International Journal of Engineering Research and Science & Technology

Real-Time Ad Click Fraud Detection Using Attention-Enhanced CNN Architecture

1Dr D.SRUJAN CHANDRA REDDY, Professor, CSE Department, PBR Visvodaya Institute of Technology and Science, Kavali-A.P

2S.SWETHA(M.Tech), Computer Science and Engineering, PBR Visvodaya Institute of Technology and Science, Kavali-A.P

Abstract: Ad click fraud poses a significant challenge to online advertising platforms by generating illegitimate clicks that distort campaign performance and increase financial losses. To address the limitations of conventional machine learning and deep learning models in capturing complex contextual patterns, this work presents an attention-based Convolutional Neural Network (CNN) framework for real-time ad click fraud detection. The proposed extension integrates an attention mechanism within the CNN architecture to emphasize the most informative features and sequential dependencies in clickstream data, thereby improving prediction reliability. Additionally, a web-based deployment using the Flask framework enables real-time fraud prediction through an interactive interface. Experimental evaluation demonstrates that the attention-enhanced CNN achieves superior accuracy and robustness compared to baseline models, highlighting its effectiveness for scalable and practical ad click fraud detection in dynamic online advertising environments.

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

1. INTRODUCTION

The rapid expansion of online advertising has intensified the impact of ad click fraud, making it a critical threat to the reliability and economic sustainability of digital advertising platforms. Fraudulent clicks generated by automated bots or coordinated malicious users lead to inflated advertising costs, inaccurate performance metrics, and reduced trust between advertisers and publishers. Although machine learning and deep learning techniques have been widely applied to address this issue, many existing models struggle to capture contextual relevance and sequential dependencies present in clickstream data, particularly in real-time environments.

Convolutional Neural Networks (CNNs) have demonstrated strong performance in learning spatial and behavioral patterns from high-dimensional data; however, conventional CNN architectures treat all extracted features with equal importance. This limitation reduces their effectiveness in complex fraud scenarios where only specific temporal or behavioral attributes are truly indicative of fraudulent activity. To overcome this challenge, attention mechanisms have emerged as a powerful enhancement that enables models to selectively focus on the most informative features, thereby improving

contextual understanding and decision-making accuracy.

In this work, an attention-based CNN framework is proposed to enhance ad click fraud detection by prioritizing critical clickstream features during training and inference. The integration of the attention layer strengthens the model's ability to identify subtle and evolving fraud patterns that traditional CNNs and machine learning classifiers may overlook. Furthermore, the system is extended with a Flask-based web interface to support real-time prediction, enabling practical deployment for advertisers and advertising platforms. This extension bridges the gap between high-accuracy deep learning models and real-world usability, offering a scalable and adaptive solution for detecting fraudulent ad click activities in dynamic online advertising ecosystems.

2. LITERATURE SURVEY

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

https://www.semanticscholar.org/paper/Fool-Me-Once%3A-Regulating-%27Fake-News%27-and-Other-Wood-Ravel/b725c99a93997321fe22377864e3c2e7f49170a3

A persistent issue in American political elections has been the opaqueness of internet political advertising. Since the 2016 campaign, American voters have been subjected to disinformation campaigns, which has increased the urgency of regulatory action.

Internet companies have just lately come around to supporting proposed transparency laws and prefer self-regulation. The government may and ought to impose openness on the political speech process, even though it is not required to control its substance. We suggest a number of measures to promote transparency. First and foremost, platforms should be required by campaign finance regulations to save and make accessible (1) advertisements that have appeared on their platforms and (2) the target demographic for those ads. under addition to meeting a crucial speech goal under the First Amendment's "marketplace of ideas" theory—that of permitting counter-speech—audience availability can be designed to circumvent privacy issues. All online political advertising, including misinformation, that is disseminated through paid social media distribution would be covered by our proposed legislation. Second, the current gaps in internet advertising transparency laws ought to be filled. Since it has stopped regulatory bodies from taking action to demand disclosure from so-called dark money organizations, Congress has a role in this. Lastly, the government need to mandate that social media companies provide an opt-in mechanism so that individuals can read information that is challenged or adverts that are specifically targeted against them.

ii) Understanding fraudulent activities in online ad exchanges

https://dl.acm.org/doi/10.1145/2068816.2068843

Advertisers may successfully target Web users with the use of online advertisements, or ads. Ads can be tailored according to a user's interests, location, and browsing habits. Online advertising, which is presently worth billions of dollars, is the main source of income for some of the most well-known websites

on the Internet. Marketplaces referred to as "ad exchanges" are used to handle the intricate interactions between publishers (i.e., the websites hosting the advertising) and advertisers in order to fulfill the enormous market demand. Through these exchanges, advertisers (who purchase this ad space) and publishers (who sell it) may dynamically broker traffic through ad networks in order to effectively optimize earnings for everybody involved. Regretfully, the intricacies of these systems allow for a significant amount of misuse by hackers, who make money off of the advertising.

From the perspective of a member ad network, we provide a thorough analysis of the workings of one of the biggest ad exchanges and the related security concerns in this paper. More precisely, we examined a dataset that included transactions for this ad network's incoming and outgoing ad traffic. Furthermore, we looked at data gathered from a command-and-control server that runs a botnet in order to commit ad fraud against the same ad exchange.

iii) Click fraud detection for online advertising using machine learning

https://www.sciencedirect.com/science/article/pii/S1110866523000294

As digital technology and social media have grown, advertising companies have shifted their attention to online and in-app ads. The main source of income for advertising networks, which act as a bridge between advertisers and ad publishers, is online advertising. Using the pay-per-click (PPC) payment model, the advertising networks compensate the ad publisher according to the quantity of clicks that go to advertisers. Click Fraud is a rising security risk associated with this payment method. The unlawful practice of clicking on pay-per-click ads in order to boost publishers' profits or drain advertisers' budgets is known as click fraud. Artificial intelligence methods have been used more and more to address complex problems in several fields of study, including cybersecurity, with surprising results. In order to determine if a user is a person or a bot, this article built a number of machine learning models and used a set of assessment measures to compare their performances. We made use of an actual collected dataset that described how people behaved when they were on websites. We retrieved a collection of user behavior-related characteristics, such as the quantity of webpages seen, the length of the browsing session, and the actions taken. All of the models under consideration produced good outcomes, according to the empirical data, with the random forest algorithm outperforming all other algorithms across all assessment measures.

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

https://www.scientific.net/AMR.859.586

The purpose of this research is to create an efficient method for identifying fake clicks in commissioners' click records. We wish to anticipate if a person is harmful or not based on the information they provide about their ad clicks. We construct our classification models for fraudulent click detection based on the training data. For the raw log data we provide above, we first generate and extract features. After assessing classifiers, we select models and construct an

ensemble model. Lastly, we used the actual dataset to test our model for human judge reference.

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

https://link.springer.com/chapter/10.1007/978-3-319-03844-5_38

Click fraud is a major drain on advertising budgets and may substantially undermine the profitability of the online advertising industry by taking money away from genuine partners, often known as publishers. Because of this, fraud detection algorithms that are able to recognize fraudulent activity by analyzing user click patterns are quite useful. We suggest a unique click fraud detection method based on the BuzzCity dataset. This method is built on a collection of new characteristics that are generated from existing properties. The resultant accuracy, recall, and area under the ROC curve are used to assess the suggested model. Based on six distinct learning methods, the final ensemble model demonstrated stability across all three performance metrics. Our finished model demonstrates its generalizability to other datasets by exhibiting improved outcomes on training, validation, and test datasets.

3. METHODOLOGY

a) Proposed Work:

In order to increase the precision and dependability of ad click fraud detection in intricate advertising contexts, the suggested system presents an attention-based Convolutional Neural Network (CNN) architecture. In order to minimize noise and improve learning efficiency, clickstream data is first preprocessed by encoding category properties, normalizing numerical features, and choosing pertinent features. While the integrated attention mechanism dynamically gives increased priority to important elements and temporal interactions that clearly suggest fraudulent activity, the CNN architecture is intended to extract significant spatial and behavioral patterns from the processed data. This makes it possible for the model to better capture sequential relationships and contextual relevance than traditional CNN-based methods.

The suggested model is implemented as a real-time fraud detection system with a Flask-based web interface to guarantee practical applicability. With the deployment, customers may stream or upload click data and get real-time predictions about the validity of ad clicks. Because it allows for constant monitoring and prompt response to fraudulent actions, this real-time connection improves usability for platform administrators and advertisers. A scalable, flexible, and effective solution that can manage big datasets and changing fraud trends in online advertising ecosystems is offered by the attention-enhanced deep learning architecture and web-based deployment.

b) System Architecture:

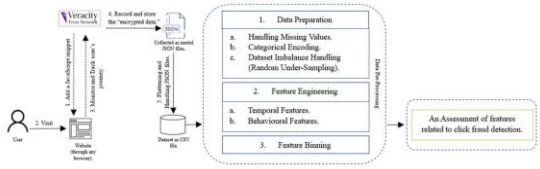


Fig 1 Proposed Architecture

The suggested extension's system architecture is intended to facilitate real-time fraud detection and attention-based deep learning. To guarantee consistent data representation, raw clickstream data is

first gathered from online advertising platforms and then routed via a preprocessing layer that handles missing values, encodes categorical properties, and normalizes numerical features. Only the most pertinent characteristics for fraud detection are kept after feature refining. After processing, the data is put into a Convolutional Neural Network (CNN), which uses click interactions to identify patterns in both space and behavior. To enable the model to concentrate on crucial click behaviors that point to fraudulent activity, an attention layer is added on top of the CNN feature maps. This layer assigns dynamic weights to significant features and temporal patterns.

A classification layer receives the attention-enhanced CNN outputs in the next step and classifies each click instance as either authentic or fraudulent. To facilitate real-time inference, the trained model is implemented within a web framework built with Flask. Users may rapidly acquire fraud prediction findings by using this tool to stream live click records or upload test data. Because of the architecture's scalability and adaptability, it can be seamlessly integrated with extensive advertising systems and updated continuously as fraud trends change. For reliable ad click fraud detection, our end-to-end architecture successfully combines cutting-edge attention-based deep learning with useful, real-world implementation.

c) 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 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.

d) 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.

4. EXPERIMENTAL RESULTS

A labeled ad click dataset was used to test multiple machine learning and deep learning models for fraudulent click detection. Managing missing values, label encoding categorical variables, normalizing numerical features, and selecting features using Recursive Feature Elimination preprocessed the dataset. The processed data was split 80:20 between training and testing sets. Several models, such as 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 including accuracy, precision, recall, and F1- Ensemble-based models capture complex feature interactions and manage data imbalance better than traditional classifiers, according to experiments.

Random Forest, Decision Tree, Gradient Boosting, and XGBoost all outperformed 98% in detecting fraudulent click patterns. The Recurrent Neural Network simulated sequential click activity with 97% accuracy, competing with deep learning models. The attention-enhanced Convolutional Neural Network performed best with 99.73% accuracy, demonstrating the importance of attention mechanisms for contextual feature weighting and sequential pattern recognition. These results show that feature selection and attention-based deep learning improve fraud detection, reduce false positives, and provide a viable and scalable solution for online advertising.

Good data preprocessing, feature selection, and an attention mechanism that emphasizes discriminative temporal and behavioral characteristics give the attention-based CNN 99.73% accuracy. The attention layer focuses on significant click patterns to reduce noise, while CNN-based hierarchical feature extraction improves generalization and fraud detection.

Accuracy: The capacity of a test to accurately distinguish between healthy and sick cases is a measure of its accuracy. We can determine a test's accuracy by calculating the proportion of reviewed

International Journal of Engineering Research and Science & Technology

cases with true positives and true negatives. In mathematical terms, this looks like:

Accuracy = TP + TN TP + TN + FP + FN.

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

F1-Score: One way to evaluate a model's performance in machine learning is via its F1 score. It takes a model's recall and precision and mixes them. Using the whole dataset, the accuracy metric determines the frequency with which a model produced accurate predictions.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

Precision: The proportion of occurrences or samples that are correctly classified out of all the ones that are classified as positive is known as precision. Consequently, the following is the formula for determining the accuracy:

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

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

Recall: The capacity of a model to detect all significant instances of a given class is measured by recall, a metric in machine learning. It sheds light on how well a model captures instances of a certain class and is calculated as the ratio of the total number of positive observations predicted to the number of actual positive observations.

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

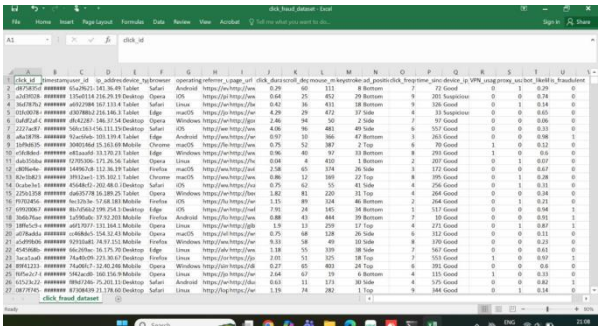


Fig 2: dataset

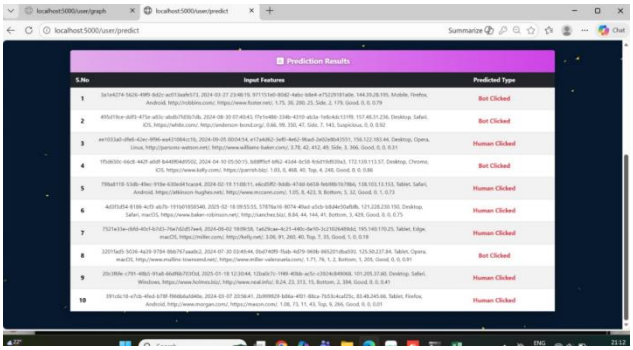


Figure 3 : Prediction Results

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.

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Logistic Regression	97.267	96.573	96.343	96.457
Decision Tree	98.933	98.044	99.277	98.639
Random Forest	98.333	97.010	98.888	97.939
KNN	73.200	48.649	49.902	43.215
ANN	67.267	53.533	52.713	52.569
Gradient	98.000	96.454	98.645	97.475

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Boosting				
LightGBM	98.533	97.349	99.006	98.137
XGBoost	98.667	97.579	99.097	98.304
Naïve Bayes	89.333	85.492	92.527	87.606
SVM	73.800	36.900	50.000	42.463
CNN	95.667	93.212	96.326	94.599
DNN	73.800	36.900	50.000	42.463
RNN	98.733	97.625	98.978	98.382
CNN with Attention (Proposed)	**99.7**	**99.61**	**99.87**	**99.74**

Table 1: Performance Comparison table

The attention-based CNN represents the proposed model and achieves the highest performance among all evaluated algorithms.

5. CONCLUSION

Through the integration of an attention-based Convolutional Neural Network with real-time deployment capabilities, this study offers an efficient extension to ad click fraud detection. Compared to traditional deep learning techniques, the attention mechanism improves the model's capacity to concentrate on important clickstream characteristics and contextual patterns, resulting in more accurate and trustworthy fraud detection. Additionally, by facilitating useful, real-time prediction, the Flask-based web interface helps to close the gap between high-performance deep learning models and practical advertising applications. All things considered, the suggested extension provides a scalable, flexible, and effective way to counteract changing ad click fraud in dynamic online advertising ecosystems.

6. FUTURE SCOPE

To further improve long-range dependency modeling in clickstream data, future developments of this system can concentrate on integrating sophisticated sequence-learning models, including Transformer architectures. The system may be expanded to include incremental and online learning, allowing the model to continually adjust in real time to newly discovered fraud tendencies. Furthermore, federated learning and anomaly detection approaches may be used to improve detection across various advertising platforms while preserving privacy. Cloud-based and edge computing technologies can increase deployment scalability, enabling the system to function effectively in expansive, heavily trafficked online advertising settings.

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.