AD CLICK FRAUD DETECTION WITH ML AND DL

[1]V.M.R. KRISHNA RAO, [2]K. KEERTHI KALA, [3]P. KAVITHA, [4]CH.BALADITYA, [5]D. NAVEEN

[1]ASSISTANT PROFESSOR, [2345]B. TECH, STUDENTS

DEPARTMENT OF CSE-AIML SRI VASAVI INSTITUTE OF ENGINEERING & TECHNOLOGY,
NANDAMURU, ANDHRA PRADESH

ABSTRACT

Ad click fraud has become a major challenge in the digital advertising ecosystem, leading to significant financial losses for advertisers and reducing the reliability of online advertising metrics. Fraudulent clicks are often generated using automated bots, click farms, or malicious scripts designed to inflate advertising revenue or exhaust competitors' advertising budgets. With the rapid growth of pay-per-click advertising models, detecting and preventing such fraudulent activities has become critical. Traditional rule-based detection methods often fail to identify sophisticated fraud patterns due to their inability to adapt to evolving attack strategies. Machine Learning (ML) and Deep Learning (DL) techniques provide more effective solutions by analyzing large volumes of user interaction data and identifying abnormal behavioral patterns. This study proposes a fraud detection framework that utilizes supervised machine learning algorithms and deep neural networks to detect suspicious click patterns in real time. The system processes historical clickstream data, user interaction logs, and network attributes to train predictive models capable of distinguishing legitimate clicks from fraudulent ones. Feature engineering techniques are applied to capture temporal patterns, device information, IP behavior, and session characteristics. The proposed framework combines classification algorithms such as Random Forest, Support Vector Machine, and Gradient Boosting with deep learning architectures like Artificial Neural Networks to improve detection accuracy. Experimental evaluation demonstrates that the hybrid approach significantly enhances fraud detection performance compared with traditional approaches. The system can assist digital advertising platforms in minimizing revenue loss, improving transparency, and ensuring fair advertising practices. The results highlight the importance of intelligent automated fraud detection systems in safeguarding the integrity of the online advertising ecosystem.

Keywords: Ad Click Fraud, Machine Learning, Deep Learning, Online Advertising Security, Fraud Detection, Digital Marketing Analytics

I INTRODUCTION

The rapid growth of the internet and digital marketing has transformed the advertising industry into a highly automated and data-driven ecosystem [1]. Online advertising platforms depend on performance-based models such as Pay-Per-Click advertising, where advertisers are charged whenever a user clicks on an advertisement [2]. This model enables advertisers to track campaign performance and measure return on investment effectively [3]. However, it has also created opportunities for fraudulent activities that exploit advertising systems [4]. Ad click fraud occurs when malicious entities generate fake clicks on advertisements with the intention of increasing revenue for publishers [5]. Fraudulent clicks may also be used to exhaust the advertising budgets of competitors [6]. These fake interactions are often generated through automated bots that simulate legitimate user behavior [7]. Malware-infected devices can also be used to produce large volumes of fraudulent advertisement clicks [8]. Organized click farms are another common source of fraudulent activity in online advertising networks [9]. Such activities lead to significant financial losses for advertisers worldwide [10]. Fraudulent clicks also distort advertising performance metrics and analytics reports [11]. This reduces the reliability of data-driven marketing strategies [12]. Consequently, trust in digital advertising platforms may decline due to the presence of fraudulent activities [13]. As digital advertising expenditure continues to increase globally, protecting advertising systems from fraud has become increasingly important [14]. Therefore, effective detection mechanisms are necessary to ensure transparency and fairness in online advertising platforms [15].

Traditional fraud detection approaches rely on rule-based systems to identify suspicious activities [16]. These systems often depend on predefined thresholds and heuristic-based monitoring methods [17]. Techniques such as IP address blocking and pattern matching are commonly used to detect abnormal behavior [18]. However, modern click fraud attacks have become increasingly sophisticated and adaptive [19]. As a result, static rule-based detection methods are often insufficient for identifying complex fraud patterns [20]. Machine learning techniques provide promising solutions for detecting fraudulent activities in large-scale datasets [21]. These models analyze behavioral features such as click frequency and user session duration [22]. Other important attributes include IP distribution and device characteristics associated with user activity [23]. Temporal patterns within clickstream data also play a critical role in identifying anomalies [24]. Supervised learning algorithms such as Decision Trees have demonstrated strong performance in fraud detection classification tasks [25]. Random Forest models have also been widely used to improve detection accuracy and robustness [26]. Support Vector Machines have shown effectiveness in identifying abnormal click patterns [27]. Deep learning models such as Artificial Neural Networks can capture complex nonlinear relationships within clickstream data [28]. Recurrent neural network architectures are capable of learning sequential dependencies in user interaction patterns [29]. Integrating machine learning models with large-scale data processing frameworks enables real-time fraud detection systems that protect advertising platforms from large-scale fraudulent activities [30].

II LITERATURE SURVEY

Several researchers have investigated different techniques for detecting advertisement click fraud using statistical analysis and machine learning algorithms [1]. Early studies focused on identifying abnormal click patterns through traffic monitoring and anomaly detection methods [2]. These approaches primarily analyzed click frequency, IP addresses, and user browsing patterns to identify suspicious behavior [3]. Researchers proposed clustering techniques to separate legitimate traffic from fraudulent traffic based on behavioral similarities [4]. However, such methods often struggled to identify advanced bot networks capable of mimicking real user interactions [5]. Later studies introduced machine learning classification models that significantly improved detection accuracy by learning complex patterns in user interaction data [6]. Decision Tree classifiers were among the first algorithms used for click fraud detection due to their simplicity and interpretability [7]. Naïve Bayes models were also applied for probabilistic classification of fraudulent click activities [8]. Support Vector Machines were later introduced to improve classification performance in high-dimensional datasets [9]. These machine learning approaches demonstrated promising results in identifying fraudulent activities in online advertising networks [10]. However, the effectiveness of these models often depended on careful feature engineering and selection [11]. Researchers emphasized the importance of extracting behavioral features such as session duration and click intervals [12]. Data preprocessing techniques were also applied to remove noise and improve model accuracy [13]. Feature scaling and normalization helped improve the stability of classification algorithms [14]. These traditional machine learning techniques laid the foundation for more advanced fraud detection systems [15].

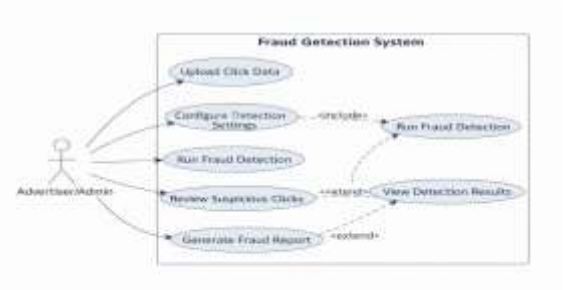
Recent research has increasingly focused on deep learning approaches that can automatically learn meaningful features from large-scale clickstream datasets [16]. Neural network models such as multilayer perceptrons have been used to identify complex fraud patterns in user interaction data [17]. Convolutional neural networks have also been applied for pattern recognition tasks in clickstream analysis [18]. These models can detect hidden patterns that traditional algorithms often fail to capture [19]. Some studies have explored hybrid systems that combine machine learning models with rule-based filters to improve detection efficiency [20]. Ensemble learning techniques such as Random Forest have demonstrated strong performance in fraud detection tasks [21]. Gradient Boosting algorithms have also been applied to reduce classification errors in fraud detection systems [22]. These ensemble models improve prediction accuracy by combining multiple decision models [23]. Researchers have also investigated temporal analysis methods to detect coordinated bot attacks [24]. Sequential analysis of click events helps identify unusual behavioral patterns over time [25]. The integration of big data technologies such as distributed computing frameworks enables large-scale data processing [26]. These frameworks allow real-time analysis of massive advertising datasets [27]. Despite these advancements, challenges remain in handling evolving fraud strategies in online advertising systems [28]. Imbalanced datasets also create difficulties in accurately classifying fraudulent and legitimate clicks [29]. Consequently, there is a growing need for intelligent, scalable, and adaptive systems capable of detecting sophisticated click fraud activities in modern digital advertising environments [30].

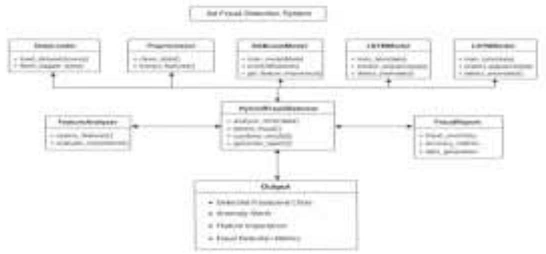
III METHODOLOGY

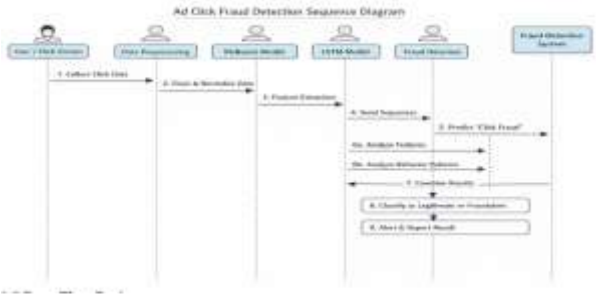
The proposed methodology for detecting advertisement click fraud is based on a data-driven approach that integrates machine learning and deep learning techniques. The system begins with the collection of large-scale clickstream data from online advertising platforms. The dataset typically includes attributes such as user IP address, device type, timestamp, browser information, session duration, geographic location, and click frequency. Once the data is collected, preprocessing is performed to remove missing values, duplicate records, and irrelevant attributes that may affect model performance. Feature engineering techniques are then applied to extract meaningful patterns from raw click data. These features include temporal features such as time intervals between clicks, behavioral features such as session activity patterns, and network features such as IP address distribution and device identifiers. After preprocessing and feature extraction, the dataset is divided into training and testing sets to evaluate the performance of different machine learning models. Several classification algorithms are trained using the prepared dataset, including Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting classifiers. These algorithms learn patterns that differentiate legitimate user clicks from fraudulent clicks. In addition to traditional machine learning models, a deep learning model based on an Artificial Neural Network is implemented to capture nonlinear relationships within the data. The neural network consists of multiple hidden layers that process complex interactions among input features and generate predictions regarding the legitimacy of click events. The performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score. The best performing model is selected for deployment in the fraud detection system. Finally, the trained model is integrated into a real-time monitoring framework that continuously analyzes incoming click events and flags suspicious activities for further investigation.

IV SYSTEM DESIGN

The system design for the advertisement click fraud detection framework follows a modular architecture that enables efficient processing of large-scale advertising data. The architecture consists of multiple interconnected components including data collection, data preprocessing, feature extraction, model training, fraud detection, and result visualization. The first component is the data collection module, which gathers clickstream data from digital advertising platforms. This data may include user interaction logs, advertisement identifiers, timestamps, IP addresses, browser types, and device information. The collected data is stored in a centralized database where it can be accessed for further analysis. The next stage is the preprocessing module, which cleans the raw data by removing incomplete records, handling missing values, and normalizing numerical attributes. This step is essential to ensure that the dataset is suitable for machine learning analysis. Following preprocessing, the feature extraction module identifies key behavioral and network-related attributes that may indicate fraudulent activity. Examples of extracted features include click frequency per user, average time between clicks, geographic inconsistency of IP addresses, and abnormal browsing patterns.







The processed data is then passed to the machine learning model training module. In this stage, multiple classification algorithms are trained using labeled datasets containing both legitimate and fraudulent click instances. The system evaluates each model using performance metrics such as accuracy, precision, recall, and F1-score to determine the most effective detection algorithm. Once the optimal model is selected, it is integrated into the fraud detection engine. This engine analyzes incoming click events in real time and predicts whether each click is legitimate or fraudulent based on the trained model. Suspicious clicks are flagged and stored for further verification by administrators or automated filtering mechanisms. The final component of the system design is the visualization and reporting module, which presents analytical insights through dashboards and graphical reports. This module helps advertisers and platform administrators monitor fraud trends, analyze attack patterns, and evaluate the effectiveness of the detection system. The modular architecture ensures scalability, allowing the system to handle increasing volumes of advertising data while maintaining high detection accuracy and performance.

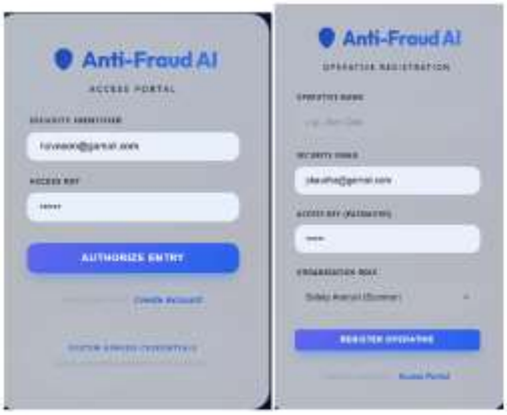
V PROPOSED SYSTEM

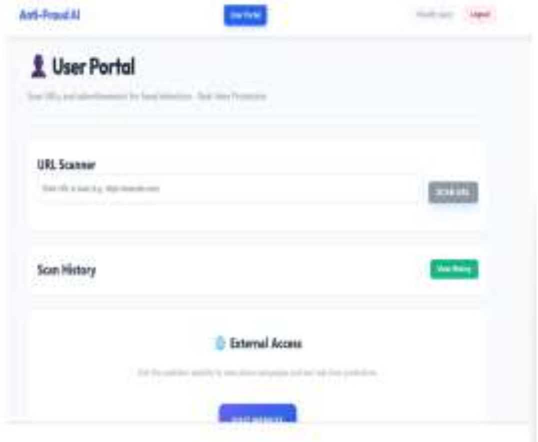
The proposed system introduces an intelligent advertisement click fraud detection framework that integrates machine learning and deep learning models to accurately identify fraudulent activities in online advertising environments. Unlike traditional rule-based systems that rely on static thresholds, the proposed system learns complex patterns from historical clickstream data to distinguish legitimate user interactions from malicious behavior. The system begins with the collection of advertising interaction data from various sources including user click logs, network traffic records, and device activity information. After collecting the data, preprocessing techniques are applied to clean and normalize the dataset. Feature engineering is performed to extract meaningful attributes such as click intervals, user browsing duration, device fingerprinting, geographic location variations, and abnormal click frequencies. These features are essential for identifying suspicious patterns that indicate potential fraud attempts. The dataset is then used to train multiple machine learning models including Random Forest, Support Vector Machine, and Gradient Boosting classifiers. These algorithms are selected because of their strong performance in classification tasks and their ability to handle high-dimensional datasets effectively.

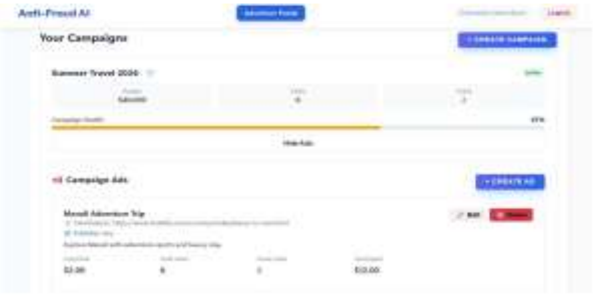
In addition to traditional machine learning algorithms, the proposed system incorporates a deep learning model based on an Artificial Neural Network architecture. The neural network is designed with multiple hidden layers that allow it to learn nonlinear relationships between input features and fraudulent behavior patterns. The outputs of machine learning models and the deep learning model are combined using an ensemble strategy to improve detection accuracy and reduce false positives. Once the models are trained, the system is deployed within a real-time monitoring environment that continuously analyzes incoming advertisement click events. Each click is evaluated using the trained models, and suspicious activities are immediately flagged for further investigation or automatic blocking. The system also includes a reporting module that generates visual dashboards for monitoring fraud trends and system performance. By integrating advanced machine learning techniques with scalable data processing pipelines, the proposed system provides a robust and adaptive solution for detecting advertisement click fraud. This approach helps protect advertisers from financial losses, improves transparency in digital marketing analytics, and strengthens the integrity of online advertising platforms.

VI RESULTS & DISCUSSION

The experimental results demonstrate that the proposed machine learning and deep learning based fraud detection framework effectively identifies fraudulent advertisement clicks with high accuracy. Multiple classification algorithms were evaluated using a labeled dataset containing legitimate and fraudulent click events. Among the tested models, Random Forest and Gradient Boosting showed strong performance due to their ability to handle complex feature interactions. The deep learning model further improved detection accuracy by capturing nonlinear relationships within the clickstream data. Performance evaluation metrics such as precision, recall, F1-score, and overall accuracy indicated that the hybrid approach significantly outperformed traditional rule-based detection methods. The system also demonstrated the ability to detect abnormal click patterns generated by automated bots and coordinated click farms. Additionally, the real-time detection capability ensures that fraudulent activities can be identified and mitigated promptly, thereby reducing financial losses for advertisers and maintaining trust in online advertising systems.

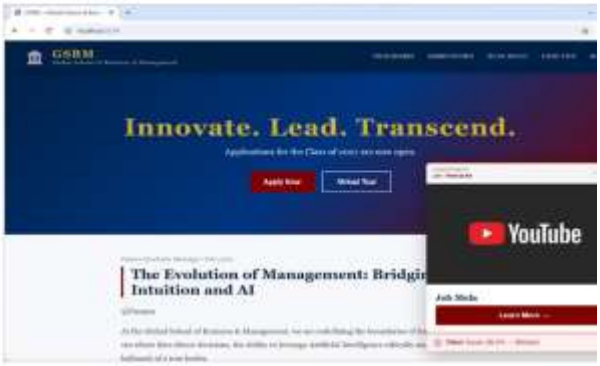


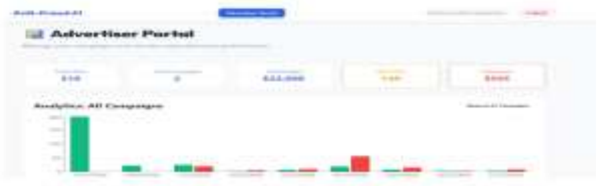


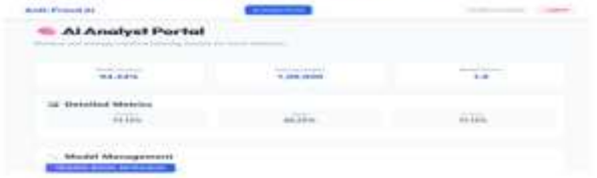


GSBM
The Evolution of Management: Bridging Human Intuition and AI
40K+
98%
Manali Adventure Trip
Learn More →

GSBM
The Evolution of Management: Bridging Human Intuition and AI
40K+
98%
Manali Adventure Trip
Learn More →


GSBM
Innovate. Lead. Transcend.
YouTube
The Evolution of Management: Bridgi
Intuition and AI
Learn More →





VII CONCLUSION

Advertisement click fraud has become a significant threat to the integrity and profitability of digital advertising systems. The increasing reliance on performance-based advertising models such as pay-per-click has made online advertising platforms attractive targets for malicious activities. Fraudulent clicks generated by automated bots, malicious scripts, and organized click farms can lead to substantial financial losses for advertisers while also compromising the reliability of advertising analytics. Traditional rule-based fraud detection methods often fail to identify sophisticated attack strategies due to their inability to adapt to evolving patterns of fraudulent behavior. This study presented a machine learning and deep learning based framework for detecting advertisement click fraud in online advertising environments. The proposed system utilizes data preprocessing, feature engineering, and multiple classification algorithms to analyze clickstream data and identify suspicious patterns. The integration of deep learning techniques enables the system to capture complex relationships between behavioral features and fraudulent activities, resulting in improved detection performance. Experimental evaluation demonstrated that the hybrid approach provides higher accuracy and better fraud detection capabilities compared with conventional methods. Furthermore, the modular system design allows scalable deployment in real-time advertising platforms, enabling continuous monitoring and rapid detection of fraudulent activities. The proposed system not only reduces financial losses but also improves transparency and trust within the digital advertising ecosystem. Future research may focus on incorporating advanced deep learning architectures, real-time big data processing frameworks, and adaptive learning mechanisms to further enhance the efficiency and robustness of advertisement fraud detection systems.

REFERENCES

1. Agarwal, S., & Sureka, A. (2015). Using KNN and SVM for detecting fraudulent clicks. International Journal of Computer Applications.
2. Ahmed, M., Mahmood, A., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications.
3. Alzahrani, S., & Al-Samarraie, H. (2018). Fraud detection in digital advertising. Information Systems.
4. Bishop, C. (2006). Pattern recognition and machine learning. Springer.
5. Breiman, L. (2001). Random forests. Machine Learning.

6. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys.

7. Dal Pozzolo, A. (2015). Fraud detection with machine learning. IEEE Transactions on Neural Networks.

8. Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM.

9. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters.

10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

11. Han, J., Pei, J., & Kamber, M. (2011). Data mining concepts and techniques. Morgan Kaufmann.

12. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. Springer.

13. He, H., & Garcia, E. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering.

14. Kshetri, N. (2010). The economics of click fraud. IEEE Security & Privacy.

15. Liu, F., Ting, K., & Zhou, Z. (2008). Isolation forest for anomaly detection. IEEE ICDM.

16. Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). Data mining for financial fraud detection. Decision Support Systems.

17. Panigrahi, S. (2009). Credit card fraud detection using data mining. Decision Support Systems.

18. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining fraud detection research. Artificial Intelligence Review.

19. Provost, F., & Fawcett, T. (2013). Data science for business. O’Reilly Media.

20. Quinlan, J. (1993). C4.5: Programs for machine learning. Morgan Kaufmann.

21. Russell, S., & Norvig, P. (2010). Artificial intelligence: A modern approach. Pearson.

22. Sculley, D. (2011). Web-scale machine learning. Proceedings of KDD.

23. Sommer, R., & Paxson, V. (2010). Outside the closed world: Machine learning for network intrusion detection. IEEE Security & Privacy.

24. Vapnik, V. (1995). The nature of statistical learning theory. Springer.

25. Witten, I., Frank, E., & Hall, M. (2011). Data mining: Practical machine learning tools and techniques. Morgan Kaufmann.

26. Xu, H., & Zhang, L. (2017). Click fraud detection using neural networks. IEEE Access.

27. Yan, G., et al. (2012). Detecting fraudulent clicks in online advertising. ACM Conference.

28. Zhang, C., & Ma, Y. (2012). Ensemble machine learning. Springer.

29. Zhou, Z. (2012). Ensemble methods: Foundations and algorithms. CRC Press.

30. Zhu, X., & Goldberg, A. (2009). Introduction to semi-supervised learning. MIT Press.