Abstract— In today's world there are many fraud ways through which app developers try to put their app at the first position. The Fraud ranking in mobile phones lead to download of the false app which allows damaging the mobile phones and falsely getting famous by that false apps. Fraud ranking in mobile phones are very important and this paper shows the misinterpretation of the apps information and configured apps position. Also a framework is used for fraud detection in apps. The work is grouped basically into three categories. Very first is web ranking spam detection, second is the online review spam detection and third one is mobile app recommendation. The first method Web ranking spam refers to any kind of actions which bring to selected Web pages an unjustifiable favorable relevance or give much importance. The second one is Review spam which is designed to give unfair view of some objects so as to influence the consumers' perception of the objects by directly or indirectly damaging the object's reputation. The third one is mobile app recommendation which tells users to check the app usage record. This paper shows the final result weather the app is true or not by using their respective historical records which is done by the administrator by logging into the system.

Keywords— Apps in mobile phones, Fraud ranking Detection, Evidence, Aggregation of evidence, Ranking Records, Rating, Review

I. INTRODUCTION

The mobile apps are growing rapidly and everyday many apps get launched as well as many apps get closed. And in these many apps there are many fraud apps which can completely damage mobile phones. There are many apps supporting many operating system such as android and Mac. As apps are growing daily and many new apps are launched everyday so it gets difficult for viewer to select the best apps, so many App stores launches daily App leader boards, which shows the rankings of various popular Apps[1, 3]. The App leader board is one of the most important ways for detecting weather that app is true or not. A higher rank on the leader board usually shows that a large number of downloads have occurred and have million dollars in revenue. Many App developers try to find out various ways like advertising to promote their Apps so that their Apps is ranked as high as possible in such App leaderboards. Instead of moving on old marketing techniques, false App developers have started fraud apps to increase their position of apps and ultimately manipulating the rankings at the store[4].

The evidences like ranking, rating and review are being used. Ranking evidences means the ranking of the apps according to their popularity. User can see the popularity list and can decide the level of apps in play store. In rating evidences, resolving the problem of “restriction of time reduction” is done, identifying of fraud evidences is also surveyed by the app rating records [2, 5]. As it is known that rating is been done after downloading by the user, and if the rating is high in leader board then it is attracted by most of the users. The review based aggregation that contains some text as comments as reviews by the app user and before downloading or using the app user mostly refer to the reviews given by most of the users. This is usually done by downloads, ratings and reviews in short time [6].

There are works, like web positioning spam recognition, online survey spam identification and portable App suggestion, but the problem of difference in the position is misrepresentation in the mobile Apps is still being researched [7]. The problem of detecting fraud ranking apps in mobile
Apps is still not developed much. To overcome the various problems regarding the fraud app is the build of such a system for positioning misrepresentation by discovering framework for portable apps that is the model for detection of fraud ranking apps in mobile phones. For this, we have to identify several important challenges. The very first, fraud is happening at any time during the whole cycle of app, so the identification of the exact time of fraud is important. The Second is due to the large number of mobile Apps, it is very difficult to manually label ranking fraud for each App, so it is very important for automatically detection of fraud without any help of any basic information. In some leading events ranking that is fraud usually happens in leading sessions. The main target is to detect the fraud ranking of mobile Apps within leading sessions. The First is an propose of an effective algorithm to identify the leading sessions of each App based on its ranking records. With the analysis of Apps ranking behaviors, finding out the fraud Apps generally have different ranking patterns in each and every leading session compared with normal Apps. Some fraud evidences are characterize from Apps’ historical ranking records. The three functions are then developed to extract ranking based fraud evidences. Further again two types of fraud evidences are proposed based on Apps rating and review, which reflects some patterns from Apps rating and review. In addition to integrate these three types of evidences, an unsupervised evidence-aggregation method is developed which is used for evaluating the credibility of leading sessions from mobile Apps. Although some of above processes can be used for anomaly detection from rating and reviewing records, they are not able to extract fraud evidences in a limited given time period.

Overview of the paper is as follows: The literature survey is shown in section 2 which shows the various work of various scientists done on the topic of ranking fraud application. The next section is section 3 which shows the proposed system, in which the architecture system is mentioned. After this there is algorithm which is mentioned in section 4, in this minning leading session algorithm is used and also along with it aspect minning algorithm is used. Implementation is described in section 5 and have sub point of data flow. And last section 6 is conclusion.

II. LITERATURE SURVEY

Page Lay Hengshu Zhu et al. [1] introduces ranking fraud detection system for mobile Apps which accurately identify the ranking fraud by mining the active periods, like leading sessions, of mobile Apps. This paper displays three types of evidences mainly ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Hengshu Zhu also proposes an optimization based aggregation method for gathering all the evidences for fraud detection.

H. Zhu et al. [2] illustrates extracting of personal context-aware preferences device logs that is context logs for evaluating personalized context-aware recommender systems. Users download these application and installed it, besides, it is not only a indicator of whether user actually wants these application. Sometimes users only download and install the applications to check them out. So it becomes necessary to try the context logs of users to mine personal context-aware preferences of users. Yong Get al introduced the system for detection of taxi driving fraud which is misleader by fraudulent taxi drivers to earn the money. They take unwanted detours to passengers to commit the fraud by overcharging them. So this type of fraud are detected using GPS traces collected from numerous of taxi’s and from these GPS traces different evidences are collected and finally these evidences are merged using dempster-shafer theory.

Maksims et al. [3] solves Meta search and collaborative filtering problems mainly by using flexible probabilistic model over pearly comparisons, in which various preferences over objects must be aggregated into a consensus ranking.

Jeevanandam et al. [4] introduces Opinion mining using Learning Vector Quantization classifier. It is inefficient to manually take in account the large amount of opinions which created during online. To solve this Opinion mining uses automatic processes for collecting reviews. Inverse document
frequency is used to collect features from review document and Principal Component Analysis is used for feature selection. Alexandre Klementiev et al. [5] introduces unsupervised learning framework for rank aggregation over votes of rankers with domain specific expertise and also gives EM-based algorithm. This paper shows framework on collecting full rankings and aggregating top-k lists that increases over a domain-agnostic baseline. This approach is used to solve various problems in Information Retrieval and Natural Language Processing such as meta-search or collection of machine translation systems. Vyas Krishna Maheshchandra, and Prof. Ankit P. Vaishnav introduce “A survey on web spam detection” done an overview on an eminent source of gathering the analysis on particular thing where people will write their review seeing that thing. Some people will does some wrong writing and wrong comments. This condition is known as review spam. So, they used the different methods acquainted with to recognize the Review spam with their outcome, various methodologies like Vector Space, SVM, SLM, LM and I-match. In this way, web crawlers have a solid method to get rid of spam website pages. It used a Language model for an identification of effective detection of web spam which gathers latest link-based features build on a classifier. For executing this SVMs calculation is utilized and then send for scanning for a hyper plane that merges two classes of information with the biggest edge.

Michael Crawford*, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada [6] introduces “A taxi driving fraud detection” does the survey on Online spam recognition to give a solid impact on discriminating survey spam utilizing various machine learning strategies and to use approach for leading further examination. They used the noticeable machine learning methods that have been determined to take response of the issue of audit spam recognition and the execution of various methods for merging and recognition of audit spam. Also proposed “a Mobile app recommendation on security awareness” novel method recommendation by producing suggestions by both examining the metadata and by measuring the closeness between the applications, using the Latent Semantic Index strategy. They also proposed an assorted qualities based-measurement advanced structure for the improving the versatile application recommender frameworks. To execute the structure, they needs to show the framework advancement as a multi-criteria streamlining issue and plan a rank accumulation scheme to settle it.

Kun Wang, Yang Xiao, and Zhen Xiao[7] proposed the “Internet Water Army’s” conduct from numerous methods. At that point it needs to choose a few successful components as the preparation item and utilize machine learning strategies in order. In view of the conduct of clients review and remark, they work on a model to quantify the impact of Internet Water Army. With a specific goal end to lessen the impact coefficient, they introduced other direct time many-sided quality online calculations named MEIWA. The new calculation results determines that the impact is reduced to one 6th of the succession procedure which is used as a matter of course with giving the guaranteeing of clients' survey remarks propensities.

III. PROPOSED SYSTEM
Detection of ranking fraud for mobile Apps is still undetermined subject to research. To fill this crucial lack, we thought to develop a ranking fraud detection system for mobile Apps. We also evaluate several important challenges. First challenge, in the full life cycle of an App, the ranking fraud does not happen always, so we require to detect the time when fraud happens. This challenge can be considered as identifying the local anomaly in the order of global anomaly of mobile Apps. Second challenge, it is important to have a scalable way to correctly detect ranking fraud without using any basic information, as there are vast number of mobile Apps, it is very difficult to manually label fraud ranking for each App. Finally, due to the dynamic nature of chart rankings, it is difficult
to find and identify the evidences related with ranking fraud, which motivates us to find some implicit fraud patterns of mobile Apps as evidences.

The system architecture shows various components. The very first thing is administrator and developer login. In this the admin and developer will login. Developer is the person who develops apps while admin is the person who look after the system. User is the final client which will use this system to find out whether the app is true or not. So for this three parameters is used rating, ranking and review. These three parameters are used to detect the fraudness of the app. Ranking is the numeric number given to the app, rating is the number of stars given and review is the textual comment given. So developer will develop the app and upload it in the system. Then many users will login and rank them, rate them and review them. Then the admin will login and will rank them, rate them and review them. After this the aggregation of all the ranking, rating and reviewing is done by aspect ratio algorithm. Mining leading is used for aggregation of ranking and rating while aspect ratio is used for all three. If the aggregation value comes less than the respective values of ranking, rating and reviewing then it is a fraud app or else it is a true app.

IV. ALGORITHM

ALGORITHM 1: FRAUD RANKING BEHAVIOR DETECTION

INPUT: Mobile app rating and review evidence
Output: Fraudulent ranking behavior of apps
1. Gather user rating and review of apps
2. Mine the leading session and leading event of app
3. Leading session of app equal to the leading event of app
4. Find ranking based evidence
5. Find rating based evidence
6. Find review based evidence
7. End
8. Aggregate the evidence based on unsupervised approach
9. Output fraudulent app behavior
10. Load the user rating and review comments
11. Divide rating and review evidence
12. Calculate fraud in app and store
13. Rank app on original rating
14. Return fraud app

Above shows the step by step description of behavior of detecting fraud in mobile app. This method increase the accuracy to detect the fraud. Mathematical model-
a) Theory set
*Input-input set I containing input text file
I=records of ranking rating and review
*Process set-various processes set used
P= (P1, P2, P3, P4)
P1-review evidences
P2-rating evidences
P3-ranking evidences
P4-aggregation of evidences
*Intermediate output set-output sets are determined by two sets.
  #Very first set is-
  ∴IO-(IO1, IO2, IO3, IO4)
  IO1-review evidences generated
  IO2-rating evidences determined
  IO3-ranking evidences determined
  IO4-aggregation done
*Final output set output -
I=fraud app to be detected

b) Venn diagram
Displaying of the mapping of the input, output and processes of the system.
It is also used for showing the interaction of all the three parameter

![Venn diagram with sets I, P, IO, O, Q1, Q2, Q3, Q4, Q5 showing the interaction of processes and output.]

c) Process state diagram
This shows the interaction of the various processes.
Processes are determined by Q1, Q2, Q3, Q4, Q5.
In this Q5 is the final state in which finally the fraud app has been detected.
(Process state diagram)
5.1. Identifying Leading Sessions
Ranking fraud usually happens in leading sessions. Hence, detecting ranking fraud of mobile Apps is actually mean to detect ranking fraud within leading sessions of mobile Apps. Specifically, we first propose a simple effective algorithm to determine the leading sessions of each App based on its past i.e. historical ranking records. Then, with the analysis of Apps’ ranking ‘behaviours, we find that the fake Apps often have different ranking patterns in each leading session as compared with normal Apps.

-Mining Leading Sessions
There are two main methods for mining leading sessions. First one, we need to discover leading events from the App’s historical, ranking records. Second, we need to gather adjacent leading events for constructing leading sessions.

b) Ranking Based Evidences
A leading session is composed of many leading events. Hence, we have to first analyze the basic characteristics of leading events for extracting fraud evidences. By analyzing the Apps’ past ranking records, we observe that Apps’ ranking behaviors in a leading event always satisfy a specific and average ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase. Specifically, in each leading event, an App’s ranking first increases to a peak position in the leader board (i.e., rising phase), then keeps such peak position for a period (i.e., maintaining phase), and finally decreases till the end of the event (i.e., recession phase).
These formulas help to find the ranking of the apps.

c) Rating Based Evidences
The ranking based evidences are useful for ranking fraud detection. However, while, it is not sufficient to only use ranking based evidences. Specifically, after an App has been published in the play store, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important thing of App advertisement. An App which has higher rating may attract more users to download that particular app and can also be ranked higher in the leader board. Thus, rating manipulation is also an important perspective of ranking fraud. Intuitively, if an App has ranking fraud in a leading session $s$, the ratings during the time period of $s$ may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences.

\[
\theta_1 = \arctan \left( \frac{K^* - r^a_b}{t^*_b - t^a_c} \right), \quad \theta_2 = \arctan \left( \frac{K^* - r^a_c}{t^*_d - t^a_c} \right)
\]

(Ref: Paper-1)

\[
\bar{\theta}_s = \frac{1}{|E_s|} \sum_{e \in E_s} (\theta^e_1 + \theta^e_2),
\]

(Ref: Paper-1)

These formulas help to find the ranking of the apps.

d) Review Based Evidences
Besides ratings, most of the App stores also allow users to write some textual comments or reviews about that App. Such reviews will shows the personal perceptions and usage experiences of users for particular mobile Apps. Indeed, review manipulation is one of the most important features of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often firstly few, read its historical and past reviews to detect their decision making, and a mobile App contains more positive reviews may attract more users to download and installed it. Therefore, imposter’s often post false reviews in the leading sessions of a specific App in order to increase the App downloads, and thus increases the App’s ranking position in the leader board. Although some past works on review spam detection have been reported in recent years, the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under-explored.

\[
\Delta R_s = \frac{\overline{R}_s - \overline{R}_a}{\overline{R}_a}, \quad (s \in a).
\]

(Ref: Paper-1)
e) Evidence aggregation
In this the information gathered from all the parameters is aggregated and final result is being shown. Final score is in terms of linear equation.

\[ \Psi(s) = \sum_{i=1}^{N_p} w_i \times \Psi_i(s), \quad s.t. \sum_{i=1}^{N_p} w_i = 1, \]

(Ref. Paper-1)

5.2. EXPERIMENTAL SETUP
We will take an example from paper (1) which shows the apps table for paid apps and free apps available at the apple store. The app in which these two apps are stated is known as leader board. In this leader board the top paid apps and free apps are shown.

| Statistics of the Experimental Data |
|-----------------------------------|
| Top Free 300 | Top Paid 300 |
| App Num.     | 9,764        | 5,281        |
| Ranking Num. | 285,900      | 285,900      |
| Avg. Ranking Num. | 29.22 | 54.34 |
| Rating Num.  | 14,912,459   | 4,561,943    |
| Avg. Rating Num. | 1,524.17 | 867.12 |

Then, for each evidence score \(\Psi_i(s)\), we can measure its consistency using the variance-like measure,

\[ \sigma_i^2(s) = \left( \pi_i(s) - \pi(s) \right)^2. \]

![Fig. 6. The distribution of the number of Apps w.r.t. different rankings.](image)

5.3 DATA FLOW DIAGRAM
The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system. Also various parameters to take the input is also taken into serious consideration. All types of operations to be taken into performing any process to be completed is taken into consideration and
all are shown into steps wise manner. By this chart, user can easily understand the problem and find out about each and every steps in less time and can also find out the solution if something wrong or error occurs.

5.4.Screenshots

V1.CONCLUSION

This paper, gives the ranking fraud detection method for mobile apps. Now days many of mobile app developers uses numerous frauds techniques to increase their rank. To avoid this chart ranking, there are many fraud detection techniques one of which is by extracting the leading sessions of versatile Apps, we check the ranking fraud. The leading sessions works for checking the nearby inconsistency of App rankings. The framework expects to distinguish the ranking frauds taking into consideration three types of method, i.e., ranking, rating and review based method. Further, an optimization based aggregation strategy aggregates all the three proofs to distinguish the fraud. This paper proposes the time efficient system to detect the fraud Apps.
REFERENCES

1. Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen, “Discovery of Ranking Fraud for Mobile Apps”, IEEE TRANSACTION ON KNOWLEDGE AND DATA ENGINEERING, Vol. 27, NO. 1, pp. 74-87, JANUARY 2016
2. Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, “A taxi driving fraud detection system,” in Proc. IEEE 11th Int. Conf. Data Mining, pp. 181–190, 2011
3. Hengshu Zhu, Hui Xiong, Yong Ge and Enhong Chen —Mobile App Recommendations with Security and Privacy Awarenessl KDD’14, August 24–27, 2014, New York, NY, USA. Copyright 2014 ACM 978-1-4503-2956-9/14/08 ...$15.00
4. Kun Wang, Yang Xiao, and Zhen Xiao —Detection of Internet Water Army in Social Network! International Conference on Computer, Communications and Information Technology (CCIT 2014)
5. M. N. Volkovs and R. S. Zemel, “A flexible generative model for preference aggregation,” in Proc. 21st Int. Conf. World Wide Web, pp. 479–488, 2012.
6. Jeevanandam, Jotheeswaran Loganathan R. and Madhu Sudhanan B. “Feature Reduction using Principal Component Analysis for Opinion Mining”, International Journal of Computer Science and Telecommunications, Volume 3, Issue 5, May 2012.
7. Klementiev, D. Roth, K. Small, and I. Titov, “Unsupervised rank aggregation with domain-specific expertise,” in Proc. 21st Int. Joint Conf. Artif. Intell. Pp. 1101–1106, 2009.