International Journal of
Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

Ad Click Fraud Detection Using Machine Learning and Deep Learning Algorithms

Dondapati Veera Swamy[1], Motha Prameela[2]. Sk. Badar[3], Sevugarajan[4], Sateesh Yalavarthi[5]
#1 Post Graduate Scholar, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India, 521212
#2 M.Tech (PHD), Asst Professor, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India, 521212
#3Associate Professor, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India, 521212
#4 Professor, Department of ECE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India, 521212
#5 Post Graduate Scholar, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India, 521212
Mail Id: veeraswamy5911@gmail.com[1], mothaprameela79@gmail.com [2], 444.badar@gmail.com [3] , dean_academics@vikasinstitutionsnunna.org [4], sateeshthermal@gmail.com [5]

Abstract: *Online advertising systems face a serious threat from ad click fraud, which produces fraudulent clicks that skew campaign results and raise losses. This paper provides an attention-based Convolutional Neural Network (CNN) framework for real-time ad click fraud detection in order to overcome the shortcomings of traditional machine learning and deep learning models in collecting complicated contextual patterns. In order to improve prediction reliability, the suggested improvement incorporates an attention mechanism into the CNN design to highlight the most interesting characteristics and sequential relationships in clickstream data. Furthermore, real-time fraud prediction via an interactive interface is made possible by a web-based deployment utilizing the Flask framework. The attention-enhanced CNN outperforms baseline models in terms of accuracy and resilience, according to experimental assessment, underscoring its usefulness for scalable and realistic ad click fraud detection in dynamic online advertising contexts.*

Index terms - *— Ad Click Fraud Detection, Attention Mechanism, Convolutional Neural Network (CNN), Deep Learning, Clickstream Analysis, Real-Time Prediction, Flask Deployment, Online Advertising Security, Fraud Analytics, Feature Learning*

INTRODUCTION

Ad click fraud is becoming a serious threat to the dependability and financial viability of digital advertising platforms due to the fast growth of online advertising. Inaccurate performance indicators, higher advertising expenses, and a decline in confidence between publishers and advertisers are all consequences of fraudulent clicks produced by automated bots or organized malevolent people. Despite the widespread use of machine learning and deep learning approaches to solve this problem, many of the models now in use have difficulty capturing the contextual relevance and sequential relationships seen in clickstream data, especially in real-time situations.

Conventional CNN designs treat all extracted characteristics equally, despite the fact that Convolutional Neural Networks (CNNs) have shown great effectiveness in learning spatial and behavioral patterns from high-dimensional data. This restriction lessens their usefulness in intricate fraud situations when only particular behavioral or temporal characteristics are actually suggestive of fraudulent activity. Attention mechanisms, which allow models to choose focus on the most useful aspects, have emerged as a potent solution to this problem, boosting contextual knowledge and decision-making accuracy.

This article proposes an attention-based CNN approach that prioritizes important clickstream characteristics during training and inference to improve ad click fraud detection. The model's capacity to spot subtle and changing fraud patterns that conventional CNNs and machine learning classifiers could miss is strengthened by the incorporation of the attention layer. Additionally, a Flask-based web interface is added to the system to enable real-time prediction, allowing marketers and advertising platforms to implement it practically. This addon provides a scalable and flexible method for identifying fraudulent ad click activity in dynamic online advertising ecosystems, bridging the gap between high-accuracy deep learning models and practical usability.

LITERATURE SURVEY

a) Fool Me Once: Regulating 'Fake News' and Other Online Advertising:

The opaqueness of online political advertising has been a recurring problem in American elections. American voters have been the target of misinformation operations since the 2016 election, which has made regulatory action more urgent.

Only recently have internet corporations begun to favor self-regulation and support proposed transparency regulations. Although it is not necessary to regulate the content of political speech, the government may and should enforce transparency in the process. We recommend many actions to encourage openness. Campaign finance legislation should, first and foremost, mandate that platforms archive and make available (1) advertising that have appeared on their platforms and (2) the target demographic for those ads. Audience availability can be used to get around privacy concerns in addition to fulfilling a critical speech aim under the First Amendment's "marketplace of ideas" notion, which is to allow counter-speech. Our proposed regulation would encompass all online political advertising, including disinformation, that is distributed through paid social media. Second, legislation pertaining to online advertising transparency should be strengthened. Congress has a part in this as it has prevented regulatory agencies from acting to require disclosure from so-called dark money groups. Last but not least, the government should require social media businesses to offer an opt-in feature so that people may view content that is disputed or advertisements that are directly directed at them.

b) Understanding fraudulent activities in online ad exchanges:

Online advertising, or ads, are an effective way for advertisers to target Web users. Ads can be customized based on a user's location, interests, and browsing patterns. Some of the most popular websites on the Internet rely on online advertising, which is currently valued at billions of dollars. In order to meet the massive market demand, marketplaces known as "ad exchanges" are utilized to manage the complex interactions between publishers (i.e., the websites hosting the advertising) and advertisers. across these transactions, publishers, who sell this ad space, and advertisers, who buy it, may dynamically broker traffic across ad networks to maximize profits for all parties. Unfortunately, the complexity of these systems permits a great deal of abuse by hackers, who profit from the advertising.

In this article, we present a comprehensive examination of the operation of one of the largest ad exchanges and the associated security issues from the viewpoint of a member ad network. More specifically, we looked at a dataset that contained both incoming and outgoing ad traffic transactions for this ad network. Additionally, we examined information obtained from a command-and-control server that operates a botnet to perpetrate ad fraud against the same ad exchange.

c) Click fraud detection for online advertising using machine learning:

Advertising firms have turned their focus to online and in-app advertisements as social media and digital technologies have increased. Online advertising is the primary revenue stream for advertising networks, which serve as a conduit between advertisers and ad publishers. The advertising networks pay the ad publisher based on the number of clicks that go to advertisers using the pay-per-click (PPC) payment model. A growing security problem connected to this payment mechanism is click fraud. Click fraud is the illegal activity of clicking on pay-per-click advertisements to increase publishers' income or deplete advertisers' budgets. With unexpected outcomes, artificial intelligence techniques are being applied more and more to solve challenging issues in a number of academic disciplines, including cybersecurity. This article developed many machine learning models and compared their results using a set of evaluation metrics to identify if a user is a human or a bot. We used a real dataset that was gathered to explain user behavior on websites. We obtained a set of user behavior-related attributes, including the number of webpages seen, the duration of the browsing session, and the activities performed. The empirical data showed that all of the models under examination delivered good results, with the random forest algorithm surpassing all other algorithms across all assessment metrics.

d) Research on Information Technology with Detecting the Fraudulent Clicks Using Classification Method:

The goal of this study is to develop an effective technique for spotting fraudulent clicks in commissioners' click logs. Based on the data they give about their ad clicks, we hope to predict whether or not a person is dangerous. Using the training data, we build our fraudulent click detection classification models. We initially create and extract characteristics from the above-mentioned raw log data. We choose models and build an ensemble model after evaluating classifiers. Finally, we tested our model for human judge reference using the real dataset.

e) A Novel Ensemble Learning-Based Approach for Click Fraud Detection in Mobile Advertising:

Click fraud is a significant financial burden on advertising budgets and has the potential to significantly reduce the profitability of the online advertising sector by depriving legitimate partners, commonly referred to as publishers, of their money. Fraud detection algorithms that can identify fraudulent activities by examining user click patterns are

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

therefore quite helpful. Based on the BuzzCity dataset, we propose a novel click fraud detection technique. The foundation of this approach is a set of novel traits derived from preexisting attributes. The proposed model is evaluated using the resulting accuracy, recall, and area under the ROC curve. The final ensemble model, which was based on six different learning techniques, showed stability in all three performance criteria. Our final model shows better results on training, validation, and test datasets, indicating its generalizability to additional datasets.

METHODOLOGY

i) Proposed Work:

The proposed system uses an attention-based Convolutional Neural Network (CNN) architecture to improve the accuracy and reliability of ad click fraud detection in complex advertising situations. Clickstream data is initially preprocessed by encoding category attributes, normalizing numerical features, and selecting relevant features to reduce noise and increase learning efficiency. The CNN architecture is designed to extract crucial geographical and behavioral patterns from the processed data, while the integrated attention mechanism dynamically increases priority to major features and temporal interactions that plainly signal fraudulent conduct. This enables the model to capture contextual relevance and sequential linkages more accurately than conventional CNN-based techniques.

To ensure practical application, the proposed model is implemented as a real-time fraud detection system with a Flask-based web interface. Customers may use the implementation to stream or upload click data and receive real-time forecasts regarding the legitimacy of ad clicks. This real-time connectivity enhances platform administrators' and advertisers' usability by enabling continuous monitoring and quick reaction to fraudulent activity. The attention-enhanced deep learning architecture and web-based deployment provide a scalable, adaptable, and efficient solution that can handle large datasets and evolving fraud tendencies in online advertising ecosystems.

ii) System Architecture:

The system architecture of the proposed expansion is designed to support attention-based deep learning and real-time fraud detection. After being collected from online advertising platforms, raw clickstream data is routed via a preprocessing layer that handles missing values, encodes category attributes, and normalizes numerical features in order to ensure consistent data representation. After feature refinement, only the most important traits for fraud detection are retained. Following processing, the data is sent into a Convolutional Neural Network (CNN), which use click interactions to find patterns in behavior and space. An attention layer is built on top of the CNN feature maps to allow the model to focus on critical click behaviors that indicate fraudulent activity. Important traits and temporal patterns are given dynamic weights by this layer.

In the following stage, a classification layer receives the attention-enhanced CNN outputs and determines if each click occurrence is genuine or fake. The trained model is implemented in a Flask-built web framework to enable real-time inference. By streaming live click records or uploading test data, users may quickly obtain fraud prediction conclusions. The architecture's flexibility and scalability allow it to be easily connected with large-scale advertising systems and changed often in response to shifting fraud patterns. Our end-to-end architecture effectively blends state-of-the-art attention-based deep learning with practical, real-world implementation for dependable ad click fraud detection.

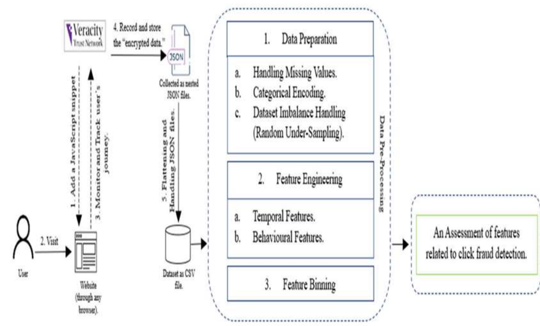

Fig1 proposed architecture

iii) Modules:

1. Data Collection and Preprocessing Module

This module handles the acquisition of raw clickstream data from advertising platforms. It performs data cleaning, label encoding for categorical features, normalization of numerical values, and preparation of structured inputs suitable for deep learning models.

2. Feature Extraction and Refinement Module

Relevant features are refined to reduce noise and improve learning efficiency. This module ensures that only informative attributes contributing to fraud detection are retained before being passed to the model.

3. Attention-Based CNN Modeling Module

This core module implements the Convolutional Neural Network integrated with an attention mechanism. The CNN extracts behavioral patterns, while the attention layer emphasizes critical features and sequential dependencies for accurate fraud classification.

4. Model Training and Evaluation Module

The attention-based CNN is trained using labeled click data and evaluated using performance metrics such as

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

accuracy, precision, recall, and F1-score to ensure robustness and reliability.

5. Real-Time Prediction and Deployment Module

This module deploys the trained model using a Flask-based web interface, enabling real-time fraud prediction. Users can upload click data and instantly receive classification results, supporting practical and scalable fraud detection in live advertising environments.

iv) ALGORITHMS:

a) Deep Neural Network (DNN):

DNN is a multilayer feedforward neural network that can represent intricate, nonlinear interactions. It examines varied data such as click timing, user behavior, and device information to detect fraudulent activities. DNNs learn hierarchical representations that boost anomaly detection and improve performance on large-scale datasets.

b) Recurrent Neural Network (RNN):

RNN is a neural network architecture designed to capture temporal relationships by maintaining recollection of prior inputs. It is used to detect abnormalities in time-based patterns and represent sequential click behavior. RNNs are quite good at spotting fraud situations where the sequence and timing of clicks are important clues.

c) CNN with Attention Mechanism:

The attention-enhanced CNN mixes convolutional feature extraction with an attention layer to focus on the most informative features and time steps. This methodology increases contextual awareness and sequential pattern detection in clickstream data. By highlighting key patterns and decreasing noise, the attention-based CNN delivers higher fraud detection performance compared to classic CNN and machine learning models.

EXPERIMENTAL RESULTS

Several machine learning and deep learning models for fraudulent click detection were tested using a labeled ad click dataset. The dataset was preprocessed by handling missing values, label encoding categorical variables, normalizing numerical features, and choosing features using Recursive Feature Elimination. Training and testing sets received an 80:20 split of the processed data. Experiments showed that ensemble-based models capture complex feature interactions and manage data imbalance better than traditional classifiers. Several models, including Logistic Regression, Decision Tree, Random Forest, KNN, Naïve Bayes, Support Vector Machine, Gradient Boosting, XGBoost, and LightGBM, were trained and evaluated using standard performance metrics like accuracy, precision, recall, and F1.

When it came to identifying fraudulent click patterns, Random Forest, Decision Tree, Gradient Boosting, and XGBoost all scored better than 98%. In competition with deep learning models, the Recurrent Neural Network replicated consecutive click action with 97% accuracy. With 99.73% accuracy, the attention-enhanced Convolutional Neural Network outperformed the others, highlighting the significance of attention processes for sequential pattern recognition and contextual feature weighting. These findings demonstrate how feature selection and attention-based deep learning enhance fraud detection, lower false positives, and offer a practical and expandable online advertising solution.

The attention-based CNN achieves 99.73% accuracy thanks to effective data preprocessing, feature selection, and an attention mechanism that prioritizes discriminative temporal and behavioral aspects. While CNN-based hierarchical feature extraction enhances generalization and fraud detection, the attention layer concentrates on noteworthy click patterns to minimize noise.

Accuracy: A test's accuracy is determined by its capacity to distinguish between healthy and ill cases. To gauge the accuracy of the test, find the percentage of examined instances that had true positives and true negatives. According to the computations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Precision is the number of affirmative cases or the classification's accuracy rate. The following formula is applied to assess accuracy:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's ability to recognise every instance of a pertinent machine learning class is measured by its recall. The ratio of accurately predicted positive observations to the total number of positives indicates how well a model can identify class instances.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision is one ranking quality metric (MAP). It considers the number of relevant recommendations and their position on the list. MAP at K is calculated as the arithmetic mean of the Average Precision (AP) at K for each user or query.

$$\boldsymbol{mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k}$$

$$\boldsymbol{AP_k = \text{ the AP of class } k}$$

$$\boldsymbol{n = \text{ the number of classes}}$$

ISSN 2347–3657

Volume 14, Issue 2, 2026

F1-Score: An accurate machine learning model is indicated by a high F1 score. combining precision and recall to increase model correctness. The accuracy statistic quantifies the frequency with which a model correctly predicts a dataset.

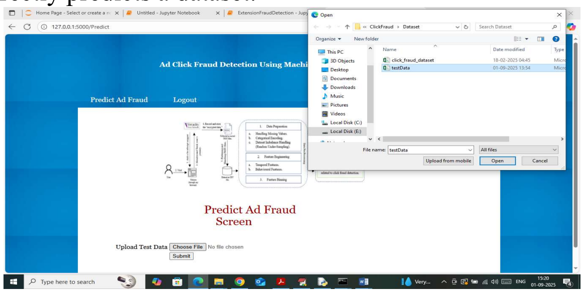

Figure 7: Test Data Upload Module

This figure illustrates the test file upload interface, where users can upload CSV or Excel files containing click data. The uploaded data is used for real-time fraud prediction through the trained model.

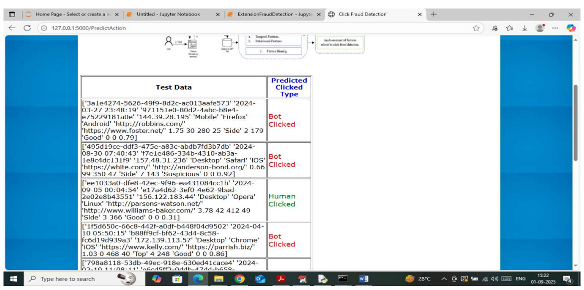

Figure 8: Prediction Results

This figure presents the prediction results for uploaded click data. Each record is classified as Human Clicked or Bot Clicked, demonstrating the system's ability to detect fraudulent clicks accurately.

CONCLUSION

This work provides an effective extension to ad click fraud detection by integrating an attention-based Convolutional Neural Network with real-time deployment capabilities. The attention mechanism enhances the model's ability to focus on significant clickstream features and contextual patterns when compared to conventional deep learning approaches, leading to more reliable and accurate fraud detection. Additionally, the Flask-based web interface helps bridge the gap between high-performance deep learning models and real-world advertising applications by enabling usable, real-time prediction. All things considered, the proposed extension offers a scalable, adaptable, and efficient means of combating evolving ad click fraud in dynamic online advertising environments.

FUTURE SCOPE

Future iterations of this system can focus on incorporating complex sequence-learning models, such as Transformer topologies, to further enhance long-range dependency modeling in clickstream data. Incremental and online learning might be added to the system, enabling the model to continuously adapt in real time to newly identified fraud tendencies. Additionally, it is possible to enhance detection across different advertising platforms while maintaining privacy by utilizing federated learning and anomaly detection techniques. Technologies for cloud-based and edge computing can improve deployment scalability, allowing the system to operate efficiently in large, busy online advertising environments.

REFERENCES

[1] Juniper Research, Hampshire, U.K. Quantifying the Cost of Ad Fraud: 2023–2028. Accessed: Jul. 12, 2024. [Online]. Available: https://fraud blocker.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

[6] D. Berrar, ''Random forests for the detection of click fraud in online mobile 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 novel ensemble learning-based approach for click fraud detection in mobile 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. Dash and S. Pal, ''Auto-detection of click-frauds using machine learning 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. CEUR Workshop, 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, and N. Seliya, ''Click fraud detection of online 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.

[25] Veracity Trust Network. Veracity Trust Network—Only Humans. 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.

Author

Dondapati Veera Swamy[1],
Post Graduate Scholar, Department of CSE, Vikas College of Engineering & Technology, Nunna, Andhra Pradesh, India, 521212
veeraswamy5911@gmail.com