# Detection of Fraud in Mobile Advertising using Machine Learning

**B.Sathyabama, Harshita Singh, Harshit Goraya, Aman Vira**

***Abstract:** With ongoing advancements in the field of technology, mobile advertising has emerged as a platform for publishers to earn profit from their free applications. An online attack commonly known as click fraud or ad fraud has added up to the issue of concerns surfacing mobile advertising. Click fraud is the act of generating illegitimate clicks or data events in order to earn illegal income. Generally, click frauds are generated by infusing the genuine code with some illegitimate bot, which clicks on the ad acting as a potential customer. These click frauds are usually planted by the advertisers or the advertising company so that the number of clicks on the ad increases which will give them the ability to charge the publishers with a hefty sum per number of clicks. A number of studies have determined the risks that click fraud poses to mobile advertising and a few solutions have been proposed to detect click frauds. The solution proposed in this paper comprises of a social network analysis model – to detect and categorize fraudulent clicks and then test sample datasets. This social network analysis model takes into consideration a wide range of parameters from a large group of users. A detailed study is conducted for analyzing these parameters in order to separate the parameters, which affect the click fraud generation process largely. These parameters are then tested and categorized into sample datasets. The mobile advertising industry forms a large part of the revenue generated by the advertising industry. Hence, detection of click fraud in mobile advertising is important to ensure that no illegitimate sources are used to generate this revenue. To be precise, the proposed method touches an accuracy of about 92%.*

***Keywords:** click fraud, add fraud, mobile advertising.*

## I. INTRODUCTION

Recently, mobile advertising has evolved expeditiously as it provides publishers a platform to expand their audience reach by putting their advertisements in mobile applications. Statistically, mobile in-app advertising [1] is estimated to surpass a revenue limit of approximately $17 billion by the end of 2020.

The mobile advertising industry is bilateral as on one hand, it helps developers boost app monetization and on the other hand, it expands the install horizons to which an application can reach. It allows advertisers to take their product out to new audiences and app developers to escalate their application reach to new markets. Another common name for mobile advertising is in-app advertising. This in-app advertising comprises of four major components, namely: 1) The advertiser, 2) The user, 3) The publisher, and 4) The ad network. A user, in mobile advertising is one who views the ad. The owner of the product which is being advertised is considered to be the advertiser whereas the person to whom the application, in which the advertisement is advertised, belongs is the publisher. Furthermore, the third party who links advertisers to publishers is called the ad network. These ad networks aim at generating as a percentage of publisher's profit.

Click fraud is a type of fraud, which puts the Cost Per Click model in jeopardy. Certain fraudulent sources came up with the idea of generating illegitimate clicks on the advertisement, which encouraged unethical groups to hire these sources to increase the amount of user action on their advertisement and generate money from it. A click fraud occurs when a bot, a computer code, an automated script or a person, pretending to be genuine user, generates a random number of clicks on an advertisement without any legitimate interest in it. Click frauds pose a huge risk to the advertising industry as businesses are estimated to lose $26 billion by 2020, $29 billion by 2021 and $32 billion by 2022(A study by cyber security company, Cheq [2]), provided that these frauds remain unchecked. Till date, multiple methods [3,5] have been proposed and various studies have been conducted for efficient detection and elimination of click fraud, for example, functionality test to characterize user engagement, UI state transition graphs to check against a set of heuristic based rules for detecting fraudulent behaviors, a DECAF method to control the execution of fraud detection policies.

In this paper, we design a systematic approach to categorize fraudulent clicks. This approach's adoption of a set of custom-designed attributes enables the effective detection of click fraud. In addition, multiple correlation techniques including heat map using correlation have been used to successfully eliminate the detected frauds.

## II. A TERMINOLOGY OF AD FRAUDS

There exists detailed literature on ad frauds in windows phone and web applications but only a tiny part of it comprises of such frauds on Android platform.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author

**B.Sathyabama*,** Assistant professor, IT department of SRM Institute of Science and Technology.
**Harshita Singh,** Pursuing bachelors, Information Technology, Chennai.
**Harshit Goraya,** Bachelor's in Information Technology,SRM Institute of Science and Technology.
**Aman Vira,** Student at SRM Institute of Science and Technology with an extensive background in Information Technology

Fig. 1 represents the terminology, which defines the ad fraud types categorized so far [4].

There are two major classes of ad fraud, which are further divided into sub-classes specifying different fraudulent behaviors which mobile advertising is vulnerable to.

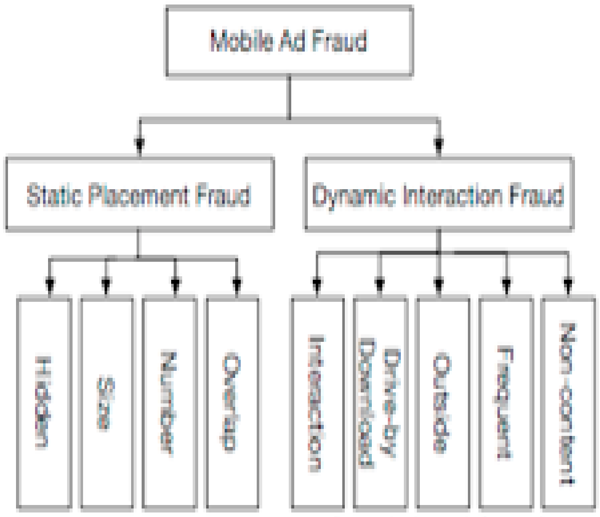


**Fig. 1. A terminology of mobile ad frauds.**

The two main classes are: Static Placement Frauds and Dynamic Interaction Frauds. Static means that these frauds can be detected by static information and usually occur in a single state of UI. When the placement aspects such as location, number of views and size of the ad are exploited by the fraud, such frauds are referred as placement frauds. Whereas, dynamic implies to ads which can be detected using runtime and interaction stands for the frauds which exploit UI scenarios involving multiple UI states.

Four specific types of behaviors have been identified by far, which fall under the category of static placement frauds. These include:

1) Ad Size frauds*:* There are not any guidelines on the size ratio or ad size. However, the size ratio of ad size to screen size is advised to be reasonable, so that the users can view ads conveniently. A fraud can be created by either making the ad smaller than the visible size so that it unintentionally gets clicked on while the user engages with app UI or by making it oversized to attract viewer’s attention.
2) Ad Overlap fraud: These ads manipulate users into clicking on the ad by placing it over any oversized ad. Scam advertisers usually place the app to cover the areas of user’s interest so that the viewer is forced into clicking on it.
3) Ad Hidden fraud: There are cases where application developers hide ads (between app’s UI components) so that users are trapped in the illusion of an ad-free app.
4) Ad Number fraud: The number of ads in an app is usually limited so that the functionality of the application is not compromised.

Five types of ads fall under the category of dynamic interaction frauds. These include:

1) Interaction ad fraud: Fraudulent behavior can be injected into an app by placing ads while entering or before exiting the application. These are ad views, which are placed when transitioning between UI states.
2) Frequent ad fraud: The probability of ad views is maximized by manipulating the application’s UI. Each time the user clicks on the app’s core content, an add is displayed as an attempt to increase the number of interstitial clicks.
3) Drive by download ad fraud: At times, the publisher is paid on the basis of cost per action model. A typical example of such fraud is placing an ad in a manner in the app which triggers unintentional downloads when clicked on it.
4) Non-content ad fraud: At times, ads are placed on non-content pages by the app developer. Pages containing error, login, thank you or exit screens are called non-content pages. These ads are placed with an intention of confusing the user between real content and ads.
5) Outside ad fraud: Fraudulent behavior is infused into the app by placing ads in app background, or sometimes even outside the app environment. Such practices are intended to carry out outside ad fraud.

## III. RELATED WORK

In the domain of mobile advertising, researches conducted previously have proposed the use of pattern recognition for categorizing fraudulent acts. Zhang et al. [6] proposed a combination of Cost-Sensitive Back Propagation Neural network (CSBFNN) and the novel Artificial Bee Colony (ABC) algorithm in their research towards detection of click fraud in mobile advertising. They optimized feature selection process with BPNN connection weights using ABC to alter the relation between feature and weights.

An empirical study conducted by Cho et al. [7] highlighted the threats posed by click generation attack on mobile advertising networks. They designed an automated click generation attack and evaluated the efficiency of eight favored advertising networks by carrying out real attacks on them. This study revealed that approximately 75% of the advertising networks were in jeopardy.

A detailed research carried out by Kingston et al. [8] disclosed three primary reasons why law enforcement oftenly falls behind fraudsters in case of detecting frauds in mobile advertising. Consequently, the research also categorized frauds into specialized fraud plans, generic fraud plans and red flags.

Furthermore, Gupta et al. [9] conducted another study, which proposed an online hybrid model for online in-auction fraud detection. This research showcased shilling as the primary reason behind online auction frauds. They suggested the use of Hidden Markov Model to counter possibilities of shilling. The Hidden Markov Model comprises of a statistical model, which generates sequence-based probability on the tenders put up by the users.

Subsequently, an alternative research carried out by Taneja et al. [10] proposed a novel framework consisting of the use of Recursive Feature Elimination (RFE) during the feature selection process and Hellinger Distance Division Tree (HDDT) during classification.

The accuracy of this method was approximated to be 64.07%, which was the best as compared to other existing methods.

Crussell et al. [11] devised an analysis tool and named in MadFraud. This tool triggers emulators to expose frauds by running simultaneously on multiple apps. They put forward a three step systematic approach: 1) built HTTP request tress, 2) identified ad request pages by machine learning and 3) detected clicks in HTTP request trees using heuristics.

Vasumathi et al. [12] experimented on a set of data mining algorithms to determine the possibilities of click spam in mobile advertising. Certain attributes were taken into consideration from application developers, user ratings, ad-control locations and publisher's applications.

## IV. PROPOSED SOLUTION

In researches carried out till now, detection of click fraud has been done by using different approaches including data mining, CFC model, and neural networks. The loophole in the process of click fraud detection is that the interest of publishers and ad networks lie opposite to each other. On one hand, publishers want to flag illegitimate ad clicks so as to avoid paying for them, whereas on the other hand, ad networks want to consider all ad clicks as legitimate so as to be able to charge the publisher for them. Thus, a detection system should exist which serves the interest of both parties.

Here, we propose a method including the use of social network analysis algorithm. The algorithm is trained at first using a dataset having various parameters such as name, age, daily time spent on site, area of interest, clicked on ad, timestamp, gender, city etc. A detailed research is conducted on these datasets and different graphs are plotted to determine a relationship between these parameters so that a precise and accurate training of the algorithm is ensured. Graphs such as distplot, scatter plot, bar chart are used for correlation purposes.

Once an accurate correlation is established between these parameters, approximate values are determined and fed into the algorithm to train it. After the algorithm is trained, the input data is fed to the algorithm. Different correlation coefficients are calculated by the algorithm for this input data and other detailed graphs are plotted which effectively produce the different values for legitimate and illegitimate ads.

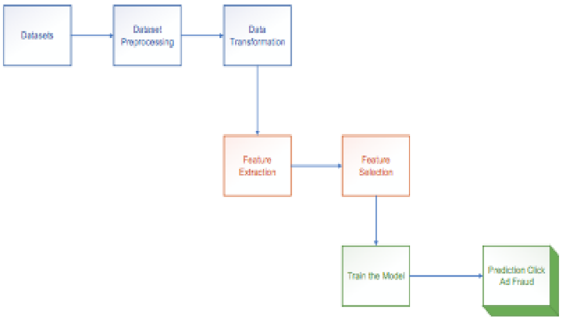


**Fig. 2. Detailed architecture of click fraud detection model..**

A dataset of approximately 1000 data is taken into consideration for training the algorithm. As shown in fig. 2, the process of click fraud detection consists of the following steps- datasets, dataset processing, data transformation, feature extraction, feature selection, train the model and Predicting click ad fraud.

## V. RESULTS AND DISCUSSION

To make the testing of our model as close as possible to testing with real ad traffic, we plotted a number of graphs and generated multiple figures among different parameters of the dataset. A multiple number of iterations were created to ease the generation of a large chunk of instances. This helped increase the precision and accuracy of our model.

Fig. 3 shows a seaborn violin plot between daily time spent on site and weekday versus clicked on ad or not. It represents a data visualization model of the daily time a user spends on site or the time user spends on a site on a weekday and whether the user clicks on an ad or not during this time.

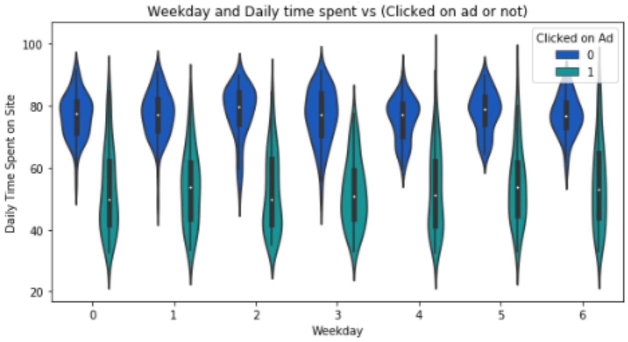


**Fig. 3. A seaborn violin plot of Weekday and Daily time spent versus Clicked on ad or not.**

Fig.4 represents four different types of styled scatter plots. First plot is a styled scatter plot of the income coming to the advertiser from a particular area versus daily time spent on site by users in that area. It gives an average estimate of how much time a user is likely to spend on a site in order to contribute to a specific percentage of advertiser's income. Second plot is a scatter plot between area income and age. It gives an idea of the approximate earning of the advertiser from a user belonging to a specific age group.

Third plot is a plot between daily internet usage and daily time spent on publisher's site. It depicts the amount of internet usage and how much of that time the user spends on publisher's site. Fourth plot gives an approximate estimation of the users internet usage according to their age.

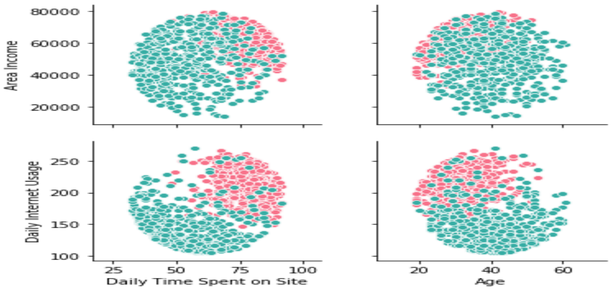
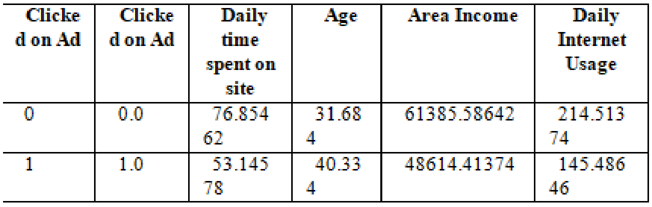


**Fig. 4. Styled Scatter Plots.**

Similarly a number of graphs are plotted between parameters of the dataset and a set of precise values is determined which are fed into the algorithm. These values are depicted in Table-I.

Table- I: Table depicting set of values to be fed into the algorithm



| Clicked on Ad | Clicked on Ad | Daily time spent on site | Age | Area Income | Daily Internet Usage |
|---|---|---|---|---|---|
| 0 | 0.0 | 76.85462 | 31.684 | 61385.58642 | 214.51374 |
| 1 | 1.0 | 53.14578 | 40.334 | 48614.41374 | 145.48646 |

## VI. CONCLUSION

The social network analysis model proposed is capable of detecting click frauds up to an accuracy of 92%. The analysis and comparison of different parameters helps put forward a clear and distinct vision towards the parameters, which impact the click fraud detection process. Furthermore, the model proposed can be made more efficient by including more parameters from the publisher's browsing data and a detailed study of the nature of the website.

## REFERENCES

1. "Total global mobile in-app advertising revenues 2015-2020", [Online]. Available: https://www.statista.com/statistics/220149/total-wporldwide-mobile-app-advertising-revenues/ [Accessed: February 2020] .
2. "Report: Ad Fraud to hit $23 billion isn't going down",[Online]. Available: https://adage.com/article/digital/report-ad-fraud-hit-23-billion-isnt-going-down/2174721 [Accessed: February 2020].
3. H. Xu, D. Liu, A. Koehl, H. Wang, A. Stavrou((2014, Sepetember),"Click fraud detection on the advertiser Side", in proceedings of the 19th European Symposium on Research in Computer Security(ESORICS), Poland, Europe.
4. F. Dong, H. Wang, L. Li, Y.Guo, T.F.Bissyande, T.Liu, G.Xu, J.Klein(2017, July),"FraudDroid: automated ad fraud detetcion for android apps", in proceedings of ACM Symposium on Principles of Distributed Computing , Washington DC, US.
5. B. Liu, S. Nath, R. Govindam, J.Liu,(2014, April), "DECAF: Detetcing and characterizing ad fraud in mobile apps", in proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation, Seattle, WA, US.
6. X. Zhang, X. Liu, H. Guo(2018, December), "A Click Fraud detection scheme based on cost sensitive BPNN and ABC in mobile advertising", 4th IEEE International Conference on Computer and Communications(ICCC), Chengdu, China.
7. G. Cho, J. Cho, H. Kim, Y. Song(2015, August), "An empirical study of click fraud in mobile advertising Networks", in proceedings of the 10th International Conference on Availability, Reliability and Security, France.
8. J.K.C. Kingston(2017,May), "Representing, Reasoning and Predicting fraud using fraud plans", in IEEE 11th International Conference on Research Challenges in Information Science(RCIS), Brighton, UK.
9. P. Gupta, A. Mundra(2015, May), " Online in-auction fraud detection using online hybrid model", published in International Conference on Computing, Communication and Automation, Ghaziabad, India
10. M. Taneja, A. Panwar, K. Garg, S. Sharma(2015, August), "prediction of click fraud in mobile advertising", in IEEE 18th International Conference on Contemporary Computing(IC3), Noida, India.
11. J. Crussell, R. Stevens, H. Chen(2014,June),"MadFraud: Investigating ad fraud in android applications", in proceedings of the 12th annual international conference on Mobile systems, applications and services, Bretton Woods, NH, US.
12. D. Vasumathi, M.S. Vani, R. Bhramaramba, O.Y. Babu(2015,April), "Data Mining approach to filter click-spam in mobile ad networks", in international Conference on Computer Science, Data Mining and Mechanical Engineering(ICCDMME'2015), Bangkok.

## AUTHORS PROFILE

**B.Sathyabama** completed B.tech from KCG College of Technology in 2014 and Mtech from Sathyabama University, Chennai in 2016. I currently work at the IT department of SRM Institute of Science and Technology as an Assistant professor. I have guided more than 50 undergraduate students with their projects and have a valuable teaching experience of 3 valuable years. I have published more than 10 papers in National and International journals till date.

**Harshita Singh** is a student, pursuing bachelors in Information Technology, Chennai. I am a tech enthusiast who loves to write and design code. I believe an individual's growth escalates by seeking discomfort and innovation, as a result of which, I have been an active part of numerous tech symposiums and college events. Alongside practicing programming and writing code, I also focus on combining creative ideas for the design of applications and websites, keeping customer's convenience in mind. My areas of interest are data structures, blockchain, artificial intelligence and human computer interaction.

**Harshit Goraya** is pursuing his bachelor's in Information Technology from SRM Institute of Science and Technology. He is constantly driven by the rapid innovation in the I.T. industry and how the merge of businesses and technologies is creating a new category of market leading products. His areas of interest include machine learning, business operations and digital marketing through the use of new gen tools that include artificial intelligence.. He has worked on several projects that include the development and usage of databases, web development and UI designing. He is currently working as an operations manager in ML driven projects at Quantiphi Inc.

**Aman Vira** is a talented student at SRM Institute of Science and Technology with an extensive background in Information Technology. He specializes in the field where the user gets to feel connected to I.T. He is focused in the fields of UI/UX, Front-End Development, Digital Marketing, and Graphic Designing. He has worked on the front-end development for the project ERP Attendance Management System. He has worked as a UI Designer and Developer at Zelican Infotech Pvt Ltd and as a Systems Engineer at Infosys Ltd.