

# Manual Fakeapp Detection Using Sentimental Analysis through Webpage

T. Sathish<sup>1</sup>, T.S.Abinaya<sup>2</sup>, B.Anupriya<sup>3</sup>, L.Uma<sup>4</sup>

<sup>1</sup> M.Tech. Assistant Professor, Department of CSE, Selvam College of Technology, Namakkal, India

<sup>2, 3, 4</sup> UG Student, Department of CSE, Selvam College of Technology, Namakkal, India.

**Abstract:** Now a day's fake app is increasing in Google play store and it is launched with the popular names which cause major challenges in current scenario. Some fraudulent developers deceptively boost the search rank and popularity of their app. We came with an idea where a webpage with PHP along with the sentimental analysis is used to process the rank and comment given by the user where the comment percentage and rate percentage accuracy automatically maintained and a graph which shows the app range is created. This analyzing method has to be done manually due to insufficiency of Google play store source. In future we will manipulate automatic analyzer to reveal more accuracy for this; process should have adequate sanction from Google to append webpage in their Google play store itself. Once we will get adequate sanction from Google, we append the webpage and the sentimental analyzer which work directly in Google page itself.

**Keywords:** Mobile phones, ranking, rating, review and sentimental analysis.

## I. INTRODUCTION

In this paper, a detailed analysis of various works related to sentimental analysis is taken. Sentimental analysis basically involves classification, but the content that actually contains the sentiment must be identified. The emotional analysis is sometimes merged with sentimental analysis in which case multiple tuples come into use. The mobile applications are increasing in day by day. Different types of apps can be used in mobile phones. These kinds of apps can be downloaded freely in Google play store and the apps have no cost. In Google play store all apps are not original. They introduce some fake apps the user cannot identify the fake app. Ranking fraud in the mobile app market refers to fraudulent or deceptive activities which have a purpose of bumping up the apps in the popularity list. App downloads appraisals and surveys in a brief timeframe. For instance, an article from Venture Beat reported that, when an App was advanced with the assistance of positioning control, it could be moved from number 1,800 to the main 25 in Apple's sans top pioneer board and more than 50,000-100,000 new clients could be gained inside of a few days. Truth be told, such positioning extortion raises awesome worries to the versatile App industry. 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.

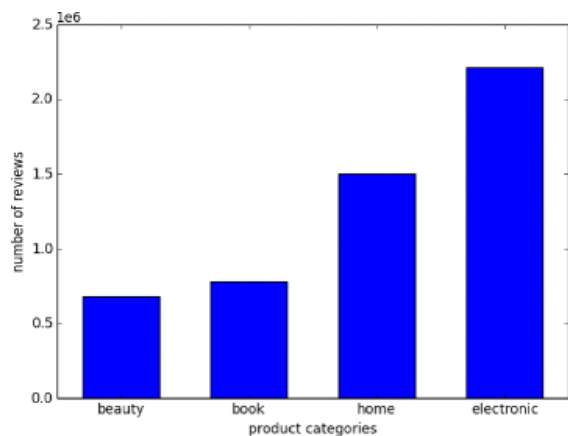


Moreover, we propose a streamlining based accumulation technique to incorporate every one of the proofs for extortion discovery. At last, we assess the proposed framework with true App information gathered from the iOS App Store for quite a while period. In the trials, we approve the viability of the proposed framework, and demonstrate the versatility of the discovery calculation and also some normality of positioning extortion exercises. To this end, in this paper, we give an all-encompassing perspective of positioning extortion and propose a positioning misrepresentation recognition framework for portable Apps. There are a lot of social websites that provide whole bulk of data that is informative from various view points. Twitter is one of the main sources of such data.

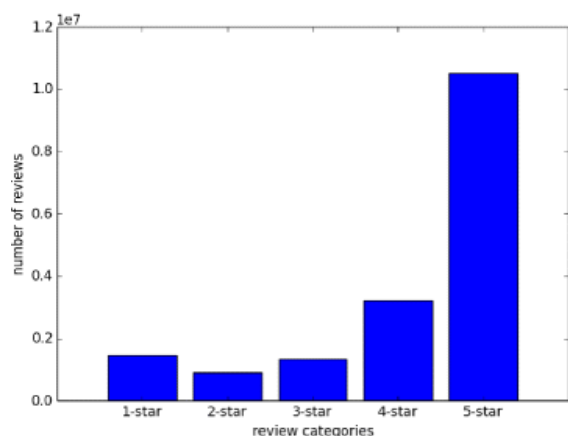
Twitter sentiments help in getting information about various current trends and can be used further in deciding usefulness of some tasks, products and themes. This paper aims to present the implementation of sentiment analysis methods in the area of scientific paper reviews as a proof of concept for future applications. A significant research work has been done since 1980 on different aspects of twitter sentiment analysis. Now we will see major work in this field and their contributions. Human life is filled with emotions and opinions. Having an access to large quantities of data through internet and its transformation into a social web is no longer an issue, as there are terabytes of new information produced on the web. As our project based on evidences collected from app data; the one of the mostly judgment by people is rating based evidences which can be used to rate the app while downloading it or we can rate it after seeing its performance.

Every day that is available to any individual. Even more importantly, it has changed the way we share information section 'Review-level categorization'. Section 'Conclusion'

concludes the paper. The use of social media is increasing day by day. Increasing growth of social media users over internet has also increased their participation in various discussions and activities simultaneously. In case of a product, reviews of users will help to take many important decisions about the services of the product.



(a)



(b)

This paper tackles a fundamental problem of sentiment analysis, namely sentiment polarity categorization. A flowchart that depicts our proposed process for categorization as well as the outline of this paper. Experimental results are presented in section 'Results and discussion'. Hence such applications must be identified, so that they will be identifiable for play store users.

The fake app can be identified based on rating, ranking and review. To increase the download of the app, developers manipulate the chart ranking, using human water armies and bot-farms. Apple got serious about various questionable promoting firms that utilize programming bots and multitudes of human clients to download applications as a group, pushing the titles to prominent situations inside of the App Store's "Top Free" rankings outline. The ranking of apps on leader board is done on the basis of views of user. Higher rank in the leader board causes million dollars in income. The

rest of the paper is organized as follows: Section II, presents the literature survey over the related work.

Getting higher position on leader board by using unwanted ways affects the reputation of app when fraud is caught, but when app is developed by known or famous app developer then that app automatically becomes famous due to this other app gets affected. Ranking evidences means the ranking of the apps according to their popularity. Rating is done after downloading by the user, and if the rating. In particular, we first propose to precisely find the mining so as to position extortion the dynamic periods, in particular driving sessions, of portable Apps. Such driving sessions can be utilized for recognizing the neighbourhood abnormality rather than worldwide irregularity of App rankings. Moreover, we research three sorts of proofs, i.e., positioning based confirmations, modelling so as to rate based proofs and audit based proofs, Apps' positioning, rating and survey practices through factual speculations tests. Moreover, we propose a streamlining based accumulation technique to incorporate every one of the proofs for extortion discovery.

At last, we assess the proposed framework with true App information gathered from the iOS App Store for quite a while period. In the trials, we approve the viability of the proposed framework, and demonstrate the versatility of the discovery calculation and also some normality of positioning extortion exercises. The object of sentiment analysis has typically been a product or a service whose review has been made public on the Internet. This might explain why sentiment analysis and opinion mining are often used as synonyms, although, we think it is more accurate to view sentiments as emotionally loaded opinions.

In general, opinion mining helps to collect information about the positive and negative aspects of a particular topic. Finally, the positive and highly scored opinions obtained about a particular product.

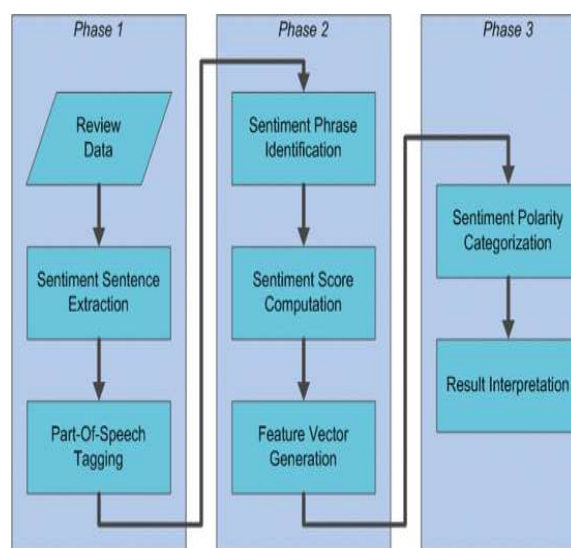


Fig: Block Diagram

The rest of this paper is organized as follows: In section 'Background and literature review', we provide a brief review towards some related work on sentiment analysis. Software package and classification models used in this study are presented in section 'Methods'. Our detailed approaches for sentiment analysis are proposed in section 'Background and literature review'. Experimental results are presented in section 'Results and discussion'. Discussion and future work is presented in section 'Review-level categorization'.

In order to promote marketing, large companies and business people are making use of opinion mining. There are several challenges in Sentiment analysis. The first is a opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. There are used in comment, rate and review.

Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great". In order to have their Apps ranked as high as possible, app developers promote their apps using various ways such as advertising, offers.

## II. LITERATURE SURVEY

**Jyoti Nandi math** has studied that the consumers or viewers before buying a product goes through a review, through the increase of e-commerce has increased the number of counterfeiting comments. Thus the need of today is to eliminate inappropriate and fraudulent reviews from a site or an application. The sentimental analysis could be done on the texts as well as on the sentences. The user of an iphone having the reviews represent the positive, negative, objects, feelings and holder of the review. The negation words like no, not, and more are to identify their phrases.

**Veenna Dubey** the target is eliminating irrelevant content through different techniques. A system is evolved by using customer reviews is known as opinion mining.

**Bharat R.Naiknaware** knew that social media monitoring has been increased in the past. In addition to that e-commerce has also played its role. With huge indulgence of people in social media the research paper mines the reviews, comments, status of twitter. Document level analysis, which analyze the document reviews wholly. Sentence analysis, a single sentence is focussed upon and Feature analysis is better than other to as it does not look for a sentence at document moreover an entity is consider ,thus making this type of level fine –gained. Here the data is collected and results from each modal are processed against all and then combine, organizes, and stores these results in SQL database.

**Md.Daiyan** explained that the sentimental analysis involves a scheme to collect and classify opinions about a product in order to track the humour of the people for a certain product. But positive document of any product does not signifies that

the reviewer hold positive opinion on all the aspects of the products.

**M.S.Patil** explained review spamming that is illegal activities like fake reviews writng. It may mislead readers by producing false positive or false negative opinion to entities to promote them or to run their reputation. False opinion: reviews that contain false opinion on products. Non-Reviews that have no opinions and are not affected by the reader of reviews, affect the automated mining systems only. It is required to find features in reviews by using sentimental analysis to identify the spam.

**Sushan Kokate** discussed the difficulty in finding review spam or to recognize fake reviews. Authors also found many unexpected rules to identify unusual review patterns ,to analyze a review data set and to indicate spam activities .In the result section are discussed the results obtained the system for detecting fake and truthful reviews given by the users. Diagrams or charts are shows the number of reviews of user.

**Qingxi peng** discussed that the sentiment analysis techniques to detect review spam. Authors proposed the method to calculate score for sentiments from natural language text and analyse the relationship between spam reviews and sentiment score. various rules are generated using the relationship between spam reviews and sentiment score. Detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps.

## III. RELATED WORK

Sentiment Analysis User-generated- content in the Web such as review, blog and micro blog usually express the author's emotion. The extraction, analysis and summarization of sentiment become an important research field. A summarization system has been built to show the user's sentiment to every feature of the product. sentiment analysis is classified at word and sentence level. comparative sentence was identified, and comparative relationships were extracted from the identified comparative sentence.

A series of definition about sentiment, sentiment holder and sentiment polarity are presented in Senti Word Net lexical resource has been applied in automatic sentiment classification. In order to study the sentiment analysis, it is important to extract product features. Two approaches have been proposed to solve this problem. One method is based on dependency parser. According to previous studies, product features are almost noun phrase. Therefore, product features can be extracted by phrase dependency parsing. Many mobile app stores launched daily app leader boards which show the chart ranking of popular apps. The leader board is the important for promoting apps. Original application grade level decreases due to the duplication arrival in the mobile apps. In recent activities duplicate version of an application not burned or blocked. This is the major defect. Higher rank leads huge number of downloads and the app developer will get more profit. In this they Allow Fake Application also. User not

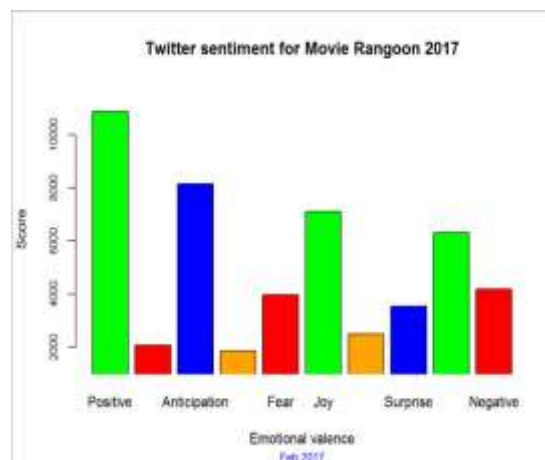
understanding the Fake Apps then the user also gives the reviews in the fake application. Exact Review or Ratings or Ranking Percentage are not correctly Calculated. In general, existing studies mainly focus on the analysis, summarization and visualization of the sentiment. Several behavioural models derived from the collusion phenomenon among fake reviewers and relation models are been used to detect the spam groups. The previous studies mentioned above generally detect the spam review by means of rating score. Sentiment analysis has never been used in spam review detection. In the unusual review patterns which can represent suspicious behaviours are identified, and unexpected rules are formulated. A novel concept of review graph is proposed in which capture the relationships among all reviewers, reviews and stores that the reviewers have reviewed as a heterogeneous graph. Then an iterative computation model is proposed to identify suspicious reviewers.

In which the user can securely obtain their private keys from the group manager without any certificate authority due to the verification for the public key of user. This scheme can achieve fine grained access control. This scheme uses the polynomial function for user revocation so it protect form collusion attack. This methods do not need training and also cost effective. A real data set is used to evaluate the effectiveness and the efficiency [1]. Therefore it is illustrated from those ranking records that some fraud is taking place in mobile app market and to restrict those frauds, evidences are developed to detect such fraud. As only ranking based evidences does not seems to be much sufficient to detect the fraud of mobile app, based on apps rating and review history some fraud evidences were discovered which showed anomaly patterns by those history. Specifically, an unsupervised evidence aggregation method is also proposed for evaluating the trustworthiness of leading sessions. And finally, the proposed system is estimated with real world app data gathered from the Google Play store for time consuming period. With the increase in the number of web Apps, to detect the fraudulent Apps, we have propose a simple and effective algorithm which identifies the leading sessions of each App based on its historical ranking of records.

#### IV. PROPOSED SYSTEM

The existing technique integrate the results of various evidences to find genuineness of an app. We again integrate a recommender system and the ranking score generated previously such that, the recommender system will recommend the most genuine app that are most relevant. The apps suggested by recommender system will be checked for whose score is high, and the app with highest score will be suggested. With the increase in the number of web Apps, to detect the fraudulent Apps, we have propose a sentimental analysis which identifies the leading sessions of each App based on its historical ranking of records. Audit based confirmation is utilized to check the surveys of the application.

The KNN calculation is utilized to enhance effectiveness and precision of the these all develop three functions to extract such ranking based fraud evidences. Ranking fraud usually happens in leading sessions. Therefore, detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps. Specifically, we first propose a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking 'behaviours, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps.



The rating based confirmation is utilized to rate by any client who downloaded it.

Audit based confirmation is utilized to check the surveys of the application.

To start with the mining driving sessions is utilized to find driving occasions from the application's chronicled positioning records and after that it blends nearby driving occasions for building driving sessions.

#### *Ranking by evidence*

A leading session is composed of several leading events. Therefore, we should analyze the basic characteristics of leading events for extracting fraud evidence. By analyzing the Apps 'historical ranking records, we observe that Apps 'ranking behaviours in a leading event always satisfy a specific ranking pattern. The leading session contain much more leading events compared with other leading session of apps in the leader board, it has high probability of having ranking frauds.

#### *Rating Based Evidence*

The ranking based evidences are useful for ranking fraud detection. However, sometimes it is not sufficient to only use ranking based evidences. For example, some Apps created by the famous developers, such as Game loft, may have some leading events with large values of 01 due to the developers' credibility and the "word-of mouth" advertising effect. Moreover, some of the legal marketing services, such as "limited time discount", may also result in significant ranking



based evidences. To solve this issue, we also study how to extract fraud evidences from Apps' n App has been published, it can be rated by any user who downloaded it. For a normal App, the average rating in a specific leading session should be consistent with the average value of all historical ratings. 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. In contrast, an App with rating manipulation might have surprisingly high ratings in the fraudulent leading sessions with respect to its historical ratings.

**Fake AppDetection**

With those limitations in mind, our future work is to focus on solving those issues. Detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Apps. Specifically, we first propose a simple yet effective algorithm to identify the leading sessions of each App based on its historical ranking records. Then, with the analysis of Apps' ranking behaviors, we find that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we characterize some fraud evidences from Apps' historical ranking records, and develop three functions to extract such ranking based fraud evidences. More future work includes testing our categorization scheme using other datasets. By analyzing the ranking behaviours of Apps, we discover that the fraudulent Apps often have different ranking patterns in each leading session compared with normal Apps. Thus, we identify some fraud evidences from Apps' historical ranking records and develop three functions to obtain such ranking based fraud evidences. Proposed a scheme that provides a secure way for key distribution without secure communication channels. In which the user can securely obtain their private keys from the group manager without any certificate authority due to the verification for the public key of user.

*Review Based Evidences:*

Besides ratings, most of the App stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Apps. Along with rating users are allowed to write their reviews about the app. Such reviews are showing the personalized experiences of usage for particular mobile Apps. The review given by the user is one of the most important factors for the popularity of the app. As the reviews are given in natural language so pre-processing of reviews and then sentiment analysis on pre-processed reviews is performed. The system will find sentiment of the review which can be positive or negative. Positive review adds plus one to positive score, if negative it will add one to negative score. In this way it will find outscore of each of the reviews and determine whether app is fraud or not on the basis of review based evidences. Indeed, review manipulation is one of the most important perspectives of App ranking fraud. Specifically, before downloading or purchasing a new mobile App, users often firstly 5, read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposters often post fake reviews in the leading sessions of a specific App in order to inflate the App downloads, and thus propel the Apps ranking position in the leader board. Although some previous 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. 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

#### *Rating Score Calculation*

Ratings on App store are generally between one to five, in this module we compute the average rating of particular app and set a threshold and compare with it. The rating which are less than or equal to three are considered as negative ratings

### V. FUTURE WORK

In future we will manipulate automatic analyzer to reveal more accuracy for this, process should have adequate sanction from Google to append webpage in their Google play store itself. Once we will get adequate sanction from Google, we append the webpage and the sentimental analyzer which work directly in Google page itself. At that point, it is recognized that positioning based confirmations, rating based proofs and survey based confirmations are used for identifying positioning extortion.

In addition, a unique model is proposed which is an improvement based total system to incorporate every one of the proofs for assessing the validity of driving sessions from

portable Apps. First, ranking fraud does not always happen in the whole life cycle of an App, so we need to detect the time when fraud happens.

Such challenge can be regarded as detecting the local anomaly instead of global anomaly of mobile Apps. Second, due to the huge number of mobile Apps, it is difficult to manually label ranking fraud for each App, so it is important to have a scalable way to automatically detect ranking fraud without using any benchmark information. Finally, due to the dynamic nature of chart rankings, it is not easy to identify and confirm the evidences and rating above three are considered as positive ratings. In future we will manipulate automatic analyzer to reveal more accuracy for this, process should have adequate sanction from Google to append webpage in their Google play store itself.

Linked to ranking fraud, which motivates us to discover some implicit fraud patterns of mobile Apps as evidences. fraud evidences from Apps' historical rating records. Specifically, after an App has been published, it can be rated by any user who downloaded it.

A novel point of view of this methodology is that every one of the proofs can be displayed by measurable theory test, in this way it is anything but difficult to be reached out with different confirmations from space information to distinguish positioning misrepresentation.

At last, the proposed framework is accepted with broad examinations on certifiable App information gathered from the Apple's App store. In Future a study can also be done to make a detection system that tracks the online campaigning on social media. Online campaigning may be used to gain a particular benefit that maybe of business, politics or something else. Exploratory results demonstrated the adequacy of the proposed methodology.

Later on, to concentrate more viable misrepresentation confirms and dissect the idle relationship among rating, survey and rankings is planned. In addition, amplification of positioning misrepresentation location approach is performed with other portable App related administrations, for example, mobile Apps suggestion, for improving client experience. We have given a framework using which we can detect fake profiles in any online social network with a very high efficiency as high as around 95%. Fake profile detection can be improved by applying NLP techniques to process the posts and the profile. Once we will get adequate sanction from Google, we append the webpage and the sentimental analyzer which work directly in Google page itself.

## VI.CONCLUSION

This paper introduces more effective fraud evidences and analyze the latent relationship among rating, review and rankings. We extended our ranking fraud detection approach with other mobile app related services, such as mobile app recommendation for enhancing user experience. By extracting

the leading sessions of versatile Apps, we intend to find the ranking fraud. The leading sessions works for identifying the nearby inconsistency of App rankings. The framework expects to distinguish the ranking frauds taking into account three sorts of confirmations, for example, ranking, rating and review based proofs. Further, an optimization based aggregation strategy joins all the three proofs to distinguish the fraud. This paper introduces a system which is built up and it is actually a positioning extortion discovery framework for mobile Apps. In particular, initially it is demonstrated that positioning misrepresentation happened in driving sessions and gave a system to digging driving sessions for each App from its chronicled positioning records. As increase popularity, mobiles are major target for malicious applications. Main challenge is to detect and remove malicious apps from mobile app market. Main aim of the fraudulent is to knock the fraud mobile apps in the popularity list. Thus, there is need to have novel system to effectively analyze fraud apps. In proposed system, system performance can be enhanced by adding the recommendation based on the ranking.

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed. This article has studied the application of sentiment analysis techniques in the domain of paper reviews

## REFERENCES

- [1]. Dubey Veena, G. D. (2016). Sentiment Analysis Based on Opinion Classification Techniques: A Survey . International Journal of Advanced Research in Computer Science and Software Engineering , 53-58.
- [2]. Nandimath Jyoti, K. B. (2017). Efficiently Detecting and Analyzing Spam Reviews Using Live Data Feed. International Research Journal of Engineering and Technology (IRJET) , 1421-1424.
- [3]. Kokate Sushant, T. B. (2015). Fake Review and Brand Spam Detection using J48 Classifier. International Journal of Computer Science and Information Technologies, 3523-3526.
- [4]. Prajakta Gayke and prof. Sanjay Thakre —Detection of Ranking Fssraud for Mobile Appl IOSR Journal of Computer Engineering (IOSR-JCE)2015
- [5]. Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321- 7308 Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 9, September 2015
- [6]. A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh. Spotting opinion spammers using behavioral footprints. In Proceedings of the 19th ACM SIGKDD International conference on Knowledge discovery and data mining, KDD '13, 2013.
- [7]. Bin Liu, Deguang Kong, Lei Cen, Neil Zhenqiang Gong, Hongxia Jin and Hui Xiong—Personalized Mobile App Recommendation: Reconciling App Functionality and User Privacy Preference February 2–6, 2015, Shanghai, China. Copyright 2015 ACM 978-1-4503-3317-7/15/02 ...\$15.00.

- [8]. Donghwan Bae, Keejun Han ; Park, J. ; Yi, M.Y. —, AppTrends: A graph-based mobile app recommendation system using usage history/ Big Data and Smart Computing (BigComp), 2015 International Conference.
- [9]. Wang Xiang, Zhang Zhilin ; Yu Xiang ; Jia Yan ; Zhou Bin ; Li Shasha —Finding the hidden hands: a case study of detecting organized posters and promoters in SINA weibo/ Communications, China (Volume:12 , Issue: 11 ) November 2015.
- [10]. Ranking fraud Mining personal context-aware preferences for mobile users. H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. In Data Mining (ICDM), 2012 IEEE 12th International Conference on, pages1212–1217, 2012.
- [11]. Detection for mobile apps H. Zhu, H. Xiong, Y. Ge, and E. Chen. A holistic view. In Proceedings of the 22nd ACMinternational conference on Information and knowledge management,CIKM '13, 2013.
- [12]. Detecting product review spammers using rating behaviors. E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw In Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10, pages 939–948, 2013.