# A SURVEY ON IDENTIFICATION OF RANKING FRAUD FOR MOBILE APPLICATIONS

**Nandini B[1], A.Ananda Shankar[2]**

[1]*MTech Scholar Department of Computer Science & Engineering, Reva ITM, Kattigenahalli, Yelahanka, Bangalore-560064*

[2]*Associate Professor Department of Computer Science & Engineering, Reva University, Kattigenahalli, Yelahanka Bangalore-560064*

## Abstract

*Now a day, mobile App is an exceptionally prevalent and surely understood idea because of the quick progression in the portable innovation and cell phones. Because of the extensive number of versatile Apps, ranking fraud is the key test before the versatile App market. In this paper we are proposing a ranking fraud discovery framework for portable Apps. The proposed framework mines the leading sessions, for example, leading sessions of portable applications to precisely find the ranking fraud. Other than this, by displaying Apps ranking, rating and review practices utilizing measurable theories tests, we examine three sorts of confirmations, they are ranking based proofs, rating based proofs and review based confirmations. Proposed an aggregation method to combine all the proof for fraud detection.*

***Keywords:*** *Ranking Fraud, Fraud For Mobile, Identification Mobile Fraud*

----------***----------

## 1. INTRODUCTION

The mobile applications are growing in a faster rate now a day, for instance, at the end of 2013 April, there are more than 1.6 million Apps at Apple's App store and Google Play. For simulating the growth of the app, many app stores released leaderboards. This became the most important way to promote mobile apps. If ranking is high in the leaderboard, number of downloads and revenue increases. To increase the download of the app, developers manipulate the chart ranking, using human water armies and bot-farms. For example, for a little engineer who's just discharged an iOS application, a position close to the highest point of Apple's App Store rankings could mean the contrast between a million dollars in income and Top Ramen for supper. So it's nothing unexpected that a few coders attempt to cheat the framework — and that Apple tries to stop it. Apple got serious about various questionable promoting firms that utilization programming bots and multitudes of human clients to download applications as a group, pushing the titles to prominent situating inside of the App Store's "Top Free" rankings outline. It isn't clear what number of applications has been influenced by the bans, yet in any given month, Apple brings down 5,000 applications for an assortment of reasons, as per application seek firm Xyologic. The designers' worries came during a period when general App Store downloads were strikingly diminishing. Matthäus Krzykowski, the CEO of Xyologic, said that the volume of downloads on the U.S. Application Store has fallen 25 percent since January. The diminishing, he says, can to a limited extent be ascribed to Apple's crackdown on bots and its choice to battle "incentivized" introduce. Recently, Apple made expansive move against outsider promoting administrations firms since proof demonstrated a few advertisers were controlling the App Store's top rankings utilizing robotized PC frameworks — that is, software bots — and additionally multitudes of human clients. The advertisers utilized these instruments to download their customers' iOS applications altogether, falsely blowing up Apple's store ranking. The ranking fraud will not happens always, so need to find the exact time when it happens. This respects to detect the local anomaly. Detecting of fraud manually is difficult, so have to use the scalable method to automatically detect the ranking fraud. it is difficult to recognize the evidence and pledge the evidences, which leads to discover implicit fraud patterns. The ranking of the mobile apps are not always high, it will be only in particular leading events, which leads several leading sessions. The ranking fraud takes place in these sessions. Using “Mining Leading Sessions” will distinguish the leading events and leading session by scanning the historical ranking records only once. The ranking of the fraud apps will differ in each leading session comparing to the normal apps. The ranking pattern differs in three phases: rising, maintaining and recession phase. The app which is fraud will have a rise to peak in short time, but it will not stay for long in maintaining phase. Using Guassian approximation and classic maximum-likelihood estimation (MLE) will find the ranking fraud. To find the fraud not only ranking is important but also rating and review is needed. After getting the information, will use the unsupervised evidence aggregation to integrate the evidences.

## 2. LITERATURE SURVEY

While there are some associated work regarding ranking fraud, such as:

### 2.1 Web Ranking Spam Detection

Vyas Krishna Maheshchandra, and Prof. Ankit P. Vaishnav done an overview on an eminent source of assembling the analysis on particular item where people will write their reviews based on the item. Some people will mislead by writing the wrong comments. This leads to the review spam. So, they used the diverse methods acquainted with recognize the Review spam with their outcome, methods such as Vector Space, SVM, SLM, LM and I-match. Some of the time individuals might run over the off-base conclusions called as survey spam [1].

Ms. Meenal M. Shingare, and Prof. S. R. Chaudhary worked on the recognition of the spam in web which makes trick or deceive to web search tool. Clients have a harder time finding the data they need, and internet searchers need to adapt to a swelled corpus, which thus causes their expense per question to increment. In this manner, web crawlers have a solid impetus to get rid of spam website pages from their file. They used a Language model for an effective detection of web spam which merges latest link-based features build on a classifier. For executing this SVMs calculation is utilized which go for scanning for a hyper plane that isolates two classes of information with the biggest edge [16].

Shrijina Sreenivasan and B.Lakshmipathi proposed the work which identifies with the correlation of web spam location utilizing three unsupervised learning strategies, SOM, HMM and ART as opposed to the directed methods SVM. The regulated procedures experiences the disadvantage that it functions admirably just with vast datasets and is not expected to be a genuine –time application. After a near study, the proposed strategy is found to yield a higher execution than the current supervised learning techniques [21].

### 2.2 Online Review Spam Detection

Michael Crawford*, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada did the survey on Online spam recognition to give a solid and thorough near investigation of momentum exploration on distinguishing survey spam utilizing different machine learning strategies and to devise approach for leading further examination. They used the noticeable machine learning methods that have been proposed to take care of the issue of audit spam recognition and the execution of various methodologies for grouping and recognition of audit spam [2].

Sushant Kokate, and Bharat Tidke worked on finding the online trumped up feedback for the movies. For doing this they used a classifier called J48, it will produce ARFF from the unmistakable components to identifying the untruthful audits. As a feature of future work, they fuse survey spammer discovery into the survey location and the other way around. Investigating approaches to learn conduct designs identified with that spamming in order to enhance the precision of the present relapse model is moreover [3].

Amir Karam, and Bin Zhou proposed to utilize classifications of lexical semantic and linguistic features in the recognition of online spam audits. The examination results appeared that by consolidating numerous linguistic features of surveys, the recognition execution of spam audits can be incredibly enhanced, contrasting and the best in class strategies [4].

### 2.3 Mobile App Recommendation

Xiao Xia, Xiaodong Wang, and Xingming Zhou proposed a novel method recommendation by making us of global details about applications, where the dangerous development of portable applications offers meet people's high expectations of application revelation. They also produced suggestions by both examining the metadata and measuring the closeness between applications, utilizing the Latent Semantic Index strategy. They also proposed an assorted qualities measurement–based advancement structure for the improvement of versatile application recommender frameworks. To execute the structure, they assist show the framework advancement as a multi-criteria streamlining issue and plan a rank accumulation scheme to settle it [22].
Hengshu Zhu, Hui Xiong,Yong Ge and Enhong Chen developed a phony application recommender system by using the privacy and security alertness. They composed a versatile and programmed approach for evaluating the security dangers of Mobile Apps. To examine both Apps 'reputation and clients' security inclinations for suggestions, they presented an adaptable App suggestion strategy in light of the advanced portfolio hypothesis. Especially, they too built up an App hash tree to effectively turn upward Apps in suggestion [17].

Bin Liu, Deguang Kong, Lei Cen, Neil Zhenqiang Gong, Hongxia Jin and Hui Xiong presented the first methodical study on consolidating both interest-usefulness collaborations and clients' security inclinations to perform customized Application suggestions. They recommended a method for taking the trade off among functionality & client’s privacy inclination [5]

Donghwan Bae, Keejun Han ; Park, J. ; Yi, M.Y., proposed AppTrends, which consolidates a chart based strategy for application suggestion in the Android OS environment. Their trial results got from the field utilization record of more than 4 million applications unmistakably demonstrate that the proposed chart based suggestion model is more precise than the Slope One Model [6]

### 2.4 Internet Water Armies

Kun Wang, Yang Xiao, and Zhen Xiao quantified the Internet Water Army's conduct from numerous measurements. At that point they chose a few successful components as the preparation model and utilize machine learning strategies for order. In view of the conduct of clients review the remark, they proposed a model to quantify the impact of Internet Water Army. With a specific end goal to lessen the impact coefficient, they proposed another direct time many-sided quality online calculations named

MEIWA. The new calculation results demonstrated that the impact is decreased to one 6th of the succession procedure which is utilized as a matter of course with guaranteeing clients' survey remarks propensities [18].

Wang Xiang, Zhang Zhilin ; Yu Xiang ; Jia Yan ; Zhou Bin ; Li Shasha consider the individual and gathering qualities of sorted out notices. A classifier is developed taking into account the individual and gathering attributes to recognize them. Broad test results on three genuine datasets show that our strategy in light of individual and gathering qualities utilizing SVM model (IGCSVM) is compelling in recognizing sorted out publications and superior to anything existing techniques. They examine finding the promoters taking into account the distinguished sorted out publications of our IGCSVM technique. Their analyses demonstrate that it is viable in identifying promoters [7].

Survey on some important challenges for detecting fraud in mobile application is as follows:

Positioning extortion does not continuously happen in the entire life cycle of an App, so we require identifying the time when misrepresentation happens. Such test can be viewed as identifying the local anomaly rather than global anomaly of portable Apps.

## 2.5 Automatically Detecting of Ranking

Vaishali Date, Dipali Dongare, Pooja Jadhav, Tejal Wayal, and Asmita Mali did the survey for finding out the positioning scam founded in the mobile applications. They proposed a system which collects all the positioning details along this feedback and rating results are collected. Then they used the aggregation procedure to do the hypotheses for finding the actual position of the application. They will recover framework with information gathered structure application particle play store for long stretch of time [8].
Tejaswini B. Gade, Prof. Nilesh G. Pardeshi takes a look on different existing procedure for online and web spam based on the positioning of the apps. They find out the local anomaly in the leading sessions and based on these they find out the ranking of the app correctly in the leader board. Finally combined all the evidences and optimized the results [9].

Raghuveer Dagade, Prof. Lomesh Ahire reviewed the detection of ranking for the mobile. They first find that positioning misrepresentation happen in driving sessions and gave a technique to digging driving sessions for each App from its chronicled positioning records. Moreover, they proposed an improvement based accumulation strategy to coordinate all the confirmations for assessing the dependability of driving sessions from portable Apps. That every one of the confirmations can be displayed by factual speculation tests for the novel viewpoint of this methodology, in this way it is anything but difficult to be stretched out with other confirmations from space learning to identify positioning extortion [10].

Prof. Amruta Gadekar, Rajani Gupta, Mamta Kumari, Monika Munswamy worked on fraud detection from user's recommendation. A novel point of perspective of the procedure is that each one of the verifications can be shown by quantifiable hypothesis test; along these lines it is definitely not hard to be connected with various affirmations from space data to recognize situating deception. They did intensification of situating distortion area methodology is performed with other convenient App related organizations [11]

Prajakta Gayke and prof. Sanjay Thakre add to a situating blackmail disclosure structure for versatile Apps. They proposed a change based aggregate framework to consolidate each one of the verifications for evaluating the legitimacy of driving sessions from compact Apps. A novel perspective of this approach is that each one of the verifications can be shown by quantifiable hypothesis tests; along these lines it is definitely not hard to be connected with various affirmations from space data to recognize situating deception. They plan to think more practical deception affirms and dismember the unmoving relationship among rating, overview and rankings. They want to enhance their situating distortion area approach with other convenient App related organizations [12].

Catarina Moreira, Bruno Martins, P´avel Calado contended that unsupervised rank total techniques give a sound methodology for consolidating different estimators of mastery, got from the printed substance, from the diagram structure of the group of specialists, and from master profile data. Probes a dataset of scholastic productions indicate extremely aggressive results in terms of P@5 and MAP, bearing witness to for the ampleness of the proposed approaches. This is especially fascinating to the application area of scholastic master hunt; subsequent to the significance judgments required by directed methodologies are just barely accessible [13].

V.Mural Krishnan, J.Joseph Elango, T.Mohanraj, Ms.Pushpalatha tend to diagram up an arranging confusion conspicuous verification framework for adaptable Apps. They tend to saw arranging based by and large certifications, rating affirmations and audit based attestations for various arranging compulsion moreover, they tend to foreseen accomplice change based generally accumulation procedure to sort out every one of the proofs for assessing the way of driving sessions from moveable Apps [14].

## 2.6 Latent Dirichlet Allocation (LDA)

Ostrowski, D.A. investigate subject considering so as to demonstrate the systems of Latent Dirichlet Allocation which is a generative probabilistic model for a gathering of discrete information. They assess this procedure from the point of view of grouping and also distinguishing proof of foremost subjects as it is connected to a separated accumulation of Twitter messages. Tests demonstrate that these strategies are powerful for the recognizable proof of

sub-points and in addition to bolster order inside of substantial scale corpora [15].

Bhutada.S, Balaram, V.V.S.S.S. ; Bulusu, V.V. attempted an endeavor is made to propose a multilevel grouping model utilizing Latent Dirichlet Allocation (LDA) approach. In spite of the fact that the presence of Latent Dirichlet Allocation (LDA) is seen in the writing yet a changed model for multilevel arrangement is exhibited which is free of any dialect. Keeping in mind the end goal to accomplish such model numerous current recommendations were viewed as like PLSI, which utilizes the Exceptional Maximization (EM) technique just to prepare the inert classes. The iterative procedure of Latent Dirichlet Allocation (LDA), which yields the multilevel order of the corpus. Theme Modeling is utilized to find the concealed things that swarm the gathering to clarify the reports as indicated by the new subjects [19].

Guolong Liu, Xiaofei Xu ; Ying Zhu ; Li Li propose a multi-characteristic inert dirichlet allotment (MA-LDA) model, a subject examination model in which the time and label properties of miniaturized scale online journals are consolidated into LDA model. By presenting a period variable about the time characteristic, MA-LDA model can choose whether a word ought to show up in hotly debated issues or not. Applying label trait permits MA-LDA model to rank the center words high in results so that the expressiveness of results can be enhanced over the customary LDA model. Exact assessment on genuine information sets show our strategy can identify hotly debated issues precisely and productively with more terms connected with each hotly debated issue found. Our study gives solid proof of the significance of the worldly figure hotly debated issues extraction [20].

## 3. EXISTING SYSYTEM

In the literary texts, while there are some associated work, for example, web ranking spam recognition, online review spam identification and mobile App suggestion, the issue of identifying ranking fraud for versatile Apps is still under-investigated. As a rule, the related works of this study can be assembled into three classifications. The primary classification is about web spam detection, second is identifying online review spam and finally the last classification incorporates the study on mobile app suggestion.

## 4. PROPOSED SYSTEM

With the expansion in the quantity of web Apps, to identify the fake Apps, we have proposed a basic and powerful calculation which recognizes the leading sessions of each Application in light of its chronicled positioning of records. By examining the ranking behavior of apps, we come across that the fraud apps frequently has dissimilar patterns for ranking compared with the normal apps in every leading sessions. Subsequently, will perceive few extortion confirmations from applications chronicled records and expounded to three capacities to get such positioning from misrepresentation confirmations.

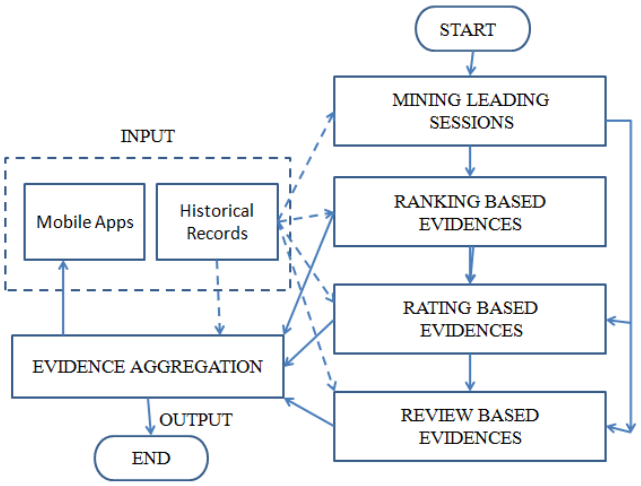


**Fig 1**: System Framework

Further we propose two sorts of fraud evidence taking into account App's review and ratings. It mirrors some peculiarity designs from Apps' authentic rating and survey records. Fig. 1 shows the structure of our positioning extortion framework for versatile applications.

The leading sessions of mobile applications are evidence of interval of popularity, so these driving sessions will include just positioning control. Subsequently, the issue of recognizing positioning extortion is to recognize dangerous driving sessions. Together with the essential errand is to take out the main sessions of a versatile application from its chronicled positioning records.

There are two principle stages for identifying the ranking fraud:

1. Recognizing the leading sessions.
2. Recognizing the evidences of the ranking fraud. Recognizing the leading sessions:

First and foremost mining driving sessions has two sorts of ventures to do with portable extortion applications. To start with, from the Applications authentic positioning records, divulgence of driving occasions is done and after that second joining of adjacent driving occasions is done which appeared for building driving sessions. Completely, some specific count is appeared from the pseudo code of mining sessions of given versatile App also, that estimation can recognize the particular examining so as to drive events and sessions verifiable records one by one.
Recognizing the confirmation of the fraud in ranking.

### Evidence for Ranking

It infers different leading events will be present in a particular leading session. Subsequently by examination of key behavior of driving occasions for finding blackmail affirmations moreover for the application chronicled positioning records, it is been watched that a specific

positioning example is always satisfied by application positioning conduct in a main occasion.

### Evidence for Rating

Past ranking found proofs are useful for recognizable proof reason yet it is definitely not adequate. Determining the "limit time consumption" matter, misrepresentation confirmations acknowledgment is arranged because of app authentic records of rating. While we probably am aware that rating is been finished in the wake of downloading it by the client, also, in the event that the rating is high in leaderboard impressively i.e., pulled in via the greater part of portable application clients. All of a sudden, the assessments in the midst of the fundamental session offers climb to the abnormality plan which happens in the midst of rating deception. These credible records can be used for making rating based affirmations.

### Evidence of Reviews

We are familiar with the survey which contains a few printed remarks as audits by application customer and some time as of late download or using the application customer generally jump at the chance to escape the surveys given by a vast bit of the customers. In this way, albeit due to some past deals with survey spam revelation there still issue on finding the adjacent quirk of audit in driving sessions. So in light of uses audit practices, using so as to position extortion in portable application are identified misrepresentation confirmations.

These three verifications will be combined by an unsupervised evidence-aggregation strategy for assessing the validity of leading sessions from portable apps. To extract the evidences we are using the statistical hypotheses tests for mobile applications. The ranking fraud detection structure is adaptable and can be reached out with other area created confirmations for ranking fraud detection. Finally, we will evaluate the proposed structure with certifiable App data assembled from the Apple's App store for a long time range, i.e., over two years.

## CONCLUSION

This paper surveys different existing strategies utilized for web spam recognition, which is identified with the positioning extortion for portable Apps. Additionally, we have seen references for online survey spam identification and versatile App suggestion.

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.

## REFERENCES

[1]. Vyas Krishna Maheshchandra, and Prof. Ankit P. Vaishnav “A Survey on Review Spam Detection techniques” International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS040505 www.ijert.org , Vol. 4 Issue 04, April-2015.

[2]. Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada “Survey of review spam detection using machine learning techniques” Crawford et al. Journal of Big Data (2015) 2:23 DOI 10.1186/s40537-015-0029-9

[3]. Sushant Kokate, and Bharat Tidke “Fake Review and Brand Spam Detection using J48 Classifier” International Journal of Computer Science and Information Technologies, Vol. 6 (4) , 2015, 3523-3526

[4]. Amir Karam, and Bin Zhou “Online Review Spam Detection by New Linguistic Features” In iConference 2015 Proceedings.

[5]. 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.

[6]. 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.

[7]. 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.

[8]. Vaishali Date, Dipali Dongare, Pooja Jadhav, Tejal Wayal, and Asmita Mali “A Survey on Recognize the Ranking Scam Occurred in Mobile Apps” International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 12, December 2015

[9]. Tejaswini B. Gade, Prof. Nilesh G. Pardeshi “A Survey on Ranking Fraud Detection Using Opinion Mining for Mobile Apps” International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015

[10]. Raghuveer Dagade, Prof. Lomesh Ahire “Review: A Ranking Fraud Detection System for Mobile Apps” International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 11, November 2015

[11]. Prof. Amruta Gadekar, Rajani Gupta, Mamta Kumari, Monika Munswamy “Detection of Ranking Fraud for Mobile App and Prevention from User’s Recommendation” International Journal of Innovative Research in Computer and Communication

Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 9, September 2015
[12]. Prajakta Gayke and prof. Sanjay Thakre "Detection of Ranking Fraud for Mobile App" IOSR Journal of Computer Engineering (IOSR-JCE) 2015
[13]. Catarina Moreira, Bruno Martins, P´avel Calado "Using Rank Aggregation for Expert Search in Academic Digital Libraries" arXiv:1501.05140v1 [cs.IR] 21 Jan 2015
[14]. V.Mural Krishnan, J.Joseph Elango, T.Mohanraj, Ms.Pushpalatha "Tracking and Altering Duplicate Mobile Apps in Website" International Journal for Research in Technological Studies| Vol. 2, Issue 4, March 2015 | ISSN (online): 2348-1439
[15]. Ostrowski, D.A "Using latent dirichlet allocation for topic modelling in twitter" Semantic Computing (ICSC), 2015 IEEE International Conference
[16]. Ms. Meenal M. Shingare, and Prof. S. R. Chaudhary "Web Spam Recognition through Classification Algorithms" © 2014, IJARCSSE
[17]. Hengshu Zhu, Hui Xiong,Yong Ge and Enhong Chen "Mobile App Recommendations with Security and Privacy Awareness" KDD'14, August 24–27, 2014, New York, NY, USA. Copyright 2014 ACM 978-1-4503-2956-9/14/08 ...$15.00
[18]. 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)
[19]. Bhutada.S, Balaram, .V.S.S.S. ; Bulusu, V.V. "Latent Dirichlet allocation Based multilevel classification" Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference
[20]. Guolong Liu, Xiaofei Xu ; Ying Zhu ; Li Li "An Improved Latent Dirichlet Allocation Model for Hot Topic Extraction" Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference
[21]. ShrijinaSreenivasan and B.Lakshmipathi "An Unsupervised Model to detect Web Spam based on Qualified Link Analysis and Language Models" International Journal of Computer Applications (0975 – 8887) Volume 63– No.4, February 2013
[22]. Xiao Xia, Xiaodong Wang, and Xingming Zhou "Evolving Recommender System for Mobile Apps: A Diversity Measurement Approach" Smart Computing Review,vol.3,no.3,June213.