IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

Ad Fraud Detector for Mobile Applications

Mr. Aniruddha Ambekar, Mr. Hrithik Kucheria, Ms. Nikita Aher,
Mr. Hrishikesh Sidwadkar, Mr. G. J. Navale
Department of Computer Engineering
AISSMS Institute of Information Technology, Pune, Maharashtra, India

Abstract: *Ad fraud happens in cost per click ad networks where publishers charge advertisers for each click. Ad fraud is posing the huge loss to the mobileadvertising industry. The conventional technologies use ensemble machine learning methods, neglecting the cost of incorrect classification for a fraud publisher is higher than a normal publisher. An effective classification model for variable ad fraud is proposed in this paper. Cost-sensitive Back Propagation Neural Network is combined with the novel Artificial Bee Colony algorithm in this research (CSBPNN-ABC). Feature selection is synchronously optimized with BPNN connection weights by ABC to reduce the interaction between features and weights. Cost Parameters are added to BPNN by correcting the error function. Experiments on real world click data in mobile advertising show that its superior classification performance compared with the state-of-the-art technology.*

Keywords: Click Fraud, Cost-sensitive Back Propagation Neural Network, Artificial Bee Colony, Feature Selection.

I. INTRODUCTION

The online fraud detector identifies click requests and examines related user input events to detect click fraud. AdFraudDetector employs an ad request tree model to identify click requests accurately and efficiently. An ad request tree is a tree whose root is the ad request and other nodes are causally related HTTP requests. To build the tree, AdFraudDetector monitors the traffic through the pattern matching module and identifies ad requests using the offline-generated patterns. Then, ad requests are passed to the module of ad request tree construction to start building trees. Click can be identified from the ad request tree simply by finding the node of the request whose response header contains “location” field. Finally, the identified click request passes through the fraud checker.

II. MOTIVATION

Ad fraud detector is an inherently nebulous field. In broad terms, it consists of identifying the intention behind the received clicks, given only technical data (such as the IP address and other information provided by HTTP requests) and contextual information (previous accesses from the same IP,for example).Thus, malicious click detectioninvolves comparing every access behavior with what’s expected from normal users, but, that is difficult to formalize and context dependent behavior is non deterministic and context dependent. Moreover, current literature is lacking in certain aspects. Many studies focus on the advertiser’s side, such as, and few on how to apply click fraud detection techniques on ad.

III. OBJECTIVE

The objective is to propose a system of click fraud detection and prevention applied to anad network. Ad network’s interest in preventing frauds lies in their relationship with advertisers: the latter wishes for the best click quality possible (or, the biggest number of interested users over the least of resources used possible), and directly pays the ad network for such traffic. Bad clicks directly hurt the advertisers’ goals, and thus their wish to interact with networks that may bring such clicks.

IV. PROPOSED SYSTEM

This section overviews of the structure of the proposed system. It mainly consists of Dataset Preprocessing, Feature Extraction, SVM and Decision Tree Algorithm and Detection of the fraud as shown below:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

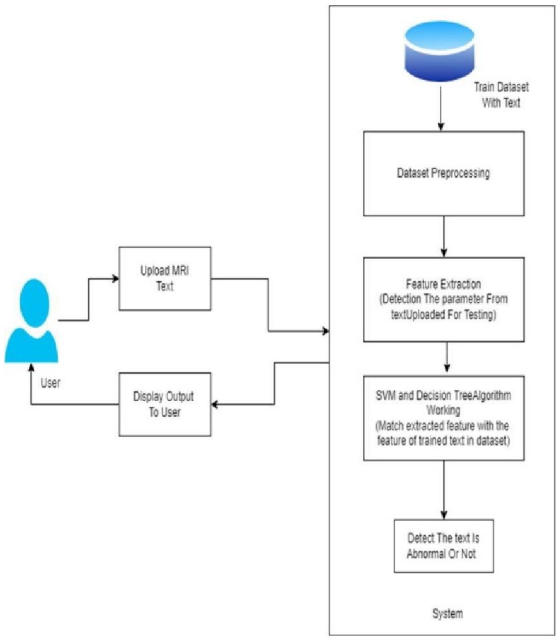

Figure 1: Schematic Diagram of Ad FraudDetection System

V. MODULES

5.1 Admin

In this module, the Admin has to log in by using valid user name and password. After login successful he can do some operations such as View All Users and Authorize, View All E-Commerce Website and Authorize, View All Products and Reviews, View All Products Early Reviews, View All Keyword Search Details, View All Products Search Ratio, View All Keyword Search Results, View All Product Review Rank Results.

5.2 View and Authorize Users

In this module, the admin can view the list of users who all registered the. In this, admincan view the user's details such as, username, email, address and admin authorizes the users.

5.3 View Chart Results

View All Products Search Ratio, View All Keyword Search Results, View All Product Review Rank Results.

5.4 E-Commerce User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login byusing authorized user name and password Once Login is successful user will do some operations like Add Products, View All Products with reviews, View All Early Product's reviews, View All PurchasedTransactions.

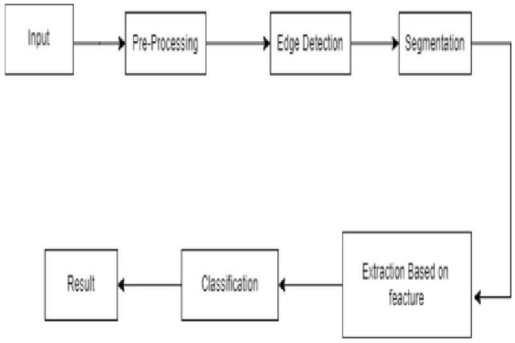

Figure 2: Process flow of Ad Fraud Detection System

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

5.5 End User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will best or to the database. After registration successful, he has to login byusing authorized user name and password. Once Login is successful user will do some operations like Manage Account, Search Products by keyword and Purchase, View Your Search Transactions View.

VI. ALGORITHM

6.1 SVM Algorithm

Support Vector Machine also known as SVM, in machine learning it is a supervised algorithm that can be used for regression, modification and classifications. Classification Issues are the main example of support vector machine algorithm. It's a supervised learning algorithm that is mainly used to classify data into different classes. SVM trains on a set of label data. The main advantage of SVM is that it can be used for both classification and regression problems. This is exactly what SVM does! It tries to find a line/ hyperplane (in multidimensional space) that separates these two classes. Then it classifies the new point depending on whether it lies on the positive or negativeside of the hyperplane depending on the classes to predict.

6.2 Decision Tree

Decision tree is a classifier where the feature of datasets are represented by the internal nodes of the decision tree, decision rules and outcomes are represented by branches and leaf node respectively. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision trees help you to evaluate your options. Decision Trees are excellent tools for helping you to choose between several courses of action. They provide a highly effective structure within which you can lay out options and investigate the possible outcomes of choosing those options.

VII. WORKING MODULE

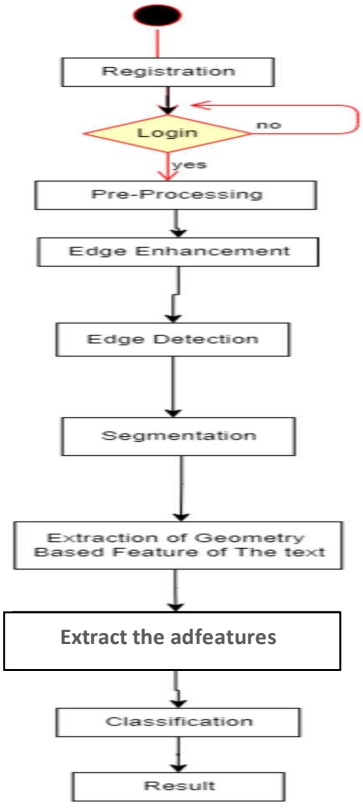

Figure 3: Activity Diagram of Ad FraudDetection System

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

The activity diagram explains the execution of Ad Fraud Detection system in detail, showing every step in the system. The organization first need to register to the system, so as it will be able to login to the distributed environment of the ad fraud detection. The pre-processing is done by the application itself and helps it to train the module using the dataset. The dataset hasmultiple features on which the fraud detection is dependent upon. The classification module is used to differentiate the features using PCA algorithm. The algorithm only selects the required features and drops the unnecessary ones. Then finally the organization checks whether the ad fraud is detected or notdetected. This system helps the organization detect the malicious behaviour of the fraudulent ad.

VIII. RESULT

It is generating random clicks on a link in order to extract illegitimate revenue from the advertisers. It is the attempt to defraud advertisers, publishers or supply partners by exploiting mobile advertising technology. The objective of fraudsters is to steal from advertising budgets. Below are the results of the application used to detect ad fraud using the features like ip, app, device, os, channel.

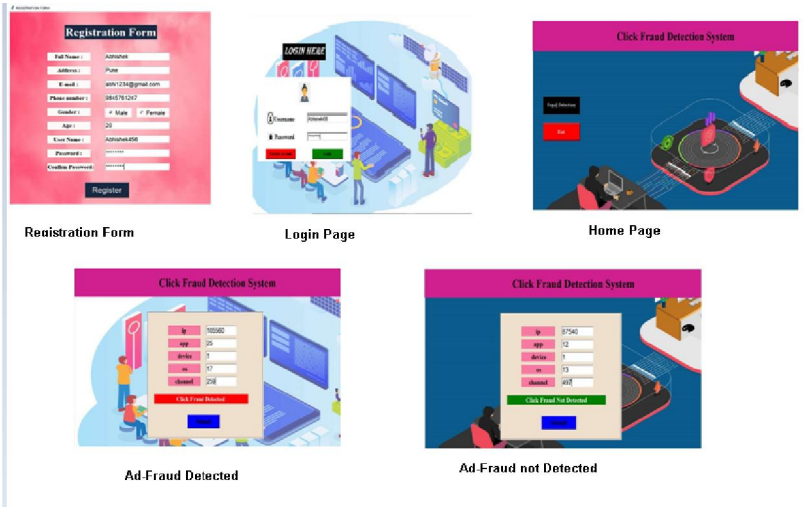

Fig. 4. Results

IX. CONCLUSION

The working approach of Ad fraud detection is very frisk and efficient for deploying to any organization. Ad fraud Detection system generates the fraudulent ad behaviour and if the add is valid and ready to publish. The server side go through for the system is very approachable. The tree request model is formidable to ad fraud detection system. Exact and Probabilistic patterns aregenerated using this system. The former mentioned patterns are used in the detection of fraudulent Ad .Research shows Ad Fraud Detector achieves high click fraud detection accuracy with a negligible runtime overhead. In the future, we plan to combine static analysis with the traffic analysis to improve the accuracy of ad request identification and explore attacks designedto evade Ad Fraud Detector.

REFERENCES

[1]. C. M. R Haider, A. Iqbal, A. H. Rahman and M. S. Rahman, "An ensemble learning based approach for impression fraud detection in mobile advertising,"Journal of Network Computer Applications, vol.112, Jun. 2018, pp. 126-141, doi: 10.1016/j.jnca.2018.02.021.

[2]. MS. Iqbal, M. Zulkernine, F. Jaafar, and Y. Gu, "Protecting Internet users from becoming victimized attackers of click- fraud," Journal of Software: Evolution and Process, vol. 30, Apr. 2017, doi:10.1002/smr.1871.

[3]. M. Alauthaman, N. Aslam ,Z. Li, R. Alasem, and M. A. Hossain, "A P2P Botnet detection scheme based on decision tree and adaptive multilayer neural networks," Neural Computing and Applications, vol. 29, Jun. 2018, pp. 991-1004, doi: 10.1007 / s00521-016-2564-5.

[4]. D. Ye, and Z. Chen, "A new approach to minimum attribute reduction based on discrete artificial bee colony."Soft Computing, vol. 19, Jul.2015, pp.1893-1903, doi: 10.1007 / s00500-014-1371-0.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

[5]. D. Karaboga, and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm." Journal of Global Optimization, vol. 39, Nov. 2007, pp. 459-471, doi: 10.1007/s10898-007-9149-x.

[6]. Y. Xiong and R. Zuo, "Effects of misclassification costs on mapping mineral prospectivity." Ore Geology Reviews, vol.82, Nov. 2016, pp. 1-9, doi: DOI:10.1016/j.oregeorev.2016.11.014.

[7]. M.A. King, A. S. Abrahams, and C. T. Ragsdale, "Ensemble learning methods for pay-per-click campaign management," Expert Systems with Applications, vol.42, Jun. 2015, doi:10.1016/j.eswa.2015.01.047.