IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Fraud Advertisement Click Detection

Click Fraud Detection

Sandhya Ch[1]
Assistant Professor, St. Peter's Engineering College,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

Sneha P[2]
B.Tech. 4th Year Students, Department of CSE,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

Thumu Nayana[3]
B.Tech. 4th Year Students, Department of CSE,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

Chilipireddy Hasini Reddy[4]
B.Tech. 4th Year Students, Department of CSE,
Opposite TS Forest Academy Dullapally, Maisammaguda
Medchal, Hyderabad, Telangana 500043

***Abstract*— In this project we are detecting fraud click from internet website advertisements. Website owners will get money from online advertisers upon each click from owner's website to advertiser's site. Each click will lead user from owner's website to advertiser's website. This moneymaking process may encourage fraud users to go for more clicks on this websites to get more money. Fraud users just visits the website and does not do any operations such as app downloading or form filling or any other process make more money. To detect such fraud click we are implementing machine learning approach where application will automatically learn about genuine or fake clicks.**

***Keywords*— recurrent neural network , super vector machines, artificial neural networks**

I.INTRODUCTION

In recent times, digital advertising has gained popularity as a means for publishers to monetize their free or paid applications. One of the main concerns in the in-app advertising industry is the popular attack known as "click fraud". Click Fraud is an act of repeated clicking on an advertisement, not because of interest in this ad, but rather as a way to generate illegal money for the application publisher. Attackers make use of "click fraud" to generate money illegally. The repeated clicking on a particular advertisement not out of interest in it but solely to earn money. To avoid such fraudulent activity, a detection model has been proposed to differentiate between legal and illegal users. This has been implemented using RNN algorithm which is more efficient and provides accurate results.

II.LITERATURE SURVEY

We have referred many papers regarding click fraud detection system. In most of the cases they have implemented using classification algorithms like SVM, KNN which are providing the results with the accuracy of 95.6%, 91%, 98%.

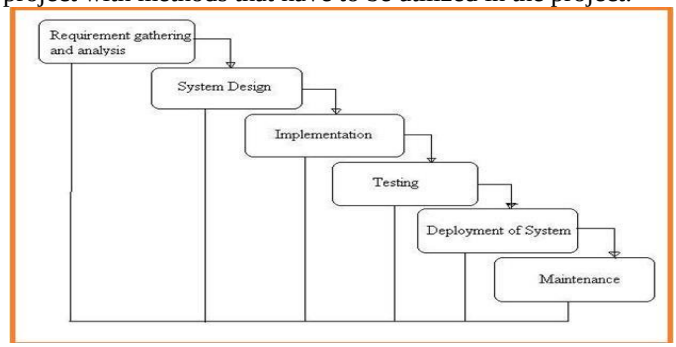
III. IMPLEMENTATION

A. *Problem Statement:*

Currently we are having lot of approaches and systems for detecting fraud clicks but they do not produce correct results when the dataset is huge. Generally when the number of users on website is in huge amount, then it becomes very difficult to predict who are a true user and a fraud. By using this system any user or client can know whether a particular website is genuine or fraudulent in order to publish their advertisements by looking at the number of fraud clicks.

B. *Sysytem Design:*

The System Design has divided into three types like GUI Designing, UML Designing with avails in development of project in facile way with different actor and it utilizes case by utilize case diagram, flow of the project utilizing sequence, Class diagram gives information about different classes project with methods that have to be utilized in the project.



C. *Online Advertisement Concepts:*

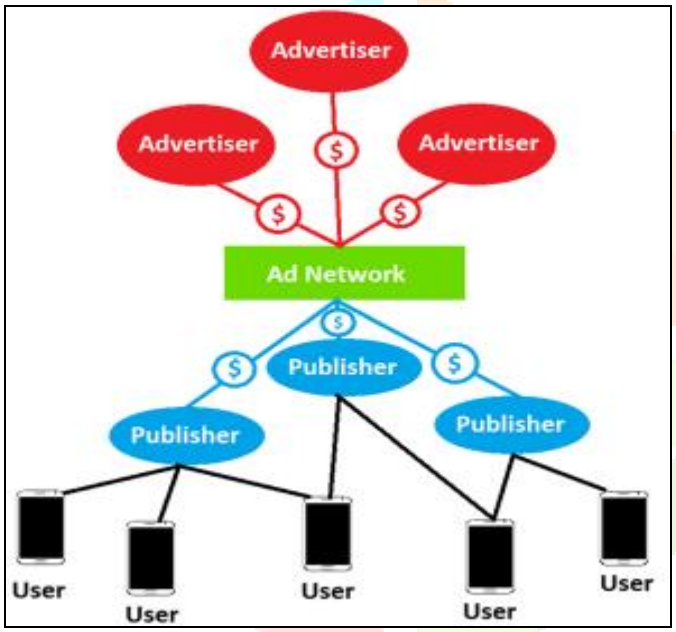
Online advertisement involves primarily four agents:

1.**Avertiser**: The advertiser or announcer wishes to publicize their product or service to a target public that may be interested in consuming what the announcer has to offer.

2.**Publisher**: A publisher is someone who has their own publicity platform, such as a site or a blog, and is able to show the advertiser's product to the visitors of said platform.Often the content in the publisher's platform and that of the advertiser's product are of similar nature, though not always.

3. **Users**: Users are any visitors to the publisher's website, who may be interested and click on the advertiser's ad,perhaps even buying their product or otherwise doing a directly financially relevant action for the advertiser,thanks to the publisher's publicity.

4.**Ad network**: At last, the ad network is a middleman between the advertisers and publishers. Their objective is to connect advertisers interested in the publicity with publishers willing to offer such service. Some ad networks take a dynamic approach by deciding which ad to show for a given user accessing a given publisher's site.



D. *Module Discription:*

There are five modules present in the system that has been developed they are:

1. Upload click dataset
2. Browse
3. Detection of click fraud
4. Genuine and fraud click chart
5. Exit

To detect such fraud clicks we will implement below steps in machine learning algorithm, first we will upload past one or half hour users click dataset and this dataset will have following details such as record, channel, OS, device, app, IP, click time, is attributed. In this step we will read all attributes and records from dataset. In this step we will apply machine learning concept the idea is we will capture time spend by user on current page after click. If this is genuine user or click then he will spend more time to fill form or download app. If he spent at least more than 10 seconds or he visit this page once in last half hour dataset then we will detect this click as genuine otherwise fake. All Fake users will spend much less time on each click or page as they don't have to download any app or fill form. So just by monitoring spend time on each click page we can detect fake or genuine click.

Many studies evaluated click fraud attacks in the literature, and some proposed solutions to detect it using, for example, random ads to detect if any automated ad clicking tool is used, or cryptographically authenticating users, etc. Because machine learning is being actively thought for network attacks, fraudulent activities, spam email we propose in this paper to classify fraudulent clicks by adopting some custom-designed features and by testing using multiple pattern recognition concepts such as KNN, ANN and SVM. In addition, differently from existing literature that adopted decision trees, SVM, and Bayesian theory to classify fraudulent clicks, our work distinguishes itself by introducing a new party trusted by both advertisers and ad network that manages the click fraud detection by crowd sourcing ad requests.

IV. PROCESS FLOW

A flow of events is a sequence of transactions (or events) performed by the system. They typically contain very detailed information, written in terms of what the system should do, not how the system accomplishes the task. Flow of events are created as separate files or documents in your favorite text editor and then attached or linked to a use case using the Files tab of a model element. In this Method, we implemented it with by using some software's.

1. **Java:**

Java is a mostly used programming language these days which is interactive, interpreted and object oriented. It gives special importance to code legibility and makes the computer specialist tasks easy by writing code in a small number of lines.

2. Recurrent neural networks:

Recurrent Neural Network is a generalization of feed forward neural network that has an internal memory. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one computation. After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input.

3. MySQL:

MySQL is an open source relational database management system. A relational database organizes data into one or more data tables in which data types may be related to each other; these relations help structure the data. SQL is a language programmers use to create, modify and extract data from the relational database, as well as control user access to the database.

A.Results:

Results obtained match the expectations. Although not yet tested in a real environment, the system has shown good performance against different types of attack, with at least two of the rules identifying each of those attempted frauds. All the attacks attempted by the electronic user were identified, but it is emphasized the importance of selecting appropriate values for rules' weights, as different values for key rules would have led to the system not correctly classifying attacks as such. The chosen weights for our tests aren't the only possible values and in a real world scenario through experimentation would be necessary. Low-frequency attacks are still a threat as

malicious clicks spread over long time intervals will most likely not be automatically detected, and if done in a large scale with multiple IP addresses could affect the agents protected by the system. There is however issues with those methods of attack themselves, namely obtaining access to this significant number of IPs and obtaining economical return from simpler attempts of the attack, given the low return for a single click

B.Figures:

The GUI interface will look like this

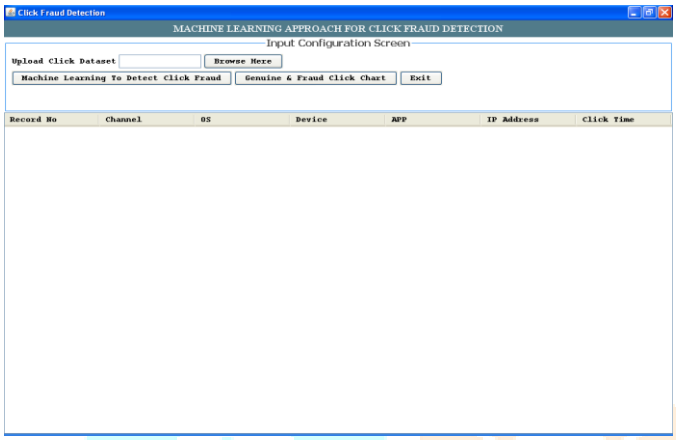


Fig. 1. Example of a Fraud click detecting using web portal

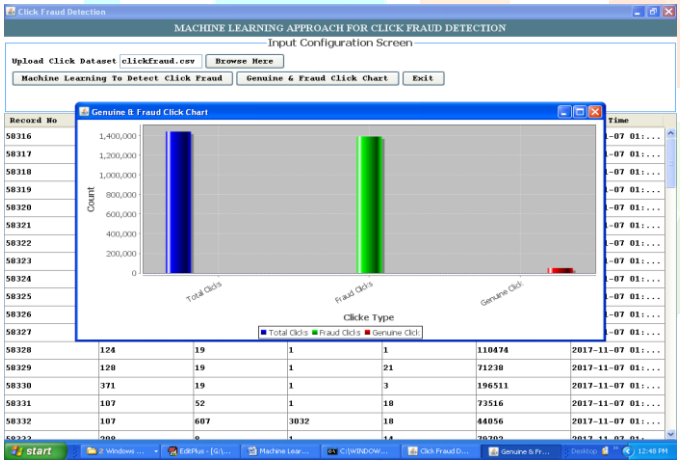


Fig. 2. Example of output of the detection

V.CONCLUSION

Fraud Advertisement Click Detection is developed to provide an analysis of a website, whether it is a genuine one or not. Just by looking at the analysis a person who wants to advertise their products can decide whether he can advertise in that particular website or not. In this system, we proposed generating ad-related information and extracting features such as the ratio of unique IPs in the total number of clicks per publishers. To classify malicious publishers, we evaluated three different classifiers: KNN, SVM, and ANN. All three classifiers gave very promising results. KNN seems to be the best classifier for our data (almost 98%) followed by SVM (almost 96%) then ANN (almost 93%). Low values of the FPR (false positive rate) are considered as an important achievement in our click fraud detection system. Future work includes the evaluation of deep learning with malicious publishers.

References

[1]Ash, cloud-iq, "In-App Advertising Growing in Popularity in UK", [Online]. Available: http://blog.cloud-iq.com/blog/in-appadvertising-growing-in-popularity-in-uk.

[2]dmnews, "Bots Mobilize", December 2015 [Online]. Available: http://www.dmnews.com/mobile-marketing/bots-mobilize/article/291566/.

[3]L. Handley, "Businesses could lose $16.4 billion to online advertising fraud in 2017: Report", CNBC, 13 April 2017. Available: https://www.cnbc.com/2017/03/15/ businesses-could-lose-164-billion-to-online-advert-fraud-in-2017.html.

[4]M. Awad, R. Khanna, "Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers", ISBN-13: 978-1430259893, 2015.