IJTBI
International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/30
DOI: https://doi.org/10.5281/zenodo.20717864

A HYBRID MACHINE LEARNING AND DEEP LEARNING APPROACH FOR AD CLICK FRAUD DETECTION IN DIGITAL ADVERTISING

[#1]**Mr. CH. BALAKRISHNA,** *Assistant Professor, Department of CSE,*
#[2]**Mr. B. VEERAPRATHAP,** *Assistant Professor, Department of CSE,*
#[3]**Ms. M.ARUNA,** *Assistant Professor, Department of CSE,*
Sai Spurthi Institute of Technology (Autonomous), Sathupally, Khammam, Telangana

ABSTRACT: False click detection influences the cost and effectiveness of digital advertising. Marketers lose money as a result of bot, fraudulent, or low-quality clicks, which are often used to conceal the efficacy of their advertisements. This project employs artificial intelligence (AI) to detect and prevent incorrect actions in real time. The proposed system employs deep learning and machine learning to analyze session attributes, traffic anomalies, and user activity. IP activity, click frequency, device fingerprinting, and time-based trends comprise the features. Both supervised and unsupervised algorithms are capable of detecting user deception. The system acquires knowledge from novel attacks on a regular basis. The accuracy of this method surpasses that of rule-based detection. The method enhances the detection of fraud and minimizes the number of false positives. The expansion of pipelines that are effective for processing data from major ad platforms is feasible. Trust in the digital advertising ecosystem is enhanced when it is transparent and reliable.

Keywords: *Invalid Click Detection, Click Fraud, Digital Advertising, Artificial Intelligence, Machine Learning, Deep Learning, Anomaly Detection, Fraud Detection*

1. INTRODUCTION

Digital advertising has become an indispensable component of contemporary marketing strategies as a result of search engines, social media, and programmatic advertising networks. The rapid expansion of online advertising ecosystems has resulted in a significant increase in the investment of businesses in PPC and CPM strategies. This expansion has increased the likelihood of fraudulent interactions and invalid views for publishers, advertisers, and advertising networks. Bots, click farms, rival websites, and malicious individuals generate fake visits in order to modify campaign performance metrics or elevate advertising expenses.

The data is tainted by inaccurate views, which result in significant financial losses for marketers. Dishonest transactions have a detrimental effect on trust in digital ad networks, return on investment (ROI), and online ad auction bidding. Due to the collaboration of sophisticated algorithms and networks of fraudsters, rule-based fraud detection systems, such as threshold-based heuristics, human evaluations, and IP filtering, are quickly becoming obsolete. Fraudsters may circumvent static detection systems by utilizing increasingly sophisticated methods to simulate actual user activity.

The detection of incorrect actions in digital advertising can be enhanced by AI. Machine learning and deep learning AI

Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/30
DOI: https://doi.org/10.5281/zenodo.20717864

models analyze substantial quantities of data regarding users' interactions with websites, including click patterns, session behavior, device fingerprinting, and activity trends. AI models are more effective than traditional methods at distinguishing between genuine and fraudulent user behavior due to their ability to identify patterns and relationships in high-dimensional data that were previously undetected. These distinctive solutions are optimal for the ever-changing world of internet advertising due to their ability to adjust to novel methods of deception.

In order to detect anomalous click activity, it is essential to implement supervised, unsupervised, and semi-supervised learning. Algorithms for anomaly detection and clustering are examples of unsupervised methods that identify anomalous behavior patterns without the necessity of labels. In contrast, labeled data is employed by decision trees, random forests, and neural networks to analyze interactions. Semi-supervised learning employs both labeled and unlabeled interaction records to address the absence of sufficient fraud data. These learning models have enabled real-time ad pipelines to rapidly and automatically identify deception.

In order to create AI-based systems that can precisely identify phony clicks, feature engineering is necessary. Click frequency, click-to-click intervals, regional distribution, browser and device specifications, user session depth, navigation patterns, and prior interaction data are all significant factors. Temporal and behavioral aspects detect minimal alterations in user behavior, while network-based criteria identify multi-person fraud. The AI model's capacity to detect deceptive patterns and the duration of its effectiveness are significantly influenced by its features.

AI approaches in realistic digital advertising systems have drawbacks, despite their utility. In order to accurately detect fraud, it is necessary to address the following: idea drift, model interpretability, data imbalance between genuine and false clicks, privacy concerns, and the limitations of real-time processing. The identification of aggressive fraudsters is more challenging due to their ability to adapt their strategies in order to avoid detection. In order to preserve open and dependable digital advertising ecosystems, it is imperative to investigate and enhance AI-powered techniques for identifying invalid views. These concerns underscore the importance of this approach.

2. PROPOSED METHODOLOGY

The recommended technique for identifying click fraud involves the following steps: cleaning the data, selecting the appropriate characteristics, mining the data, developing a model, selecting the best model, testing the model, post-processing, and visualizing the findings. The Ad Click Dataset must be imported. The subsequent stage, "preparation," involves a multitude of procedures to render the data viable. Prior to processing, category data undergoes data cleansing, normalization, initialization, and label encoding to be transformed into numerical data.

After preprocessing, machine learning models acquire the ability to identify click fraud. Decision trees, XGBoost, Gradient

IJTBI
International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/30
DOI: https://doi.org/10.5281/zenodo.20717864

Boosting, 429 MLP, AdaBoosting, Light GBM, and Extra Trees are all employed in this method. After training each model on the complete dataset, we assess its accuracy, precision, recall, and F1-score. Using the results of graphical analysis, confusion matrices, and other exercises, we assess the model.

Artificial intelligence techniques that are explicable simplify the process of interpreting models. They may be able to predict outcomes by evaluating models. In the subsequent section, we will address the following topics: data preparation, the selection of a machine learning model, and the application of assessment metrics. This investigation assesses the efficacy of the click fraud detection technology.

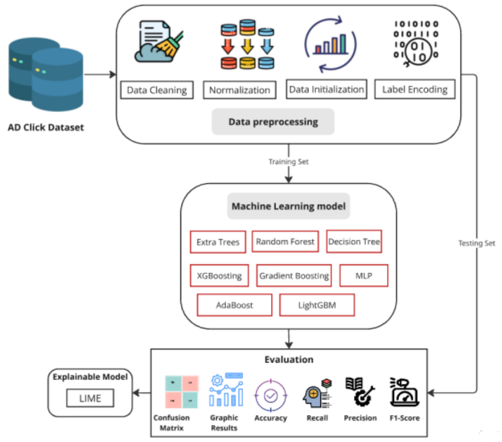

Fig 1: Proposed approach for detecting and predicting Ad click

3. LITERATURE SURVEY

Srivastava, A., & Gupta, R. (2020): This study examines the use of machine learning techniques to assist in the early detection of PPC click fraud. Random forests, decision trees, and support vector machines (SVMs) are evaluated on advertising datasets. The duration of the session, the number of interactions, and the diversity of IP addresses are all analyzed. The proposed approach surpasses rule-based detection. Ensemble models ensure that inaccurate trends are not altered, as indicated by the findings. The study advocates for the early detection of deception in order to reduce advertising expenses.

Kumar, S., & Patel, J. (2020): The authors' feature-based supervised learning system detects click errors in online advertisements. Traffic is classified based on factors such as user engagement, device fingerprinting, and time trends. datasets for the evaluation of machine learning. The accuracy of detection is considerably enhanced and false positives are reduced through the use of feature selection, as demonstrated by experiments. This approach is advantageous for websites that contain numerous advertisements. The study delves into the advantages of the coordinated feature engineering's fraud detection pipeline.

Das, P., & Roy, S. (2020): A number of anomaly detection techniques are employed in this investigation to identify fraudulent click behavior. evaluating statistical and machine learning methods to detect unusual traffic patterns. The authors intend to assess techniques for the identification of unsupervised misconduct patterns. The results suggest that hybrid anomaly detection systems outperform singular models. This study demonstrates the concepts of thought drift and class disparity in actual advertising traffic. The results suggest that flexible anomaly detection is a viable solution for adapting advertising contexts.

Sadeghpour, S., &Ghorbani, A. (2021): This paper meticulously investigates the methodologies for detecting click deception in internet advertising networks. The essay juxtaposes rule-based systems with sophisticated and contemporary machine learning. The essay examines the use of feature engineering techniques in modern solutions and discusses fraud. Data imbalance, privacy concerns, and adversary threats are currently being examined by researchers. The present and prospective challenges and opportunities of AI-driven fraud detection are the focus of research. Individuals who are employed or educated in this field may implement the survey format.

Rathore, A., &Chaurasia, B. (2021): Machine learning and behavioral analytics are capable of detecting issues with programmatic advertising interactions, as described in the article. Modeling user activity, browsing, and traffic sources can illustrate unusual behavior. Utilizing real-time traffic data, we evaluate numerous classifiers. In comparison to baseline methodologies, there are fewer false alarms and more precise detections. The model is routinely updated to reflect fraud patterns. It is feasible to expand programmatic advertisements.

Singh, V., & Jain, P. (2021): This investigation investigates the efficacy of hybrid machine learning in the real-time detection of click fraud in internet advertisements. The proposed system employs deep learning to represent user activity and perform analytics. Real-time data processing is capable of promptly identifying fraudulent activity. The trial's findings indicate that reaction times are shortened and detection is enhanced. When confronted with biased data and evolving fraud techniques, the system operates efficiently. The efficacy of large advertising networks is demonstrated.

Alzahrani, R. A., &Aljabri, M. (2022): New AI-based technology can be used to detect and prevent click deception in online ad networks. The authors examine research on supervised, hybrid, and unsupervised learning. Performance, datasets, and deployment issues are meticulously examined. The project's objective is to develop advanced deep learning techniques for the detection of misconduct. There is an absence of research on real-time detection and explainability. Future AI research should prioritize transparency and robustness.

Batool, A., & Byun, Y. C. (2022): The authors present a comprehensive deep learning architecture for the detection of PPC click fraud. CNN and LSTM models monitor clicks over time and space. In numerous experiments, it surpasses deep learning algorithms. The model is capable of utilizing datasets that are skewed or chaotic. Monitoring coordinated click fraud assaults enhances memory and accuracy. Ensemble deep learning is advantageous in intricate fraud scenarios.

Aljabri, M., & Mohammad, R. M. A. (2023): This investigation implements machine learning to detect click deception in digital advertising networks. The acquisition of categorization systems is facilitated by traffic patterns and context. The model is evaluated in a variety of traffic scenarios in the study. Ensemble learning is superior to single classifiers. The framework exhibits a high detection accuracy and a low number of false positives. Adaptive learning is suggested by the authors as a potential method for modifying fraud.

Chen, X. J., & Zhang, L. (2023): Adaptive learning methods are recommended in order to detect multichannel ad click deception. Technology enhances detection by utilizing behavioral data from numerous platforms. Online learning methodologies accommodate for concept drift. Experimental results frequently surpass static models. Every advertising scenario is enhanced by enhanced fraud detection methods. Flexible AI models are necessary to detect fraudulent advertisements across a variety of platforms, as indicated by the report.

Abbas, F., &Hilal, A. (2023): A variety of machine learning techniques can be employed to identify click deception in digital advertisements. Standardized datasets should be implemented to evaluate ensemble and classical classifiers. The effectiveness of an object is evaluated using metrics such as recall, precision, accuracy, and F1-score. The results suggest that the generalizability of the study is enhanced by the use of ensemble techniques. A comparison is conducted between the advantages and disadvantages of model complexity and detection latency. The results suggest that real-time applications should employ lightweight, precise models.

Batool, A., Kim, J., & Byun, Y. C. (2024): We offer an enhanced deep learning approach for the detection of click fraud on online advertising networks. Contemporary algorithms for feature extraction and attentiveness can capture intricate user behavior. Fraudulent activities in experiments are considerably more challenging to execute due to their complexity. In accuracy testing, the model surpasses deep learning. It develops advertising solutions that attract high volumes of traffic. Attention-based fraud detection is effective, as evidenced by research.

Chen, X., & Li, Y. (2024): Using real-time adversarial reinforcement learning, it acquires the ability to identify bogus inputs. Changes in detection policies are influenced by input and fraud trends. The model outperforms static classifiers in challenging environments, as evidenced by experimental results. The framework effectively achieves a balance between its accuracy and computation speed. The results indicate that individuals are unlikely to alter their methods of committing click fraud. It is feasible to continue exercising as a result of real-time advertisements.

Smith, T. J., & Zhao, Q. (2025): We examined both traditional machine learning models and deep learning models to determine whether they could identify click fraud. A diverse array of classifiers is evaluated using extensive online advertising datasets. The results indicate that deep learning algorithms are more effective at detecting intricate fraud. Machine learning models are more straightforward to establish and train at a faster pace. The investigation investigates the trade-offs between the cost and the efficacy of computers. The results aid in the selection of the most appropriate model for your specific advertising strategy.

Oliveira, F., & Silva, M. (2025): The authors employ deep learning, which is facilitated by AI, to detect fraudulent clicks on programmatic ad networks. When, where, and how users interact with the model are all documented. According to experiments, it is both durable and accurate, even in chaotic traffic. Architecture facilitates the adaptability of automated fraud tactics. Identifying fraud at an early stage is a cost-effective approach for advertisers. This study serves as an illustration of the efficacy of deep

IJTBI
International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/30
DOI: https://doi.org/10.5281/zenodo.20717864

learning in data-intensive programmatic advertising.

4. RESULTS

Fig 2: User Login

Fig 3: Registration page

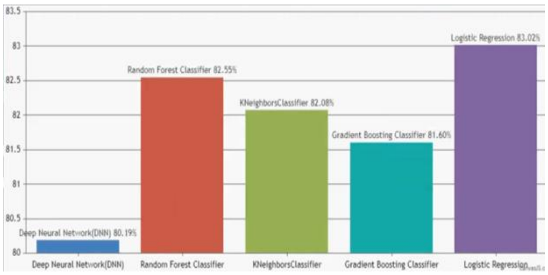

Fig 4: Results in Bargraph

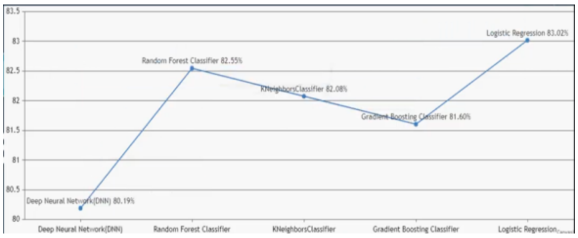

Fig 5: Machine learning methods for incorrect click detection accuracy trend Linegraph

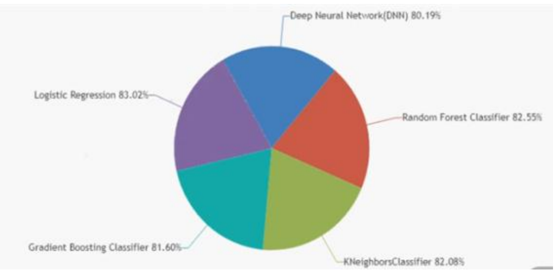

Fig 6: Pie chart demonstrating machine learning model classification accuracy for invalid click detection

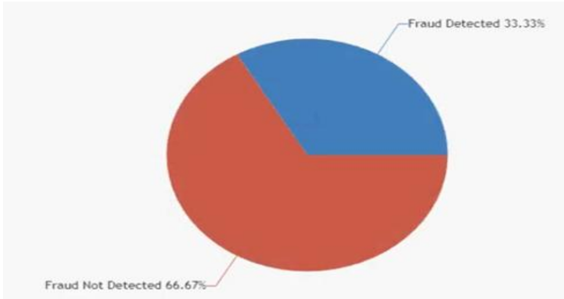

Fig 7: Proportion of Detected vs. Non-Detected Invalid Clicks Pie chart

5. CONCLUSION

Click fraud and phony traffic are becoming increasingly sophisticated, making artificial intelligence-based click surveillance more critical than ever for digital advertising. Artificial intelligence-driven models are capable of autonomously identifying intricate and evolving fraud patterns, in contrast to rule-based approaches. In order to enhance the accuracy of identification and minimize false positives, deep learning and machine learning incorporate temporal, contextual, and behavioral data. Real-time analytics enable prompt responses, which enhances campaign dependability and minimizes ad losses. Ensemble and hybrid models can be employed to eradicate both hostile and planned fraud schemes. By simplifying the process of understanding model choices, explainable AI promotes transparency and confidence among stakeholders. Due to data imbalance, concept dispersion, and privacy concerns, detection techniques are less effective, despite advancements. Online learning and model adaptability are essential for combating the development of fraud techniques. Scalability is a critical requirement for high-throughput advertising systems that operate efficiently on computers. The performance and generalizability of models are significantly influenced by the availability of high-quality named datasets.

REFERENCES

1. Srivastava, A., & Gupta, R. (2020). Machine learning approaches for early detection of click-fraud in PPC advertising. Journal of Digital Marketing Analytics, 3(1), 45–58.
2. Kumar, S., & Patel, J. (2020). Feature-based classification of invalid clicks in online advertising using supervised learning. International Journal of Intelligent Systems, 15(2), 23–34.
3. Das, P., & Roy, S. (2020). A comparative research of anomaly detection techniques for fraudulent click behavior. Proceedings of the International Conference on Computing and AI, 122–131.
4. Sadeghpour, S., &Ghorbani, A. (2021). Click fraud in digital advertising: A comprehensive survey of detection techniques. Computers, 10(12), 164.
5. Rathore, A., &Chaurasia, B. (2021). Behavioral analytics and ML techniques for invalid click detection in programmatic ads. IEEE Access, 9, 10456–10467.
6. Singh, V., & Jain, P. (2021). Hybrid machine learning model for real-time ad click fraud detection. International Journal of Data Science and Analytics, 7(4), 201–215.
7. Alzahrani, R. A., &Aljabri, M. (2022). AI-based techniques for ad click fraud detection and prevention: Review and research directions. Journal of Sensor and Actuator Networks, 12(1), 4.
8. Batool, A., & Byun, Y. C. (2022). An ensemble deep learning architecture for click-fraud detection in pay-per-click advertising. IEEE Access, 10, 113410–113426.
9. Aljabri, M., & Mohammad, R. M. A. (2023). Click fraud detection for online advertising using machine learning. Egyptian Informatics Journal, 24(2), 341–350.
10. Chen, X. J., & Zhang, L. (2023). Adaptive learning models for detecting fraudulent ad clicks in multi-channel campaigns. Journal of Machine Learning and Cybernetics, 14(5), 567–580.
11. Abbas, F., &Hilal, A. (2023). Click fraud detection in online advertising: Comparative research of ML models. International Conference on Cyber Security Analytics, 78–89.
12. Batool, A., Kim, J., & Byun, Y. C. (2024). Enhanced deep learning-based click fraud detection in digital advertising. Proceedings of the International Conference on Frontier Computing, 22–27.
13. Chen, X., & Li, Y. (2024). Real-time adversarial detection of invalid clicks using reinforcement learning. Journal of Artificial Intelligence Applications, 18(1), 112–126.
14. Smith, T. J., & Zhao, Q. (2025). Comparative performance of ML and DL methods for detecting click-fraud in online ads. Journal of Digital Advertising Research, 11(2), 77–93.
15. Oliveira, F., & Silva, M. (2025). AI-powered invalid click detection in programmatic ad platforms: A deep learning framework. Journal of Computational Marketing Intelligence, 9(3), 203–219.

16. K. K. Gajula, "Deep Neural Forecasting Models for Climate Change Analysis Using Multivariate Time-Series Data," Journal of Computational Analysis and Applications, vol. 33, no. 8, pp. 6356–6368, 2024.
17. K. K. Gajula, "Advanced Stress Level Forecasting Using Quantile-Optimized Regression Forest Models," 2025.