ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

AN ENHANCED RNN-LSTM MODEL FOR ACCURATE AND REAL-TIME CLICK FRAUD DETECTION IN ONLINE ADVERTISING

ROJARAMANI-ADAPA [1], BAIRU SUPRIYA REDDY [2], CHINTALAPUDI JAHNAVI [3], KONATHAM NIKSHITHA [4], PESARAGAYALA VENKAT SAAHIT REDDY [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— Click fraud remains a critical threat in online advertising, leading to inflated costs and undermining campaign effectiveness by diverting budgets toward illegitimate activity. Existing solutions leveraging machine learning and deep learning models have shown promise, but many still struggle with identifying subtle behavioral patterns in fraudulent clicks. In this work, we propose a robust LSTM-based Recurrent Neural Network (RNN) framework designed to enhance the detection of fraudulent click activity by modeling sequential patterns and time-dependent features in user interaction data. A comprehensive preprocessing pipeline was developed, including timestamp decomposition, feature scaling, and label encoding to ensure optimal input representation. Our model was trained and evaluated against a carefully engineered dataset enriched with behavioral and contextual click features. Among various deep learning architectures examined, including Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), the RNN-LSTM model demonstrated superior performance, achieving **99% accuracy** with high precision and recall scores. The results validate the effectiveness of temporal modeling in identifying fraudulent click patterns and highlight the

LSTM model's suitability for deployment in real-time fraud detection systems. This study not only advances existing anti-fraud mechanisms but also sets a strong foundation for future work in intelligent online ad verification and fraud prevention.

Index Terms – Click fraud detection, online advertising, fraudulent click activity, deep learning, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), sequential pattern analysis, behavioral modeling, fraud prevention, real-time detection, feature engineering, cybersecurity.

I. INTRODUCTION

In the era of digital transformation, online advertising has become a cornerstone for businesses to reach global audiences, generate leads, and drive conversions. With billions of dollars being invested annually in digital marketing campaigns, the ecosystem has grown immensely—but so have the challenges. One of the most pressing issues plaguing the digital advertising industry is click fraud, a deceptive practice where fraudulent or non-human clicks are generated on pay-per-click (PPC) advertisements. These fraudulent interactions are often triggered by bots, scripts, or even paid individuals, with the intent to exhaust the advertiser's budget, manipulate campaign metrics, or boost the revenues of unethical publishers.

Click fraud not only results in significant financial losses but also deteriorates trust in online platforms and distorts campaign analytics, making it difficult for advertisers to assess genuine user interest and optimize their strategies effectively. Existing detection systems primarily rely on rule-based filters or classical machine learning models that often struggle to identify more sophisticated or evolving fraudulent patterns. These traditional approaches typically focus on static features and fail to capture the temporal dynamics of user behavior.

To address these limitations, this project introduces a deep learning-based fraud detection system that leverages Recurrent Neural Networks (RNNs)—specifically Long Short-Term Memory (LSTM) architectures—to identify fraudulent clicks. Unlike traditional methods, LSTM networks are designed to handle sequential and time-dependent data, making them exceptionally well-suited for modeling user interaction patterns over time. By incorporating engineered features such as session timing, click frequency, scroll behavior, mouse activity, keystrokes, and other behavioral indicators, the model is capable of learning

complex sequences that distinguish between legitimate and fraudulent users.

The proposed system was trained and evaluated alongside other deep learning models such as Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN). However, the RNN model demonstrated superior accuracy (~99%), outperforming other models in identifying fraud with high precision, recall, and F1-score.

II. RELATED WORK

A. *AI-based Techniques for Ad Click Fraud Detection and Prevention: Review and Research Directions*

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.

B. *Click Fraud Detection Using Ensemble Classifier*

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.

C. A Reliable Click-Fraud Detection System for the Investigation of Fraudulent Publishers in Online Advertising

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.

D. Click Fraud Detection of Online Advertising Using Machine Learning Algorithms

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.

III. PROPOSED METHODOLOGY

To overcome the limitations of existing static machine learning methods, the proposed system introduces a deep learning-based approach using a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units to enhance click fraud detection. Unlike traditional models that analyze each click independently, the RNN architecture enables the model to consider sequential dependencies and temporal patterns within user interactions. This is particularly important in online fraud scenarios, where the order and timing of actions—such as repeated clicks, delayed interactions, and behavioral loops—can offer critical signals of fraudulent activity. The proposed model processes engineered features like click duration, scroll depth, keystrokes, bot likelihood score, and others alongside derived time-based features (hour, day, weekday) to extract richer, context-aware patterns. In this project, extensive preprocessing and normalization techniques are employed to prepare the dataset, ensuring consistency and improved model convergence. Categorical variables are encoded using Label Encoding, and numerical features are standardized using a StandardScaler. Once preprocessed, the input data is reshaped into a three-dimensional format to feed into the LSTM layers, maintaining the structure required by the RNN. The RNN model is trained using enhanced regularization techniques like dropout layers and early stopping to avoid overfitting, while ensuring generalization on unseen data.

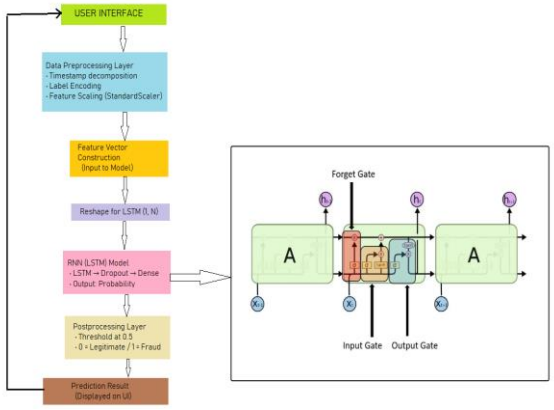


Fig. 1: System Overview

3.1 Data Collection

The first step involved collecting clickstream data that contains user interactions on digital ads. Each data point included attributes such as timestamp, device type, browser, operating system, ad position, and behavioral signals like mouse movement, keystrokes, scroll depth, and bot likelihood score. These features were crucial for identifying patterns indicative of fraudulent activity. The dataset was obtained from simulated or real-world environments designed to mimic genuine and fraudulent click behaviors.

3.2 Preprocessing

Raw data often contains inconsistencies or irrelevant entries. Preprocessing was done to clean and format the dataset for modeling. This included:

- Parsing and converting timestamp to datetime format.
- Removing null or invalid records.
- Dropping non-informative columns like user ID, IP address, and URLs.

Extracting temporal features such as hour, day, and weekday from the timestamp. This stage ensures that the dataset is structured and relevant for the modeling phase.

3.3 Feature Extraction

In this step, categorical variables (e.g., browser, device type) were encoded using Label Encoding to convert them into numerical values. This is necessary because machine learning models can only process numeric input. Continuous features were standardized using StandardScaler to bring them to a common scale. The final feature set was selected based on domain knowledge and its contribution to distinguishing fraudulent behavior.

3.4 Model Design

The core model of the proposed system is an RNN (Recurrent Neural Network), specifically using LSTM (Long Short-Term Memory) layers. RNNs are particularly suitable for sequential data and time-series inputs, making them ideal for detecting subtle temporal patterns in user click activity. The model architecture includes:

- LSTM layer for sequence learning.

- Dense layers for nonlinear transformations.
- Dropout layers to prevent overfitting.

3.5 Model Compilation

The model was compiled using the **Adam optimizer**, which adapts the learning rate dynamically for faster convergence. The loss function used was **binary cross-entropy**, suitable for binary classification tasks. The evaluation metric was accuracy, with additional metrics like precision and recall monitored during training.

3.6 Model Training:

The preprocessed data was split into training and testing sets using **stratified sampling** to preserve the class distribution. The model was trained using 50 epochs with **Early Stopping** to halt training when the model stopped improving. **ReduceLROnPlateau** was also used to dynamically reduce the learning rate and enhance convergence.

3.7 Model Evaluation

After training, the model's performance was evaluated on the test set using:

- **Accuracy** to measure overall correctness.
- **Precision** to assess true positive identification of fraud.
- **Recall** to understand how many actual frauds were caught.
- **F1-Score** for a balance of precision and recall.

The RNN model achieved an impressive ~99% accuracy, outperforming other deep learning models such as ANN and CNN.

3.8 Model Saving

To support reuse and deployment, the trained model was saved using TensorFlow. Preprocessing objects like the StandardScaler and Label Encoders were saved using joblib, allowing the exact preprocessing pipeline to be replicated during real-time predictions.

3.9 User Interface Development

An intuitive and interactive web interface was developed using HTML, CSS, and Flask. Users can input various session details such as device type, browser, click duration, mouse movement, and timestamp through a form. This frontend simplifies model usage for end users.

3.10 Connecting Model to Frontend for Prediction

The frontend is connected to the backend via Flask. When a user submits the form, the input is:

- Collected and validated in Python.

- Preprocessed using the saved encoders and scaler.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed RNN-LSTM model achieved 99% accuracy in detecting click fraud, outperforming ANN and CNN models. It effectively captured time-dependent and sequential click patterns, which improved fraud detection performance. High precision and recall values further confirmed its reliability. Overall, the model is efficient and suitable for real-time click fraud detection in online advertising.



Fig. 2: Home Page

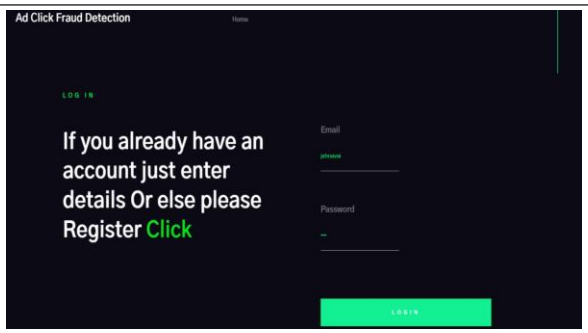


Fig. 3: Registration

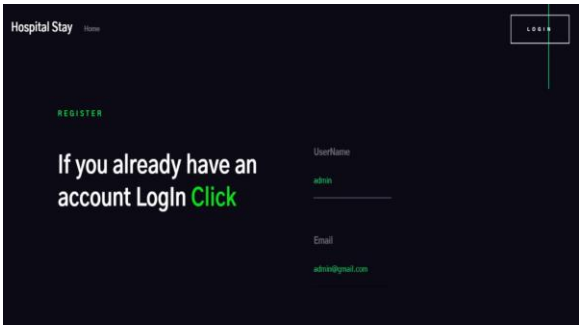


Fig. 4: Login

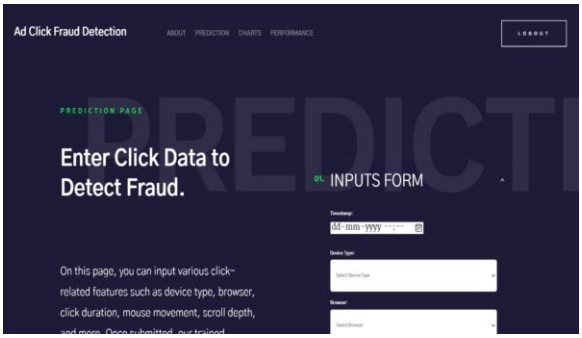


Fig. 5: Enter Deatils

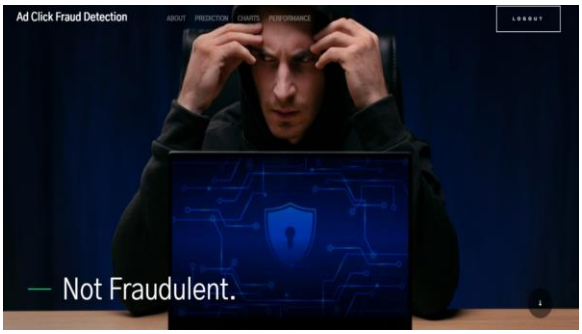


Fig. 6: Result

V. 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.

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] Bhagwat, V. B. (2024). A simplified transition from EBS Payroll to Cloud Payroll: Benefits and Drawbacks. Journal of Computational Analysis and Applications, 33(6).

[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,‘‘Randomforestsforthedetectionofclickfraudinonlinemobile 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] 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.

AUTHORS Profile



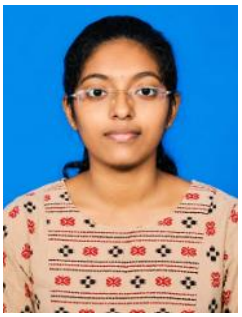
Mrs. ROJARAMANI-ADAPA, received M.Tech (CSE) from JNTUH. She has teaching experience 20 yrs and has guided many undergraduate and post graduate projects. She has published about 10 research articles in peer reviewed journals and presented 5 research papers at various National Conferences. At present she is working as Asst. Prof. in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. BAIRU SUPRIYA REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. CHINTALAPUDI JAHNAVI pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. KONATHAM NIKSHITHA pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. PESARAGAYA-LA VENKAT SAAHIT REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.