iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

CLICKSENTRY: A HYBRID MACHINE LEARNING AND RULE-BASED SYSTEM FOR CLICK FRAUD DETECTION IN MOBILE APPLICATIONS

Mrs.S.Pavani
Assistant Professor, Department of CSE, JBIET, Hyderabad, India
G.Sanjay Goud, **G.Bhargav**, **M.Nithin**, **K.Sravan**
UG Students, Department of CSE, JBIET, Hyderabad, India

ABSTRACT
Click fraud is a serious problem in the online advertising industry, especially for mobile and web applications that make money through pay-per-click (PPC) models. Fraudulent clicks often come from automated bots, harmful scripts, or fake user actions, causing large financial losses for advertisers and damaging the trustworthiness of online ad platforms. The main detection methods rely on server-side analysis and may not accurately identify widespread and advanced fraudulent activities due to limited insight into real-time user interactions. This paper introduces Click Sentry, a hybrid click fraud detection system that combines machine learning techniques with rule-based URL validation to improve accuracy and reliability in detection. The proposed system uses behavioral analysis by inspecting patterns of user interaction like click frequency along with structural analysis of URLs using keyword-based filtering and regular expression validation The hybrid approach makes sure that malformed suspicious or anomalous URLs are flagged immediately through rule-based logic while machine learning models provide predictive capabilities for identifying complex fraud patterns based on historical data. The system was implemented as a web-based application using the Flask framework which provides an interactive user-friendly interface that supports user registration authentication and real-time fraud detection. An administrator module was also added to allow secure monitoring and management of user activities ensuring controlled access to sensitive data. Role-based access control integration also increases security usability in this system Experimental evaluation shows how well this system can tell apart real clicks from fake ones by using both certain and chance-based detection methods The results show better detection performance than using either method alone with fewer missed detection and greater strength against strange URL patterns and unusual click behavior. A web-based interface was developed using Flask, which allows users to register, login, and do real-time fraud detection. There is also an admin-controlled module for secure monitoring of user activities. The suggested answer gives a flexible and effective setup for dealing with click cheating in today's ad systems. By mixing machine learning with rule-based checks, Click Sentry presents a hands-on and flexible method that could be used in real-world settings. Future upgrades might include using large-scale data, real-time traffic checks, and better deep learning methods to enhance both detection accuracy and system performance.

Keywords:
Health Click fraud, Machine Learning, URL Validation, Cyber Security, Hybrid Detection, Flask Application.

INTRODUCTION

Digital ecosystems are rapidly expanding and changing how businesses advertise their products and services. Online advertising has become the primary medium, with mobile advertising taking center stage because of high smartphone usage and internet connectivity. Among various advertising models, pay-per-click has gained much popularity since advertisers only pay when users interact with ads. However, while this model is economically efficient and performance-driven, it is also very vulnerable to fraudulent activities that take advantage of the system for illegitimate gains. One of the most critical threats in this domain is click fraud which involves generating invalid or malicious clicks on advertisements with an intention to inflate revenue or exhaust advertising budgets.

There are multiple techniques through which click fraud can be carried out such as automated bots, malicious scripts, and deceptive user interactions. Most attackers use distributed networks and proxy servers to simulate genuine user behavior making detection very complicated; some fraudulent activities also come from within applications themselves where embedded malicious code triggers unauthorized clicks without user consent.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

These challenges have made click fraud a persistent issue in digital advertising that leads to huge financial losses as well as reduced trust among stakeholders. With the increasing volume of traffic online, sophistication in fraud mechanisms has equally developed making more advanced detection strategies necessary.

Click fraud detection has traditionally been done using server-side analysis of network traffic and user behavior logs. This mainly involves the identification of abnormal patterns such as high frequency of clicks, repeated interactions from the same device, or unusual geographic distributions. Such techniques can detect certain types of fraud but they often find it difficult to identify complex or distributed attacks. Besides, server-side systems do not have visibility into fine-grained user interactions taking place within client applications which limits their ability to distinguish between genuine and fraudulent To sum up, the rising intricacy of digital ad settings calls for advanced methods in fraud detection clicks.

The Click Sentry system proposed here delivers a balanced and powerful solution through the merging of complementary techniques within a single framework. By dealing with both behavioral and structural parts of user interactions, it offers better detection accuracy and reliability. This work is part of ongoing efforts to improve the safety and transparency of online advertising systems, paving the way for more research and development in this area.

OBJECTIVES

The main goal of this study is to develop and deploy a system that is both efficient and highly effective at detecting fraudulent click behavior in mobile and web applications. This research will attempt to overcome the limitations of current detection systems by combining machine learning and rule-based approaches into one framework. The specific goals for this work are outlined below:

The first goal is to create a strong mechanism for discovering fake click behaviors based on how users act during their sessions. Parameters such as how often users click and when they interact will be used in analysis to separate real activities from potentially fraudulent ones. By adding behavioral analysis, this system aims at identifying anomalies that might not be easily spotted through traditional methods.

The second objective is to develop a rule-based system for validating URL structures. This includes identifying malformed, incomplete, or suspicious URLs using keyword matching and pattern recognition techniques. The purpose of this objective is to ensure that invalid inputs are immediately flagged, thereby improving the overall accuracy and reliability of the detection process

The Third objective is to create a hybrid detection architecture that takes advantage of the strengths inherent in both rule-based and machine learning detection systems. The rule-based component should be able to handle fraud cases that are explicit and easily recognizable, while the machine learning model deals with more complicated scenarios. This hybridization aims at reducing false negatives and improving detection performance.

Additionally, the system intends to provide a user-friendly interface via the web for interaction as well as real-time analysis. Through this interface, users can register, log in, and carry out efficient fraud detection. This objective makes sure that the system becomes accessible and practical for real-world applications. Another objective is an administrator control module that allows secure monitoring and management of user activities. The admin interface provides restricted access to sensitive data so that only authorized personnel can view user information; hence enhancing security as well as integrity within the system. Moreover, the system should be able to conduct real-time detection of fraudulent activities which means that it should immediately flag any suspicious click thereby reducing potential financial losses with improved response time.

METHODOLOGY

The Click Sentry system is based on a hybrid approach that integrates machine learning algorithms with rule-based verification to identify fraudulent click activities. Its overall architecture is designed to enhance detection accuracy through the analysis of user behavior and URL properties. The methodology consists of several stages, each one essential for creating an all-encompassing detection framework.

1) The first stage is data collection and feature engineering. A data set is generated that reflects user interaction patterns through parameters such as click frequency and URL attributes. These features are determined by their significance in differentiating between genuine and fraudulent activities. Click frequency serves as an indicator of abnormal user behavior, while URL attributes are assessed for structural irregularities. The data set is then prepared in a structured format for further analysis.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

2) The next step is data preprocessing and transformation. The gathered data undergoes cleaning to eliminate inconsistencies and ensure standard representation across the board. URL data gets converted into a numerical format using binary encoding, where any signs of suspicious patterns become represented by a flag variable. This conversion allows the machine learning model to effectively handle input data. Normalization techniques are also applied when needed to keep different ranges of data consistent with each other

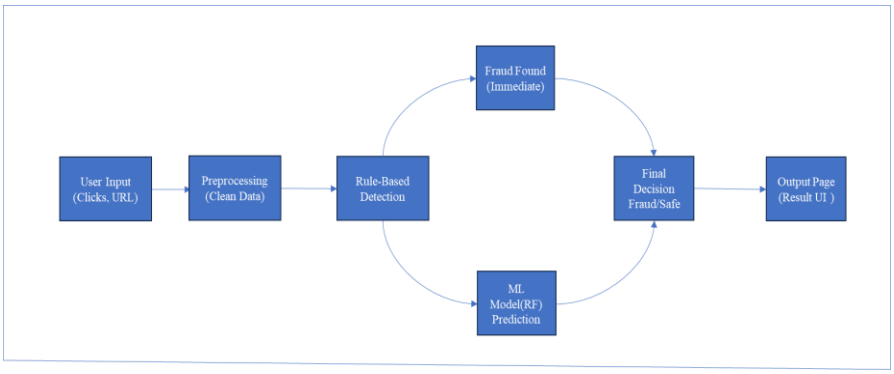

Figure 1 fraud Detection Process

3) The third stage is concerned with the machine learning model. A Random Forest classifier was chosen for its robustness and ability to achieve high accuracy on classification tasks. The model is trained on the processed data set, where input features like click frequency and URL indicators are mapped to output labels that indicate whether a click is fraudulent or legitimate. It learns patterns from the data during training and can then predict outcomes for new inputs. This stage constitutes the predictive part of the system.

4) The fourth stage introduces rule-based validation for URL analysis, which runs outside of the machine learning model and focuses on finding explicit fraud indicators. It checks if there are any suspicious keywords present in the URL such as “fraud”, “hack”, and “phishing”. Regular expression-based validation is also applied here to make sure that the URL follows a standard format; any deviation from what’s expected—for example, missing protocol identifiers or malformed structures—will be treated as a potential fraud indicator. This stage will ensure immediate detection of clearly invalid inputs.

5) The fifth stage combines both components into one hybrid detection mechanism: specifically, it integrates the machine learning model with the rule-based validation system. In this approach, rule-based validation takes precedence over more obvious cases of fraud while leaving room for deeper analysis by the machine learning model in more complicated scenarios. This layered decision-making process increases reliability by not putting all eggs in one basket when it comes to detection methods; therefore, making sure that both explicit and implicit patterns of fraud are captured effectively through this hybrid model.

6) The sixth stage of the model involves system development on a web-based platform. The Flask framework is chosen for developing an interactive front end that provides access to the modules for user registration, authentication, fraud detection, and results display. The integration of backend logic with frontend components allows seamless communication between user inputs and the detection engine. This stage is where the theoretical model transforms into a practical application.

7) The last stage deals with access control and system security. An administrator module is used for monitoring users' activities and managing data in the system. Access to sensitive information like user details will be restricted to authorized personnel through an authentication mechanism. This way, the system can keep its data integrity by preventing unauthorized access. Role-based access control has been included to further enhance the overall security of the application.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Parameter	Description	Type	Role
Click Frequency	Number of clicks per minute	Numerical	Detects abnormal behaviour
URL Input	User-entered web address	Text	Input validation
URL Flag	Indicates presence of suspicious patterns	Binary(0/1)	Rule-based fraud indicator
Keyword Match	Checks for words like fraud, hack, phishing	Boolean	Identifies malicious content
URL Format Check	Validates structure using regular expressions	Boolean	Detect mismatched URL's

Table1Figuers used in ClickSentry

The table above presents a summary of features considered in the click fraud detection system proposed in this study. Behavioral attributes, such as frequency of clicks, help in identifying abnormal user behavior, while URL-based parameters are used for discovering structural inconsistency and patterns indicating potential threats. These features together ensure efficient classification between fraudulent and legitimate clicks within the hybrid detection framework

RESULTS AND DISCUSSION

Click Sentry was implemented as a web-based system for the purposes of evaluating the effectiveness of rule-based and machine learning validation techniques in fraud detection. Performance analyses were conducted based on actual user interactions and test cases. Results indicate that ClickSentry can accurately classify legitimate/fraudulent inputs from users.

The system's first interface is the welcome page, which serves as an entry point for users to interact with it in a simple and intuitive manner. The design ensures easy navigation while introducing the system at hand in a structured way. After this, users are then taken to the home page where they can choose to register or log into their accounts. This step builds up toward user interaction with detection functionality since only authenticated users will be allowed access.

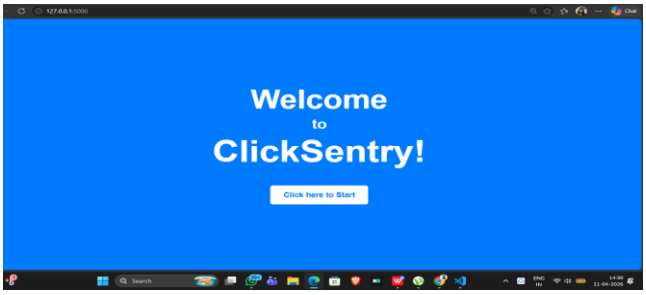

Figure 2Welcome and Home page

The registration module allows users to create an account by providing basic information such as name, email, and password. This information is stored temporarily in the system. The login module authenticates users based on these credentials. The authentication mechanism ensures that only valid users can access the fraud detection module, thus protecting the system and its data.

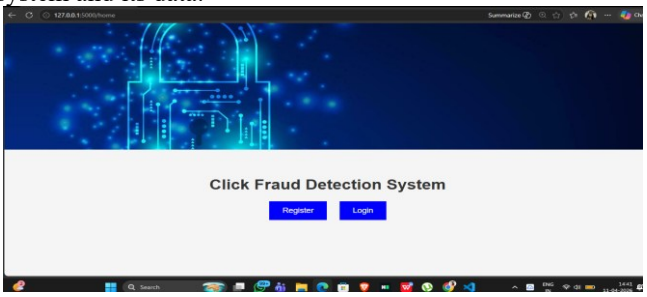

Figure 3Register and Login page

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

After logging in, the user is taken to the fraud detection page, which is the main part of the system. This page takes inputs like how often clicks happen and the URL. The layout of the detection page is set up to give a clear and friendly experience for users, making it easy for them to put in data and start the detection. The joining of backend logic with the frontend makes it easy to process user inputs.

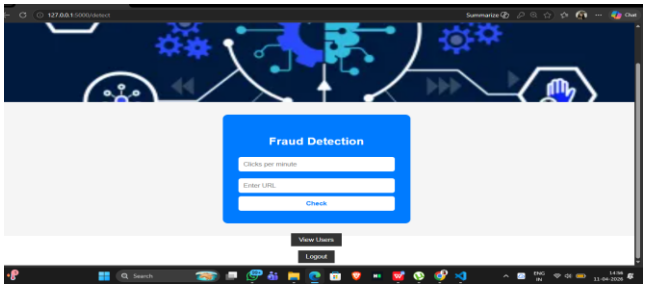

Figure 4Fraud Detection Input Interface

The detection process begins with rule-based validation of the input URL. It checks for suspicious keywords and validates the structure of the URL using pattern matching techniques. If the URL is malformed or contains malicious indicators, it is classified as fraudulent right away. This ensures that obvious fraud cases are handled efficiently without wasting computation resources on unnecessary processes. For inputs that pass rule-based validation, a machine learning model is used to perform further classification by evaluating features such as click frequency and URL indicators to establish whether the activity is legitimate. This combination of approaches enhances overall system accuracy by addressing both explicit and implicit fraud patterns.
Results from the detection process are presented in a separate result page that shows clearly if input has been classified under “Fraud Detected” or “Safe.” The output appears in a very simple and easy-to-understand format so that users can read results quickly and easily understand them. Multiple inputs were tested with this system, including valid URLs as well as those maliciously malformed or suspiciously looking ones, for performance evaluation purposes.

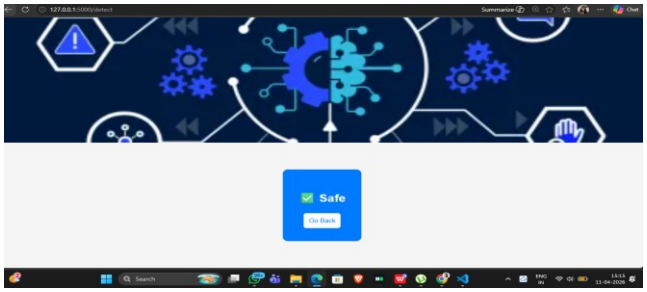

Figure 5 Detection Result

The experimental observations indicate that the system is able to successfully identify fraudulent inputs through rule-based validation. For instance, URLs containing certain keywords or invalid formats were flagged as fraudulent every time. Likewise, abnormal click frequencies were used in classifying some inputs as suspicious. The machine learning model improved detection capabilities further by looking at behavioral patterns not explicitly defined in the rule-based system.
Besides user functionality, an administrator module exists within the system that provides controlled access to user data. A redirection takes place when a user tries to reach the user details page; they are sent to the admin login interface. Only with authorized credentials can one access the user database - thus keeping sensitive information protected. This feature shows how role-based access control has been implemented in this system

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

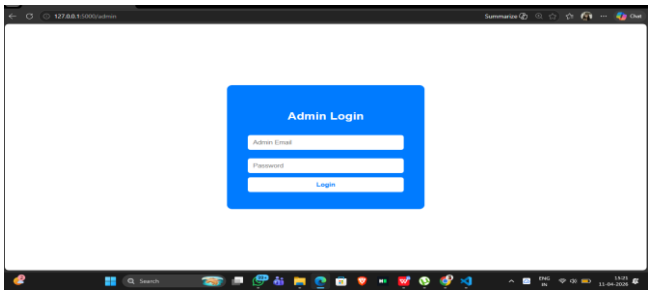

Figure 6 Admin Login

The page shows user details in a table format. The table is visually appealing with proper alignment and formatting to ensure readability. This module helps the administrator track user activity and manage data in the system efficiently. Further, including this feature improves the practical usability of the system and aligns it with real-world application requirements. In summary, results indicated that the ClickSentry system as proposed would be an efficient solution to click fraud detection since it hybridizes structural anomalies and behavioral irregularities; hence minimizing chances of non-detection of fraudulent activities. Its web-based implementation also proved its feasibility for real-world deployment.

ACKNOWLEDGEMENT

We are thankful to the management and faculty of J.B. Institute of Engineering and Technology, Hyderabad, for providing the necessary environment and facilities to complete this work in time. Their continuous support and encouragement were instrumental in the completion of this project. We express our gratitude to our project guide and faculty members for their valuable guidance, suggestions, and constructive criticism during the course of development of this work which helped us greatly in improving its quality as well as direction. We acknowledge with thanks the help received from peers and classmates in discussions, suggestions, and motivation at different stages of the project that contributed toward improving implementation and overcoming challenges effectively. We also take this opportunity to thank all developers who have contributed toward open-source technologies/tools used in developing a machine learning-based application or any web development framework that made it possible to implement successfully an application as proposed here. Last but not least, we would like to extend our thanks to family members for their constant encouragement, patience, and moral support which kept us motivated as well as focused while completing this work.

CONCLUSION

Click fraud is an increasing concern for advertisers and service providers alike due to the growing reliance on digital advertising platforms. Fraudulent clicks result in financial losses and compromise the integrity and effectiveness of online marketing ecosystems. To combat this, there is a need for detection solutions that can accurately identify both blatant and sophisticated fraud patterns.To address these issues, this study presents Click Sentry, a hybrid click fraud detection framework. The proposed solution integrates rule-based validation techniques with machine learning approaches to achieve an optimal balance between accuracy and efficiency in detection. The rule-based component focuses on validating URL patterns and input formats to catch overtly fraudulent instances quickly, while the machine learning model captures more subtle fraudulent behaviors through analysis of user interaction metrics such as click frequency. This hybrid approach ensures comprehensive coverage across different types of fraud scenarios with enhanced overall detection performance.

The system implementation as a web-based application highlights its real-world applicability. The graphical user interface has been kept simple yet effective so that users can carry out necessary tasks related to fraud detection with ease. User registration and login functionalities ensure that access to the system is regulated, while an administrator module adds another security layer by limiting access to sensitive user information. These features make the system not just efficient in terms of detection but also practical for deployment in actual scenarios Agenda goals, a service delivery network and universal health insurance are a must. Strategies include an increased health service coverage in family health, infectious, non- communicable and environmental health care services; Improved Service Delivery Capacity of frontline Health Care Providers; Enhanced Logistics Management System (Supply Chain Management) at the Provincial Health Office and Rural Health Unit to ensure that medicines and other commodities are available all the time; Enhanced timely and reliable Health Information System.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

REFERENCES

[1] Click Fraud in Online Advertising: A Review Oentaryo, R. J., Lim, E. P., Finegold, M., Lo, D., Zhu, F., Phua, C., et al. (2014).
[2] Detecting click fraud in online advertising: A data mining approach. Journal of Machine Learning Research, 15(1), 99-140. Retrieved from https://jmlr.org/papers/v15/oentaryo14a.html
[3] Haddadi, H. (2010). Fighting online click-fraud using bluff ads. ACM SIGCOMM Computer Communication Review 40(2):21-25; available from https://ccr.sigcomm.org/online/files/p22-30v40n2t-haddadiPS.pdf
[4] Alzahrani R A (2022). AI-based techniques for ad click fraud detection and prevention. Future Internet Journal. Available at *https://www.mdpi.com/2224-2708/12/1/4*
[5] Almeida PS Gondim JJ Click fraud detection and prevention system for ad networks Journal of Information Security and Cryptography in 2019.
[6] Dave, V., Guha, S., & Zhang, Y. (2012). Measuring and fingerprinting click-spam in ad networks. Proceedings of ACM SIGCOMM Conference.
[7] Nagaraja, S., & Shah, R. (2019). Clicktok: Click fraud detection using traffic analysis(2015)
[8] Batool A Byun Y C An ensemble deep learning model for click fraud detection in digital advertising.
[9] Phua C Lee V Smith K Gayler R A comprehensive survey of data mining-based fraud detection research.