ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/150-157
Syed Abdul Rahman *et. al.*, / International Journal of Engineering & Science Research

Full Length Article

An Enhanced RNN-LSTM Model For Accurate and Real-Time Click Fraud Detection in Online

Syed Abdul Rahman[1], Mohammed Umair[2], Syed Muzammil Quadri[3], Mrs. T.Anitha[4]
[1,2,3]B.E.Students; Department of Computer Science and Engineering, ISL Engineering College, Hyderabad, India.
[4]Assistant Professor; Department of Computer Science and Engineering, ISL Engineering College, Hyderabad, India.
Mail Id;syedabdulrahman0427@gmail.com,umairkaldi060@gmail.com,quadrisyedmuzammil39@gmail.com, anithat.nov26@gmail.com
Accepted 23-04-2026

ABSTRACT:
Click fraud continues to pose a significant threat to online advertising by inflating costs and diverting budgets toward illegitimate activities, necessitating more sophisticated detection methods that go beyond traditional machine learning capabilities. In this study, we propose a robust LSTM-based Recurrent Neural Network (RNN) framework specifically designed to capture subtle behavioral patterns and time-dependent features inherent in user interaction data. By implementing a comprehensive preprocessing pipeline—comprising timestamp decomposition, feature scaling, and label encoding—we ensured high-quality input representation for our model. When benchmarked against Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), our RNN-LSTM architecture demonstrated superior performance, achieving 99% accuracy alongside exceptional precision and recall scores. These results confirm that temporal modeling is highly effective for identifying fraudulent click sequences, positioning the LSTM model as a powerful solution for real-time ad verification and establishing a new benchmark for future advancements in intelligent fraud prevention.

Keywords: *Click Fraud Detection, LSTM (Long-Short Term Memory), RNN-Recurrent Neural Network, Sequential patterns, Temporal Modelling, Deep Learning, Real-Time Detection.*

INTRODUCTION

In the rapidly evolving landscape of digital transformation, online advertising has become a vital channel for global audience engagement, yet it faces a severe threat from click fraud—a deceptive practice where bots, scripts, or paid individuals generate non-human interactions to exhaust budgets and distort campaign metrics. While traditional rule-based filters and classical machine learning models often fail to capture the sophisticated temporal dynamics of these fraudulent activities, this project introduces a deep learning-based detection system utilizing Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) architectures. By modeling sequential, time-dependent data and incorporating engineered features such as session timing, click frequency, and behavioral indicators like mouse activity and keystrokes, the LSTM framework effectively distinguishes between legitimate users and fraudulent patterns. When benchmarked against Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), the RNN-LSTM model demonstrated superior performance, achieving approximately 99% accuracy with high precision and recall, thereby validating its efficacy as a robust solution for real-time fraud prevention and restoring trust in digital marketing ecosystems.

SCOPE OF THE PAPER

The scope of this project is centered on the development of an intelligent, efficient system for detecting click fraud in online advertising through advanced deep learning, specifically focusing on the analysis of user interaction data such as click duration, mouse movements, keystrokes, scroll depth, and network attributes. By employing a Long Short-Term Memory (LSTM) Recurrent Neural Network architecture, the project aims to model sequential patterns and uncover complex fraudulent behaviors that traditional methods often overlook. The study encompasses the entire machine learning pipeline—from data preprocessing and feature engineering to model evaluation and deployment via a web-based interface for real-time predictions—targeting advertisers and ad-tech companies seeking to minimize financial loss and maintain campaign integrity. Furthermore, the scope includes a comparative performance analysis of the RNN model against ANN and CNN architectures using metrics like precision and F1-score, while establishing a foundation for future scalability to larger datasets and integration with real-time ad servers.

EXISTING SYSTEM:

The current landscape of click fraud detection in online advertising primarily depends on conventional machine learning (ML) algorithms and heuristic-based mechanisms that utilize historical clickstream data, where features such as click duration, scrolling depth, mouse movement, keystrokes, browser type, device IP reputation, and usage of VPN or proxies are manually selected through feature engineering. Based on this input data, traditional ML models—such as Decision Trees (DT), Random Forests (RF), Gradient Boosting (GB), LightGBM, and XGBoost—are trained to differentiate between fraudulent and legitimate clicks, often achieving precision and recall above 98% under controlled and static datasets. However, despite their initial success, these systems exhibit several key limitations when deployed in real-world advertising platforms, most importantly by treating each user interaction as an independent event and ignoring the sequential and temporal dependencies that reveal fraud patterns over time. Additionally, these models are tightly coupled with static feature extraction processes that restrict their ability to adapt to new fraud strategies, struggle with generalization when faced with unseen or adversarial data distributions, and lack the real-time adaptability required for dynamic, large-scale advertising ecosystems where fraud patterns evolve constantly, ultimately failing to offer a scalable and intelligent solution against sophisticated or bot-driven attacks.

EXISTINGSYSTEM DISADVANTAGES:

- **Ignores temporal behavior** – Fails to consider time-sequence patterns in clicks.
- **Static feature dependency** – Relies heavily on manual feature engineering.
- **Low adaptability** – Struggles with evolving fraud strategies or unseen data.
- **Weak generalization** – Performs poorly on real-world noisy or imbalanced data.
- **Lacks real-time intelligence** – Cannot make dynamic predictions as fraud patterns shift.

PROPOSED SYSTEM

To overcome the limitations of static machine learning, the proposed system utilizes a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units to detect click fraud by capturing sequential and temporal patterns in user behavior. By processing engineered features—such as click duration, scroll depth, and keystrokes—through a pipeline of Label Encoding and standardization, the model learns to identify fraudulent loops and timing anomalies that traditional methods miss. After training with regularization to ensure generalization, the LSTM model achieved a superior 99% accuracy, outperforming CNN, ANN, and classical ML architectures. Integrated into a Flask web application for real-time classification, this solution provides a highly accurate and adaptable defense against evolving click fraud mechanisms in the digital advertising industry.

PROPOSED SYSTEM ADVANTAGES:

- **Captures temporal patterns** – Learns sequential click behavior effectively.
- **Automatic feature learning** – Reduces reliance on manual feature extraction.
- **High fraud detection accuracy** – Achieves up to 99% prediction precision.
- **Robust to noisy data** – Performs well even with unstructured or imbalanced inputs.
- **Adaptable to new threats** – Dynamically detects evolving fraudulent activities.

LITERATURE SURVEY

Title: AI-based Techniques for Ad Click Fraud Detection and Prevention: Review and Research Directions
Author: R. A. Alzahrani and M. Aljabri
Year: 2022.
Description: This paper presents a comprehensive review of artificial intelligence (AI)-based techniques applied to the detection and prevention of click fraud in digital advertising systems. The authors analyze various machine learning (ML) and deep learning (DL) approaches that have been proposed to distinguish between legitimate and fraudulent clicks. The paper highlights the growing complexity of fraud strategies—such as the use of bots, VPNs, and proxies—which make detection more challenging. It also evaluates the strengths and limitations of current detection mechanisms including supervised, unsupervised, and hybrid models. Furthermore, the study emphasizes the importance of features such as user behavior patterns, device information, IP reputation, and click timing in improving the accuracy of detection systems. It identifies gaps in current research, such as the need for real-time detection, scalability, and adaptability to evolving fraud tactics. The authors conclude by proposing several research directions, including the integration of edge computing, reinforcement learning, and blockchain for more secure and efficient click fraud prevention. This review paper serves as a valuable foundation for researchers and developers aiming to build or enhance fraud detection systems using modern AI-driven methods.

Title: Click Fraud Detection Using Ensemble Classifier
Author: A. Purwar, A. K. Jain, I. Chawla, I. Gupta, M. Raj, and D. Jain
Year: 2024.

Description: This research paper focuses on enhancing the detection of click fraud by utilizing ensemble classification techniques. The authors investigate the challenges posed by click fraud, which manipulates online advertising systems by generating fake clicks—resulting in financial losses for advertisers and degrading the integrity of ad analytics. To address these challenges, the paper introduces an ensemble-based approach, which combines the strengths of multiple machine learning classifiers to improve overall prediction accuracy and robustness. The ensemble technique integrates models like Decision Trees, Random Forest, and Gradient Boosting, allowing for better generalization and detection of complex fraud patterns that might not be captured by individual classifiers. The dataset used in the study includes diverse behavioral and contextual features, such as click time, device type, mouse activity, and IP reputation. After preprocessing and feature engineering, the models are trained and evaluated using metrics like accuracy, precision, recall, and F1-score. The results show that the ensemble model outperforms single classifiers, achieving higher detection rates and reducing false positives. The study concludes that ensemble learning is a promising technique for real-time and large-scale click fraud detection and recommends further research into model optimization, scalability, and integration with real-world ad platforms.

Title: A Reliable Click-Fraud Detection System for the Investigation of Fraudulent Publishers in Online Advertising
Author: L. Singh, D. Sisodia, K. Shashvat, A. Kaur, and P. C. Sharma
Year: 2023.
Description: This paper proposes a robust and reliable system aimed at detecting click fraud specifically carried out by fraudulent publishers in online advertising ecosystems. The authors focus on developing a detection framework that can effectively differentiate between legitimate and illegitimate click activity, with a strong emphasis on publisher-side fraud—where publishers intentionally generate fake clicks to increase their revenue. The study highlights various behavioral indicators and user interaction metrics, such as unusual click patterns, repetitive actions, device consistency, and session anomalies. These features are analyzed using machine learning algorithms to identify suspicious behavior. The proposed system incorporates advanced preprocessing and real-time monitoring to ensure timely detection. To validate the approach, the authors perform experiments on real-world advertising datasets. The results demonstrate high accuracy and reliability in identifying fraudulent publishers, outperforming traditional detection techniques. The system is also designed to scale well in dynamic and large-scale advertising environments, making it a practical tool for advertisers and ad networks. This work contributes to the ongoing efforts in safeguarding digital advertising infrastructures and suggests future enhancements like incorporating adaptive learning and integration with blockchain-based transparency solutions.

Title: Click Fraud Detection of Online Advertising Using Machine Learning Algorithms
Author: B. Kirkwood, M. Vanamala, and N. Seliya
Year: 2024
Description This research paper presents a machine learning-based approach to detecting click fraud in online advertising, focusing on improving the accuracy and efficiency of fraud detection systems. The authors address the growing issue of fraudulent clicks, which distort campaign performance metrics and result in significant financial losses for advertisers. The study explores multiple supervised learning algorithms, such as Decision Trees, Support Vector Machines (SVM), and Random Forests, to identify fraudulent behavior patterns from user interaction data. These patterns include abnormal click frequency, irregular session durations, and mismatched geolocation data. By using labeled datasets comprising both legitimate and fraudulent click instances, the models are trained to classify future click activity. A major contribution of the paper is the comparative analysis of different machine learning models based on accuracy, precision, recall, and F1 score. The results reveal that ensemble methods, particularly Random Forests, demonstrate superior performance in handling noisy and imbalanced data. Furthermore, the authors emphasize the importance of feature engineering and the use of real-time behavioral signals to improve detection reliability. The paper concludes with recommendations for integrating ML-driven detection systems into live advertising platforms and suggests future work in incorporating adaptive learning and adversarial robustness.

Title: Quantifying the Cost of Ad Fraud: 2023–2028
Author: Juniper Research, Hampshire, U.K.
Year: July 12, 2024.
Description: This whitepaper by Juniper Research provides an in-depth market analysis of the financial impact of advertising fraud globally from 2023 to 2028. It outlines the evolving tactics used by fraudsters—including botnets, fake installs, and click farms—and their growing sophistication. The report forecasts the monetary losses businesses may face due to these activities and identifies the sectors most affected. The paper emphasizes that without advanced detection mechanisms, the losses caused by ad fraud will continue to escalate, especially in mobile and programmatic advertising. It further

discusses preventive strategies, including the integration of AI-based solutions, real-time monitoring, and collaboration among advertising networks. By quantifying both direct and indirect costs, this report serves as a critical resource for stakeholders aiming to understand and mitigate the risk of fraudulent advertising activity. It also validates the urgent need for more intelligent and proactive fraud detection systems in the digital ad space.

METHODOLOGY

1. **Data Collection:** The process began by gathering clickstream data containing session attributes like device type, browser, and behavioral signals such as mouse movement and scroll depth to establish patterns of genuine versus fraudulent activity.

2. **Preprocessing:** Raw datasets were cleaned by converting timestamps to datetime objects, removing invalid records, and stripping away non-informative identifiers like IP addresses to ensure the data was structured for modeling.

3. **Feature Extraction:** Categorical variables were transformed into numerical values using Label Encoding while continuous features were normalized via StandardScaler to ensure the model could process all inputs on a uniform scale.

4. **Model Design:** A Recurrent Neural Network (RNN) utilizing LSTM layers was architected to capture temporal sequences in user behavior, complemented by Dense and Dropout layers to manage nonlinear transformations and prevent overfitting.

5. **Model Compilation:** The architecture was finalized using the Adam optimizer for dynamic learning rate adjustment and binary cross-entropy as the loss function to effectively handle the binary classification of fraud.

6. **Model Training:** The dataset was split using stratified sampling and trained over 50 epochs, employing Early Stopping and ReduceLROnPlateau to halt training at peak performance and fine-tune the convergence process.

7. **Model Evaluation:** Performance was validated using metrics like Precision, Recall, and F1-Score, where the RNN ultimately achieved an accuracy of approximately 99%, significantly outperforming standard ANN and CNN models.

8. **Model Saving:** The finalized model was exported via TensorFlow, while the preprocessing pipelines were archived using joblib to ensure consistent data transformation during live deployment.

9. **User Interface Development:** A web-based frontend was constructed using HTML, CSS, and Flask, providing a streamlined form for users to input session details for real-time analysis.

10. **Backend Integration:** The Flask backend bridges the interface and the model by capturing user input, reshaping it into a 3D format for the LSTM, and returning a definitive "Fraudulent" or "Legitimate" classification.

IMPLEMENTATION

The Hospital Management System (HMS) is a web-based application designed to manage hospital operations efficiently and securely. It supports three user roles—Administrator, Doctor, and Patient—each with specific functionalities. Administrators manage doctors, patients, and appointments; doctors view and update appointment statuses; and patients can register, book appointments, and track their history.

Built using the Django framework with a MySQL database, the system ensures secure authentication and organized data management. It includes features like appointment scheduling and prevention of double-booking.

The system is user-friendly and suitable for small to medium-sized hospitals. It can be further expanded to include features like billing, medical records, and analytics in the future.

ALGORITHM (XGBoost):

Existing Technique: **DT, RF, GB, LightGBM, XGBoost, CNN, ANN.**

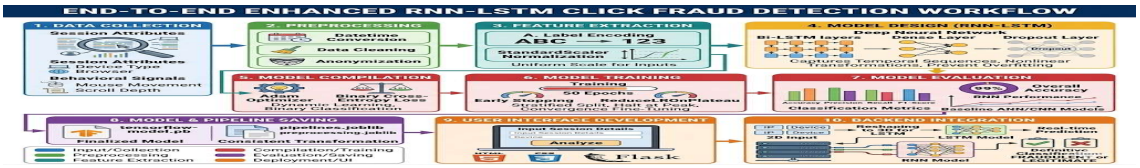
Existing click fraud detection systems primarily leverage traditional machine learning and ensemble algorithms—such as **Decision Trees (DT)**, **Random Forest (RF)**, **XGBoost**, and **LightGBM**—due to their high accuracy (often exceeding 98%) and efficiency with structured data. While these models excel at processing static features like IP addresses and device types, they treat each click as an independent event, failing to capture the **temporal dynamics** and sequential patterns essential for identifying sophisticated, evolving fraud. Consequently, while they provide a robust and interpretable baseline, their inability to account for time-based behavioral shifts limits their effectiveness in adaptive, real-time environments.

Proposed Technique: **RNN(LSTM)**

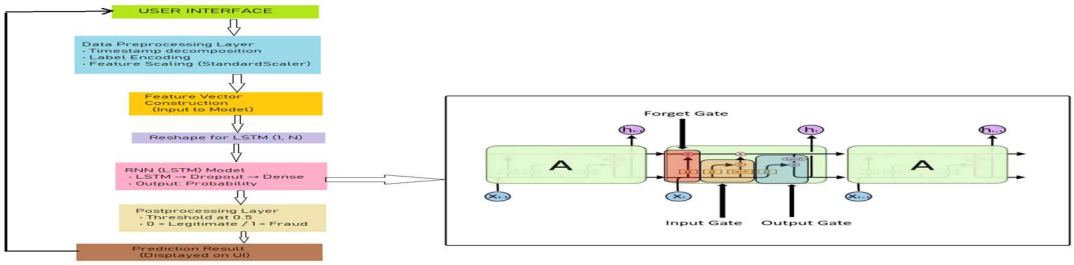
The proposed **LSTM-based model** enhances click fraud detection by analyzing **temporal sequences**

and behavioral patterns—like scroll depth and timing—that traditional static models miss. By leveraging memory cells to retain long-term dependencies in user interactions, the system achieves over **99% accuracy**. This focus on sequential dynamics allows the model to effectively identify sophisticated, evolving bot behavior in real-time advertising environments.

BLOCK DIAGRAM

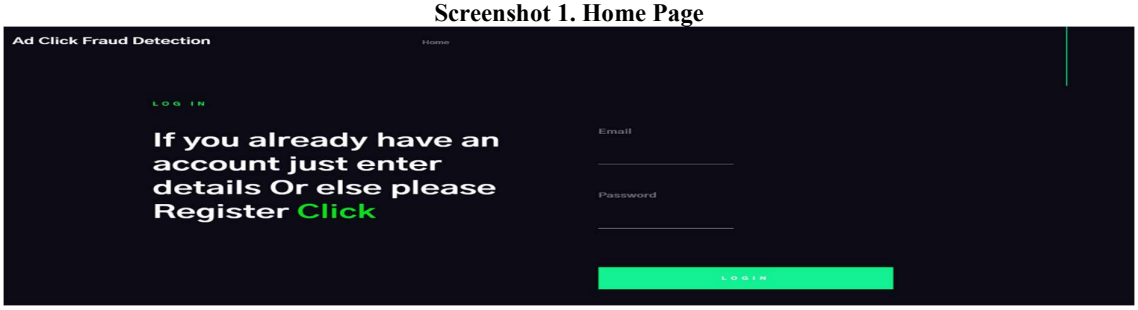


SYSTEM ARCHITECTURE



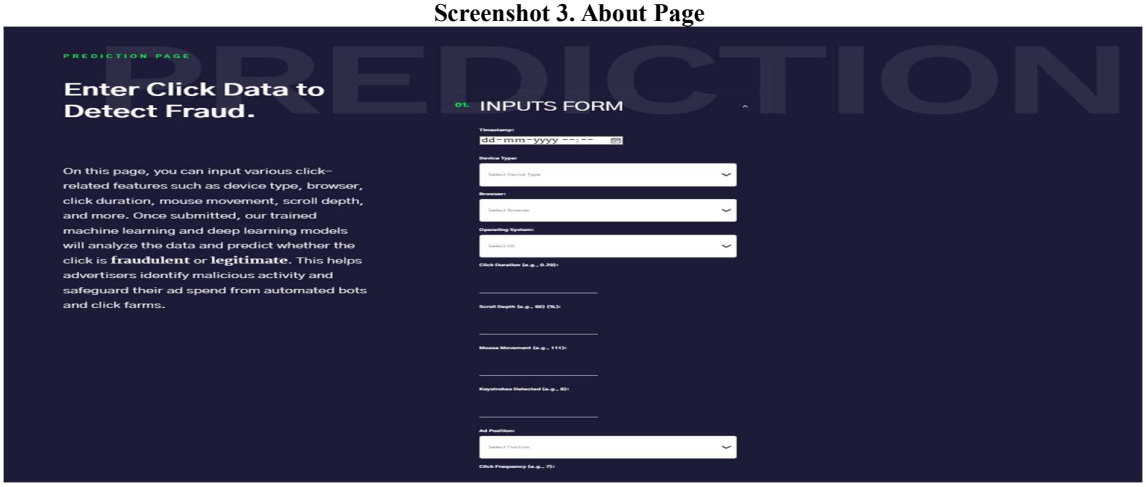
RESULT

Screenshot 1. Home Page



Screenshot 2. Login Page

Screenshot 3. About Page



Screenshot 4. Input Page

Screenshot 5. Prediction Page

Screenshot 6. Performance Page

CONCLUSION

Click fraud remains a critical issue in the online advertising ecosystem, leading to significant financial losses and undermining the trust of advertisers. This project aimed to address this challenge by developing a robust fraud detection system using advanced deep learning techniques. Through a carefully designed pipeline involving data preprocessing, feature engineering, and model development, we evaluated multiple machine learning and deep learning models, including CNN, DNN, and RNN (LSTM). Among them, the proposed RNN-based model demonstrated superior performance, achieving over 99% accuracy in detecting fraudulent clicks. Its ability to learn temporal patterns from user behavior proved highly effective in differentiating between legitimate and fraudulent activities. The model was successfully integrated into a web-based interface, allowing real-time user input and fraud prediction. Overall, this study demonstrates the power of sequence-aware architectures like RNN in identifying complex patterns of online fraud, and it lays a strong foundation for developing future real-time, scalable, and adaptive fraud detection solutions in digital advertising

FUTURE SCOPE:

To further improve the robustness and real-time effectiveness of click fraud detection, several future enhancements can be explored. One significant direction is the integration of real-time streaming data processing using platforms like Apache Kafka or Spark Streaming. This would allow the system to detect fraudulent activity as it happens, enhancing security and response time. Another enhancement involves incorporating unsupervised learning techniques such as autoencoders or clustering algorithms to detect novel fraud patterns that may not be captured in labeled datasets. The system could also benefit from advanced deep learning architectures like Bidirectional LSTM (BiLSTM) or Transformers, which can better capture the sequential dependencies and temporal patterns in user interactions. Additionally, implementing user profiling and behavioral analytics could help model long-term patterns of individual users or bots, improving the contextual understanding of clicks. Finally, deploying the model in a cloud-based scalable environment using containers (e.g., Docker) and orchestration (e.g., Kubernetes) would ensure high availability and scalability, especially for production-scale traffic. These enhancements would make the system more intelligent, adaptive, and resilient against emerging fraud tactics in online advertising ecosystems.

REFERENCES

[1] Juniper Research, Hampshire, U.K. Quantifying the Cost of Ad Fraud: 2023–2028. Accessed: Jul. 12, 2024. [Online]. Available: https://fraudblocker.com/wp-content/uploads/2023/09/Ad-Fraud-Whitepaper_Juniper-Research.pdf

[2] X. Zhu, H. Tao, Z. Wu, J. Cao, K. Kalish, and J. Kayne, Fraud Prevention in Online Digital Advertising. Cham, Switzerland: Springer, 2017.
[3] A. K. Wood and A. M. Ravel, "Fool me once: Regulating fake news and other online advertising," S. Cal. L. Rev., vol. 91, p. 1223, Jan. 2017.
[4] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna, "Understanding fraudulent activities in online ad exchanges," in Proc. ACM SIGCOMM Conf. Internet Meas. Conf., Nov. 2011, pp. 279–294.
[5] (2024). Wasted Ad Spend Report 2024. [Online]. Available: https://lp. lunio.ai/wp-content/uploads/2023/09/Lunio_Wasted_Ad_Spend _Report_ 2024_V2.pdf 12762
[6] D. Berrar,"Randomforestsforthedetectionofclickfraudi nonlinemobile advertising," in Proc. Int. Work. Fraud Detect. Mob. Advert. (FDMA), Singapore, 2012, pp. 1–10. [Online]. Available: http://berrar.com/ resources/Berrar_FDMA2012.pdf
[7] J. H. Yan and W. R. Jiang, "Research on information technology with detecting the fraudulent clicks using classification method," Adv. Mater. Res., vol. 859, pp.586–590, Dec. 2013, doi: 10.4028/www.scientific.net/amr.859.586.
[8] K. S. Perera, B. Neupane, M. A. Faisal, Z. Aung, and W. L. Woon, "A novelensemblelearning-basedapproachforclickfrauddetectioninmobile advertising," in Mining Intelligence and Knowledge Exploration (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8284. Berlin, Germany: Springer, 2013, pp. 370–382, doi: 10.1007/978-3-319-03844-5_38.
[9] C. Phua, E.-Y. Cheu, G.-E. Yap, K. Sim, and M.-N. Nguyen, "Feature engineering for click fraud detection," in Proc. Work. Fraud Detect. Mob. Advert., 2012, pp. 1–10. [Online]. Available: http://palanteer.sis.smu.edu. sg/fdma2012/doc/FirstWinner-Starrystarrynight-Paper.pdf%5Cnpapers2: //publication/uuid/9290A6CF-A861-4058-99F4-D39706B0619A
[10] E.-A. Minastireanu and G. Mesnita, "Light GBM machine learning algorithm to online click fraud detection," J. Inf. Assurance Cybersecur., vol. 2019, pp. 1–12, Apr. 2019, doi: 10.5171/2019.263928.
[11] D. Sisodia and D. S. Sisodia, "Gradient boosting learning for fraudulent publisher detection in online advertising," Data Technol. Appl., vol. 55, no. 2, pp. 216–232, Apr. 2021, doi: 10.1108/dta-04-2020-0093.
[12] A.DashandS.Pal,"Auto-detectionofclick-fraudsusingmachinelearning Auto-detection of click-frauds using machine learning," Int. J. Eng. Sci. Comput., vol. 10, pp. 27227–27235, Sep. 2020.
[13] R. Mouawi, M. Awad, A. Chehab, I. H. E. Hajj, and A. Kayssi, "Towards a machine learning approach for detecting click fraud in mobile advertizing," in Proc. Int. Conf. Innov. Inf. Technol. (IIT), Nov. 2018, pp. 88–92, doi: 10.1109/INNOVATIONS.2018.8605973.
[14] M. Aljabri and R. M. A. Mohammad, "Click fraud detection for online advertising using machine learning," Egyptian Informat. J., vol. 24, no. 2, pp. 341–350, Jul. 2023, doi: 10.1016/j.eij.2023.05.006.
[15] S. Shaik and V. Kakulapati, "Fraud detection of AD clicks using machine learning techniques," J. Sci. Res. Rep., vol. 29, no. 7, pp. 84–89, Jun. 2023, doi: 10.9734/jsrr/2023/v29i71762.
[16] D. Sisodia and D. S. Sisodia, "Stacked generalization architecture for predicting publisher behaviour from highly imbalanced user-click data set for click fraud detection," New Gener. Comput., vol. 41, no. 3, pp. 581–606, Sep. 2023, doi: 10.1007/s00354-023-00218-1.
[17] D. Sisodia, D. S. Sisodia, and D. Singh, "Evaluating feature importance to investigate publishers conduct for detecting click fraud," in Machine Intelligence Techniques for Data Analysis and Signal Processing (Lecture Notes in Electrical Engineering), vol. 997. Berlin, Germany: Springer, 2023, pp. 515–524, doi: 10.1007/978-981-99-0085-5_42.
[18] R. Dekou, S. Savo, S. Kufeld, D. Francesca, and R. Kawase, "Machine learning methods for detecting fraud in online marketplaces," in Proc. CEURWorkshop, vol. 3052, Jan. 2021, pp. 3–7.
[19] D.Sisodia and D. S. Sisodia, "Quad division prototype selection-based Knearest neighbor classifier for click fraud detection from highly skewed user click dataset," Eng. Sci. Technol., Int. J., vol. 28, Apr. 2022,Art. no. 101011, doi: 10.1016/j.jestch.2021.05.015.
[20] B.Kirkwood,M.Vanamala,andN.Seliya,"Clickfraud detectionofonline advertising using machine learning algorithms," in Proc. IEEE Int. Conf. Electro Inf. Technol. (eIT), May 2024, pp. 586–590. [Online]. Available: https://api.semanticscholar.org/CorpusID
[21] L. Singh, D. Sisodia, K. Shashvat, A. Kaur, and P. C. Sharma, "A reliable click-fraud detection system for the investigation of fraudulent publishers in online advertising," in Applied Intelligence in Human-Computer Interaction. Boca Raton, FL, USA: CRC Press, Jul. 2023.
[22] A. Batool and Y.-C. Byun, "An ensemble architecture based on deep learning model for click fraud detection in Pay-Per-Click advertisement campaign," IEEE Access, vol. 10, pp. 113410–113426, 2022, doi: 10.1109/ACCESS.2022.3211528.
[23] A. Purwar, A. K. Jain, I. Chawla, I. Gupta, M. Raj, and D. Jain, "Click fraud detection using

ensemble classifier," in Proc. Int. Conf. Artif.-Bus. Anal., Quantum Mach. Learn., Jan. 2024, pp. 15–23.
[24] R. A. Alzahrani and M. Aljabri, "AI-based techniques for ad click fraud detection and prevention: Review and research directions," J. Sensor Actuator Netw., vol. 12, no. 1, p. 4, Dec. 2022, doi: 10.3390/jsan12010004. VOLUME 13, 2025R. A. Alzahrani et al.: Ad Click Fraud Detection Using Machine Learning and Deep Learning Algorithms
[25] VeracityTrustNetwork.VeracityTrustNetwork—OnlyHumans.Accessed: Feb. 4, 2023. [Online]. Available: https://veracitytrustnetwork.com/
[26] K. Mehrabani-Zeinabad, M. Doostfatemeh, and S. M. T. Ayatollahi, "An efficient and effective model to handle missing data in classification," BioMed Res. Int., vol. 2020, pp. 1–2, 2020, doi: 10.1155/2020/8810143.
[27] J. D. Kelleher, B. Mac Namee, and A. D'arcy, Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge, MA, USA: MIT Press, 2020.
[28] D. J. Stekhoven and P. Bühlmann, "MissForest—Non-parametric missing value imputation for mixed-type data," Bioinformatics, vol. 28, no. 1, pp. 112–118, Jan. 2012, doi: 10.1093/bioinformatics/btr597.
[29] S. Hong and H. S. Lynn, "Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction," BMC Med. Res. Methodol., vol. 20, no. 1, pp. 1–12, Dec. 2020, doi: 10.1186/s12874-020-01080-1.
[30] A. K. Waljee, A. Mukherjee, A. G. Singal, Y. Zhang, J. Warren, U. Balis, J. Marrero, J. Zhu, and P. D. Higgins, "Comparison of imputation methods for missing laboratory data in medicine," BMJ Open, vol. 3, no. 8, Aug. 2013, Art. no. e002847, doi: 10.1136/bmjopen-2013-002847.
[31] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," Expert Syst. Appl., vol. 73, pp. 220–239, May 2017, doi: 10.1016/j.eswa.2016.12.035.
[32] N.JapkowiczandS.Stephen,"Theclassimbalanceproblem:Asystematic study," Intell. Data Anal., vol. 6, no. 5, pp. 429–449, Nov. 2002, doi: 10.3233/ida-2002-6504.
[33] D.FreedmanandP.Diaconis,"Onthehistogramasadensityestimator: L2 theory," Zeitschrift Für Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 57, no. 4, pp. 453–476, Dec. 1981, doi: 10.1007/bf01025868.
[34] H. A. Le Thi, V. V. Nguyen, and S. Ouchani, "Gene selection for cancer classification using DCA," in Advanced Data Mining and Applications (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 5139. Berlin, Germany: Springer, 2008, pp. 62–72, doi: 10.1007/978-3-540-88192-6_8.
[35] B.M.Randles, I. V. Pasquetto, M. S. Golshan, and C. L. Borgman, "Using the Jupyter notebook as a tool for open science: An empirical study," in Proc. ACM/IEEE Joint Conf. Digit. Libraries (JCDL), Jun. 2017, pp. 1–2, doi: 10.1109/JCDL.2017.7991618.
[36] A. Kumbhar, P. G. Dhawale, S. Kumbhar, U. Patil, and P. Magdum, "A comprehensive review: Machine learning and its application in integrated power system," Energy Rep., vol. 7, pp. 5467–5474, Nov. 2021, doi: 10.1016/j.egyr.2021.08.133.