream they are with mobile technology has become. Since there are so many mobile applications in the market, app rating manipulation is the most difficult problem for both the developers and consumers to prevent .Class expansionism is when malicious or disingenuous actions have the intention of artificially rising the numbers of apps in the rating. As a result, it is increasingly popular for application developers to use questionable strategies, like inflating revenue or posting false scores, nefarious ratings, to attempt to achieve a higher rank in the App store. That's why using statistical models to calculate, we also analyse three varieties of evidences: 1) In these cases, the models rank applications, and2) product ratings and quantitative evidences, and 3) product feedback may serve as quantitative evidence.*

I -INTRODUCTION

A vast number of smartphone applications has been developed over the past few years and is available as Downloads on phones as well For instance; there are over a Million applications on the App Store, while Google Play Already had over 1.6 million by the end of April 2013. For Better application creation, a placement, several app stores Have introduced the regular chart of most popular applications, Including Google Play, the App Store, and the Apple AppStore though small mobile Apps are essential for promoting Itself, this form of apps is the most effective ways. When your Company sits at the top of the company board, you are likely to see a large number of downloads and millions in sales. Because of this, the proliferation of different means of Application marketing, developers end up using approaches Such as their application rankings in search engines to Promote their apps of late mark expanding Software developers have Made extensive use of dishonest practices to unjustifiably Benefit their own applications by using illegitimate means. The final thing they do is to change the chart rankings on the app store. A variety of small teams of employees work tirelessly to drive the app to the top of the charts and bring it to bear on social media, both called "Internet bots" and "human water-dinosaur armies" support this process. As an example, Venture Beat reported that when an app was placed on the number one-two spot in the App Store and it had a slow-moving description, long rank could garner a substantial traffic in the highest of 50,000 to tens of thousands of new users in a couple of days. Although such rating manipulation has been favourably rewarded in the app industry, it increases a lot of doubts and worries for mobile app developers. The company has said that it will start cracking down on App developers who flout its quality standards by selling fake and Inflated app rankings in the App Store. From the varied types of mobile Apps and functions that serve different needs, mobile forms come in various types of leading (to be viewed, to be watched, to be checked, to be inspected,

https://doi.org/10.46335/IJIES.2021.6.3.5 e-ISSN: 2456-3463
Vol. 6, No. 3, 2021, PP. 21 -25

International Journal of Innovations in Engineering and Science, www.ijies.net

to be accessed). Apps that are available for smartphones and aren't necessarily at the top of the leader boards, however, by the way, happens Mobile forms come in various types of leading (to be viewed, to be watched, to be checked, to be inspected, to be accessed). Apps that are available for smartphones and aren't necessarily at the top of the leader boards, however, by the way, happens to be one of the most used apps the detection of Mob applications ranks will, therefore, detecting that there is an illegal application in the session of the mobile Apps is, in reality, the job of spotting them at the outset One may use this paper's algorithm to classify the front-running sessions from an existing mobile application's overall ranking records to determine which sessions should be ranked ahead of others. There is fraudulent proof in this example. A collection of fraud evidences was suggested by the concept of App Expanded and App's rating and background. These provide evidence of App expansion and fraud from trends in App's history and previous ratings.

II-LITERATURE SURVEY

The works within this analysis are classified in three groups, namely: the paraphrases by adjacent to this one, those below it, those that complement this one, and those that refer to it. In the first segment, we will talk about finding and removing spam on the Internet. Efforts that include unfair practice that puts unfair weighting of influence on particular Web pages within the context of the web structure, such as producing lists of favourable or negative mentions for paid inclusion in search engine results. A significant issue facing spam deterrence in the general Internet population is unsupervised identification of spam. A spam city score is implemented in order to help define whether or not a web page is considered to be spam. Measurement flexibility and adjustability are superior when classifying (a classification technique) according to users' input data, not by machine results. They provide an effective connection filter and an analysis system for spam detection that relies on spam. These procedures do not require preparation, but are easy and inexpensive to use. If it is necessary, a real, previously collected dataset is used to support the claims of that are made here. One way in which Ntoul et al. [2] have analysed different aspects of content-oriented spam on the web and given various techniques for identifying content-oriented spam is to look for suspicious patterns in web content. They perform “spam” examinations online to check for "spam dressing" pages to insert artificially. They look for “spam dynet” (spam inserted for manipulative reasons to boost rankings in search engines to increase page traffic to websites) inserted pages in the internet. The current paper looks at some previously unknown techniques and how these techniques can be used, and attempts to analyse the accuracy of these techniques in an aggregate setting. When trying to determine which sites are generating spam, they [the authors of the paper by Zhou et al. (Zhou et al.)] used the unsupervised technique of scanning the Internet to identify malicious activity. More specifically, they suggested powerful online word spam huffing and anti-spam methods using the spam city of the meanings. This survey goes over all that is known about spamming detection and jamming spam filtration in Web environments, and the results were presented recently by Spirinet al.

III- ARCHITECTURE AND WORKDONE

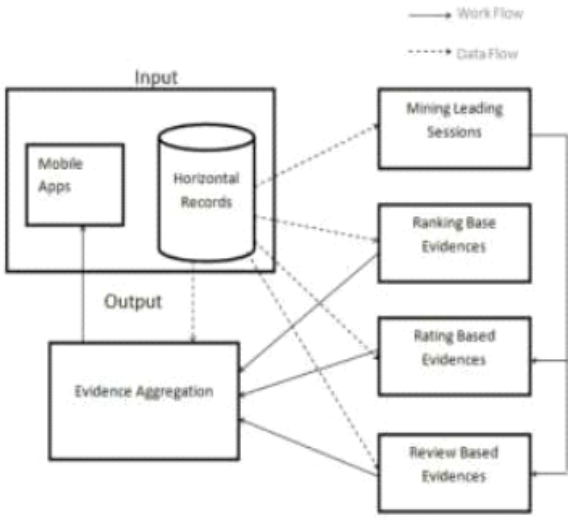

Figure 1. System Architecture

In this system architecture given above we need to consider various mobile apps and a dataset as a form of input. We have taken the dataset required from UCI machine learning repository and we have used the google play store dataset . The data that we have gathered we need to present it in a summarized format that’s why we do **evidence aggregation. The** data may be gathered from multiple data sources with the intent of combining these data sources into summary for data analysis. In **mining leading sessions** there are two types of concerns with mobile fraud apps. First, from the **apps historical ranking records**, discovery of leading events

https://doi.org/10.46335/IJIES.2021.6.3.5 **e-ISSN: 2456-3463**
Vol. 6, No. 3, 2021, PP. 21 -25

International Journal of Innovations in Engineering and Science, www.ijies.net

is done and then second merging of adjacent leading events is done which appeared for constructing leading sessions. For detecting that the app is fraud or genuine there are three main steps **Ranking based evidence, Rating based evidence, Review based evidence.**

Clustering- We divide dataset in different groups based on similarities, for this we have to use an algorithm which is "Hierarchical clustering algorithm"

1.Hierarchical Algorithm – A **Hierarchical clustering** method works via grouping data into a tree of clusters. Hierarchical clustering begins by treating every data points as a separate cluster. Then, it repeatedly executes the subsequent steps:

1. Identify the 2 clusters which can be closest together, and
2. Merge the 2 maximum comparable clusters. We need to continue these steps until all the clusters are merged together.

In Hierarchical Clustering, the aim is to produce a hierarchical series of nested clusters. A diagram called **Dendrogram** (A Dendrogram is a tree-like diagram that statistics the sequences of merges or splits) graphically represents this hierarchy and is an inverted tree that describes the order in which factors are merged (bottom-up view) or cluster are break up (top-down view).
The basic method to generate hierarchical clustering is **Agglomerative.**

Initially consider every data point as an **individual** Cluster and at every step, **merge** the nearest pairs of the cluster. (It is a bottom-up method). At first every data set is considered as individual entity or cluster. At every iteration, the clusters merge with different clusters until one cluster is formed.

- Algorithm for Agglomerative Hierarchical Clustering is: Calculate the similarity of one cluster with all the other clusters (calculate proximity matrix)
- Consider every data point as an individual cluster
- Merge the clusters which are highly similar or close to each other.
- Recalculate the proximity matrix for each cluster
- Repeat Step 3 and 4 until only a single cluster remains.

2.J48 Algorithm:

J48 **algorithm** is one of the best machine learning **algorithms** to examine the **data** categorically and continuously. When it is used for instance purpose, it occupies more memory space and depletes the performance and accuracy in classifying medical **data**.

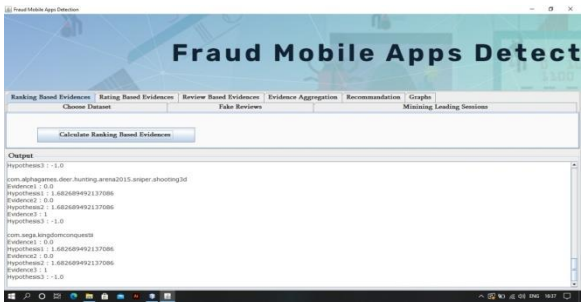

Fig.2- Ranking based Evidence

It concludes that leading session comprises of various leading events. Hence by analysis of basic behaviour of leading events for finding fraud evidences and also for the app historical ranking records, it is been observed that a specific ranking pattern is always satisfied by app ranking behaviour in a leading event.

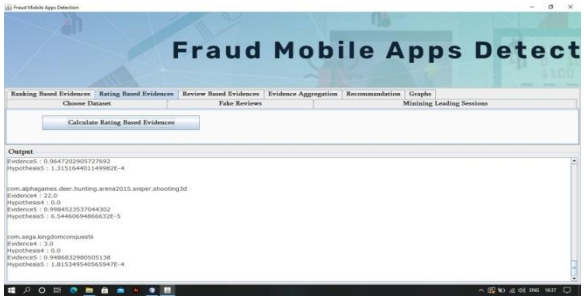

Fig.3- Rating based Evidences

Previous ranking-based evidences are useful for detection purpose but it is not sufficient. Resolving the "restrict time depletion" problem, fraud evidences recognition is planned due to app historical rating records. As we know that rating is been done after downloading it by the user, and if the rating is high in leader board considerably that is attracted by most of the mobile app users. Spontaneously, the ratings during the leading session gives rise to the anomaly pattern which happens during rating fraud. These historical records can be used for developing rating-based evidences.

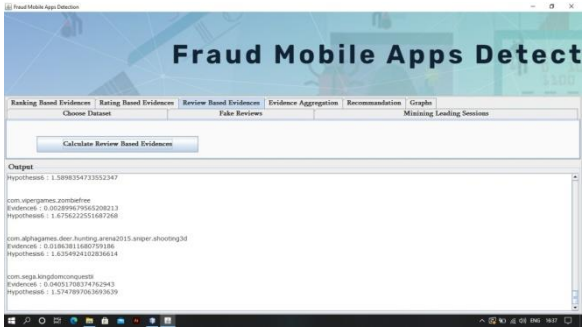

Fig 4- Review based Evidences

https://doi.org/10.46335/IJIES.2021.6.3.5 **e-ISSN: 2456-3463**
Vol. 6, No. 3, 2021, PP. 21 -25

International Journal of Innovations in Engineering and Science, www.ijies.net

We are familiar with the review which contains some textual comments as reviews by app user and before downloading or using the app user mostly prefer to refer the reviews given by most of the users. Therefore, although due to some previous works on review spam detection there still issue on locating the local anomaly of reviews in leading sessions. So based on apps review behaviours, fraud evidences are used to detect the ranking fraud in Mobile App. These three evidences will be integrated by an unsupervised evidence-aggregation method for evaluating the credibility of leading sessions from mobile Apps. The statistical hypotheses tests models Apps ranking, rating and review behaviours to extract all the evidences. The ranking fraud detection framework is scalable and can be extended with other domain generated evidences for ranking fraud detection. Finally, we will evaluate the proposed system with real-world App data collected from the Apple's App store for a long-time span, i.e., more than two years.

An **ARFF** (Attribute-Relation **File Format**) **file** is an ASCII text **file** that describes a list of instances sharing a set of attributes. **ARFF files** were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software.

Evidence aggregation is the process of gathering data and presenting it in a summarized format. The data may be gathered from multiple data sources with the intent of combining these data sources into a summary for data analysis.

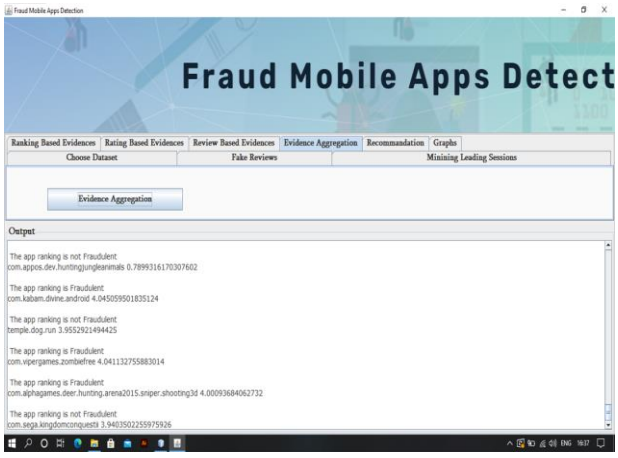

Fig 5- Evidence aggregation

IV-RESULT ANALYSIS & COMPARISON

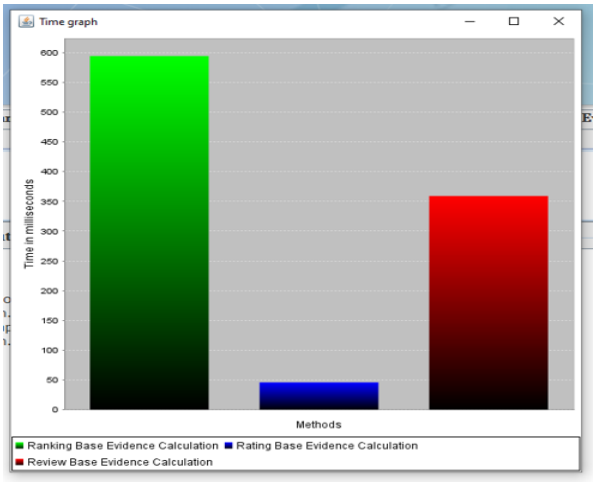

Fig 6- Ranking, Rating, Review based evidence calculation Time graph

The above graph indicates the time taken to calculate ranking, rating and review based evidence.

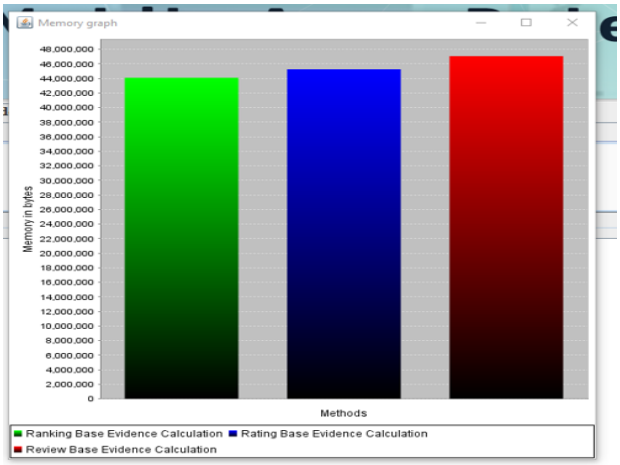

Fig 7- Ranking, Rating, Review based evidence calculation Memory graph

The above graph indicates the memory used for calculating ranking, rating and review based evidence.

Advantages: The proposed framework is scalable and can be extended with other domain generated evidences for ranking fraud detection.
Experimental results show the effectiveness of the proposed system, the scalability of the detection algorithm as well as some regularity of ranking fraud activities.
To the best of our knowledge, them is no existing benchmark to decide which leading sessions or Apps

really contain ranking fraud. Thus, we develop four intuitive baselines and invite five human evaluators to validate.

V-CONCLUSION

In this paper, we looked at the application of a rating fraud model to mobile software. App developers today are using various strategies to build their ranks right now, including Internet scam techniques like banditry. This paper presents a number of different ways to prevent fraud detection, which have been separated into various methods in order to provide clarity. there are also three types of tasks that are implemented like web ranking systems, programs for discovering web spam, application-specific ranking, and recommendations for mobile applications. Every single one of these techniques can be honestly used to tackle the problem of rating and detecting fraud. We used expansion for all the likelihood of becoming a top session to pull all together in and treat it as one collection of facts. However, there is one point of view that states that all evidences can be displayed by statistical theories, and therefore they are perfectly well-suited to expose rating fraud. Another thing we can do to clean up the dataset is to delete the fakes with the similarity algorithm, and ensure that every review is legitimate, which means that they are from people who are authorized to post on the site. Expanding on the experimental evidence found, the findings demonstrate that the expansion/synthesis/models save both time and memory.

REFERENCES

[1] *H. Zhu, H. Xiong, Y. Ge, E. Chen, Discovery of Ranking Fraud for Mobile Apps, 2015 IEEE.*

[2] *Ntoulas, M. Najork, M. Manasse, and D. Fetterly. Detecting spam web pages through content analysis. In Proceedings of the 15th international conference on World Wide Web, WWW 06, pages 8392, 2006.*

[3] *N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explore. News.13(2):5064, May 2012.*

[4] *E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviours. In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM 10, pages 939948, 2010.*

[5] *Z.Wu, J.Wu, J. Cao, and D. Tao. Hysad: a semi_supervised hybrid shilling attack detector for trustworthy product recommendation. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD 12, pages 985993, 2012*

[6] *S. Xie, G.Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD 12, pages 823831, 2012.*

[7] *Yan and G. Chen. Appjoy: personalized mobile application discovery. In Proceedings of the 9th international conference on Mobile systems, applications, and services, MobiSys 11, pages 113126, 2011*