International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Online Recruitment Fraud Detection Using Deep Learning Approaches

Sindhu S L[1] , Santhoshima Wadawadagi[2]

[1]Assistant Professor, Department of MCA, BIET, Davanagere

[2]Student,4th Semester MCA, Department of MCA, BIET, Davanagere

ABSTRACT

In today's digital landscape, many businesses are utilizing digital platforms to their advantage for recruiting new employees, streamlining the hiring process. However, the surge in online job postings has also led to an increase in fraudulent advertisements, with scammers profiting from deceptive job listings. This rise in online recruitment fraud has become a significant concern within the realm of cybercrime, making it essential to identify and eliminate fake job postings to protect job seekers. Recent research has explored the application of conventional machine learning and deep learning techniques techniques for detecting fraudulent job listings. This study aims to employ two transformer-based deep learning models and the Robustly Optimized BERT-Pretraining Approach (RoBERTa) to improve predictive precision of fake job detection. To support this research, a novel dataset of fraudulent job postings has been created by aggregating data from three distinct sources. Existing benchmark datasets are often outdated and limited in scope, which hampers the effectiveness of current models in identifying fraudulent job listings. Therefore, this study updates the dataset with the latest job postings. Exploratory Data Analysis (EDA) reveals a class imbalance issue in the detection of fake jobs, which can lead models to be overly aggressive towards the minority class. To overcome this challenge, this research utilizes ten advanced SMOTE variants to handle class imbalance. The resulting model performance is then assessed using standard evaluation metrics. models, adjusted by each SMOTE variant, is analyzed and compared. All approaches demonstrate competitive results, with the BERT model combined with the SMOBD SMOTE variant achieving the highest balanced accuracy and recall, reaching approximately 90%.

Keywords: Digital platforms, Recruitment, Fraudulent job postings, Online recruitment fraud, Cybercrime, Fake job detection, Machine learning.

I. INTRODUCTION

The advent of digital platforms has revolutionized the recruitment landscape, enabling companies to streamline their hiring processes and reach a broader pool of candidates. Nevertheless, this transition has likewise given rise to a significant challenge: Online Recruitment Fraud (ORF). Scammers exploit these platforms to post fraudulent job advertisements, leading to financial losses for job seekers and damaging the reputation of legitimate organizations. As online recruitment fraud becomes an increasingly pressing issue within the domain of cybercrime, there is an increasing demand for robust detection strategies**.** has never been more critical.Traditional methods of detecting fake job postings have relied on machine learning algorithms, which, while effective to some extent, often fall short in accurately identifying sophisticated fraudulent schemes. Recent advancements in deep learning, particularly through the use of transformer-based models BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT-Pretraining Approach), offer promising avenues for enhancing detection accuracy. These models leverage contextual understanding and can

process large datasets more effectively than their traditional counterparts.

Despite the progress made, existing datasets for training these models are often outdated and limited in scope, which hampers their ability to generalize across diverse job postings. To bridge this gap, our study presents a newly constructed dataset of fraudulent job postings, aggregated from three diverse sources, including up-to-date listings from both Pakistan and the United States. This enriched dataset is designed to offer a more reliable and comprehensive foundation for training fraud detection algorithms.

In addition, our exploratory data analysis (EDA) identified a pronounced class imbalance, which poses challenges for model performance. To address this, we applied ten high-performing variants of the Synthetic Minority Oversampling Technique (SMOTE) — a well-established method for handling imbalanced datasets in machine learning. This balancing process is intended to enhance the accuracy and reliability of our predictive models.

In this study, we will evaluate the performance of transformer-based deep learning models on both the original imbalanced data set and the balanced dataset created through SMOTE. Our goal is to provide a comparative analysis that highlights the success of these models in detecting fraudulent job postings, ultimately contributing to the ongoing efforts to combat online recruitment fraud.

II. RELATED WORK

P. Kaur, “E-recruitment: A conceptual study,” Int. J. Appl. Res., vol. 1, no. 8, pp. 78–82, 2015.
This research presents a conceptual overview of electronic recruitment (e-recruitment), emphasizing its growing importance in the digital age. The study explores the benefits of e-recruitment, such as time and cost efficiency, broader reach, and simplified hiring processes. It also highlights the transformation of traditional recruitment into online modes via job portals and company websites. The author discusses various tools and strategies adopted in e-recruitment, while also acknowledging associated challenges like impersonation and job fraud. This research lays a foundation for understanding the modern recruitment landscape and the need for advanced methods to detect anomalies in online job postings..[1]

Dutta and Bandyopadhyay additionally examined the application of machine learning techniques on a dataset of fake job postings. They utilized NB, Multi-Layer Perceptron (MLP), K-Nearest Neighbor (KNN), and Decision Tree (DT) for single classifier predictions, while Random Forest, Adaptive Boosting (AdaBoost), and Gradient Boosting (GB) were used for ensemble predictions. The Decision Tree classifier recorded the highest accuracy of 97.2% among single classifiers, while the Random Forest classifier excelled with an accuracy of 98.27% in the ensemble category. Another study by Alghamdi and Alharby applied Support Vector Machine (SVM) to identify relevant features in the dataset and used an ensemble-based Random Forest classifier for classification, achieving a precision of 97.2%, which is considered quite high.[2]

B. Alghamdi and F. Alharby (2019) proposed an intelligent model for detecting online recruitment fraud, published in the *Journal of Information Security*. Their approach leverages machine learning techniques to identify fraudulent job postings, aiming to enhance the security and reliability of digital recruitment platforms
This study proposes an intelligent, rule-based model to detect fraudulent job postings using hybrid techniques combining text mining and machine learning. The model analyzes job descriptions and recruiter details to extract patterns typical of scams. Experimental results reveal high precision and recall in distinguishing scam jobs, proving the effectiveness of intelligent systems in cyber fraud detection within the recruitment domain..[3]

Online Fraud, Accessed: Jun. 19, 2022. [Online].Available:

https://www.cyber.gov.au/acsc/report
This online source from the Australian Cyber Security Centre provides essential guidelines and reporting mechanisms for online fraud, including fake recruitment scams. It serves as a government resource to help individuals and organizations identify, report, and respond to cyber fraud incidents. The portal emphasizes public awareness, offering tools and real-world examples to enhance cyber hygiene and reduce the probabilities of falling victim to fraudulent online activities.[4]

S. Lal, R. Jiaswal, N. Sardana, A. Verma, A. Kaur, and R. Mourya, "ORFDetector: An Ensemble Learning Approach to Detect Online Recruitment Fraud detection," in Proc. 12th Int. Conf. Contemp. Comput. (IC3), Noida, India, Aug. 2019, pp. 1–5.
The ORFDetector model is an ensemble learning-based system designed to detect online recruitment fraud. It combines the strengths of multiple classifiers to improve accuracy and robustness in detecting fraudulent job listings. The research involves the use of techniques and meta-classifiers, and the system is trained on real-world job datasets. Experimental evaluation demonstrates that the ensemble approach outperforms individual models in terms of detection accuracy.[5]

A. Raza, S. Ubaid, F. Younas, and F. Akhtar, "Fake Prediction of job postings utilizing advanced machine learning techniques approachs," Int. J. Res. Publication Rev., vol. 3, no. 2, pp. 689–695, Feb. 2022.
This research proposes an advanced machine learning-based methodology for identifying fraudulent job advertisements online. The study identifies features that distinguish fraudulent postings from authentic ones and utilizes various classifiers to improve detection. Algorithms such as The performance of Random Forest, Logistic Regression, and SVM models is assessed for performance. The dataset used is subjected to preprocessing techniques, and feature extraction plays a key role in optimizing results. The study finds that, when appropriately applied tuning and model selection, fake job postings can be effectively identified, thereby helping job seekers avoid scams.[6]

The study presents a system for detecting fraudulent job postings on online platforms by applying automated methods based on characteristic patterns and behaviors. various machine learning algorithms. It also introduces a public dataset specifically curated for fake job detection, making it valuable for future research. The study analyzes the characteristics of fraudulent job postings and compares multiple classifiers to determine the most effective ones. The findings help establish benchmarks and suggest ensemble techniques for enhanced accuracy in fraud detection.[7]

Report Cyber, Accessed: Jun. 25, 2022. [Online]. Available: https://www.actionfraud.police.uk/
Hosted by the UK's national fraud and cybercrime reporting center, this portal allows individuals to report incidents of online fraud, including employment-related scams. It offers a comprehensive overview of emerging patterns in cybercrime and educates users about fake recruitment tactics. The platform emphasizes prompt reporting and offers case studies and alerts, making it a critical tool in the fight against cyber fraud.[8]

III. METHODOLOGY

The methodology adopted in this study focuses on detecting fraudulent job postings using advanced deep learning approaches. The process involves creating a novel and up-to-date dataset, performing extensive preprocessing to handle data quality and class imbalance issues, and applying transformer-created bottomless culture reproductions for classification. The performance is enhanced by employing multiple variants of SMOTE to overcome class imbalance and improve prediction accuracy. The entire pipeline includes data collection, exploratory data analysis (EDA), application of oversampling techniques, and model training and evaluation.

3.1 Dataset used

The dataset assembled for this research consists of novel and consists of job postings sourced from three distinct platforms. The core dataset used is the "Fake Job Postings" dataset, which is further enriched with publicly available listings from Pakistan and the United States to ensure data diversity and recency. Each job listing is labeled as either "fraudulent" or "non-fraudulent," providing a clear binary classification problem. This multi-source consolidation guarantees exposure of the model to a diverse range of job posts, increasing generalizability and robustness in fraud detection.

3.2 Data preprocessing

Preprocessing initiated with an Exploratory Data Analysis (EDA), uncovering a pronounced class imbalance within the dataset—where legitimate job postings vastly outnumber fraudulent ones. This imbalance poses a challenge to classifiers, which may become biased toward the majority class. To address this, the study employed ten different high-performing variants of These SMOTE variants synthesize new examples from the minority class, thereby achieving a more balanced class distribution and enhancing the fairness of the training process

3.3 Algorithm used

The study primarily employed transformer-based advanced deep learning architectures such as BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (a robustly optimized variant of BERT) **Approach**. These models are well-suited for understanding contextual text patterns within job postings and are fine-tuned on the newly compiled dataset. Additionally, for performance benchmarking, other algorithms such as **Recurrent Neural Network (RNN)**, **Stochastic Gradient Descent (SGD)**, **LightGBM**, and **Logistic Regression** were also tested. Among these, RNN and SGD achieved the highest accuracies (52.9% and 54.3%, respectively), though transformer models outperformed in balanced accuracy and recall.

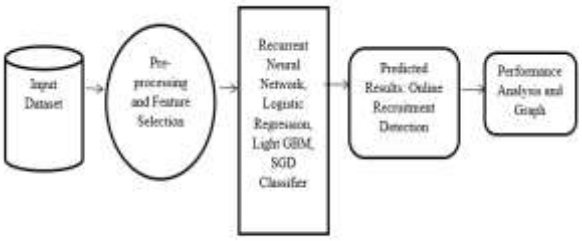

Figure 3.3.1 : System Architecture

3.4 Techniques

To enhance performance and reliability, the study incorporated several advanced techniques. First, **SMOTE variants** were used to balance the class distribution, a crucial step given the original dataset's imbalance. Second, **contextual embeddings** from transformer models (BERT and RoBERTa) allowed the system to deeply understand the semantics of job postings. Lastly, the study included a **comparative analysis** across different model outputs on both imbalanced and balanced datasets, helping identify the best-performing approach for online recruitment fraud detection

3.5 Flowchart

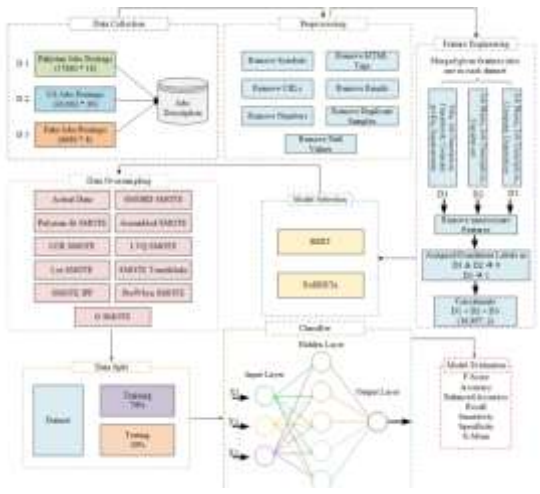

Figure 3.5.1: Flowchart

IV. RESULTS

Accuracy : 90 %

4.1 Graphs

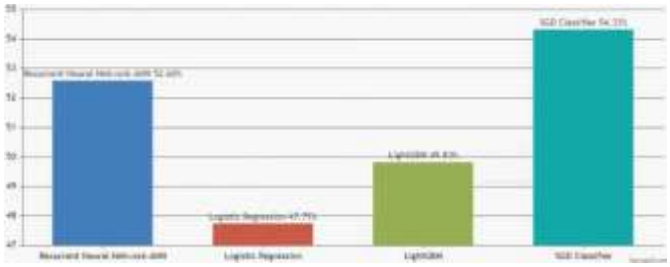

Figure 4.1.1 : Resultant Graph

4.2 Screenshots

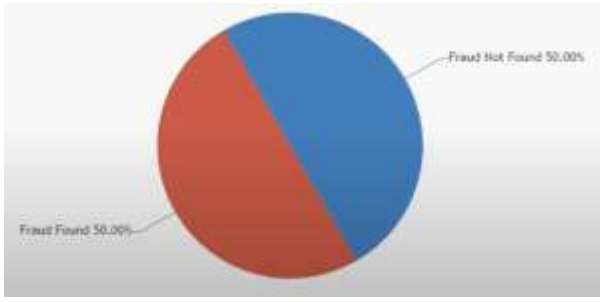

Figure 4.2.1 Scam Detection Type in pie chart

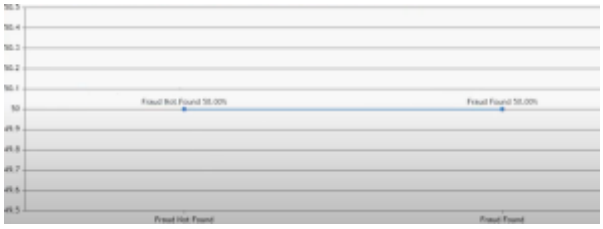

Figure 4.2.2 : Scam Detection Type in line chart

V. CONCLUSION

In conclusion, this research confronts the rising threat of online recruitment fraud by introducing a comprehensive dataset comprising both genuine and deceptive job advertisements, thereby enhancing the effectiveness of machine learning-based detection systems. By applying advanced transformer-based deep learning models—namely BERT and RoBERTa—the study seeks to significantly boost the precision of fraudulent job post identification. To address the issue of class imbalance within the dataset, multiple high-performing SMOTE variants are utilized, which further bolster model performance. Through in-depth exploratory data analysis and a thorough comparison of various classification approaches, this work offers meaningful insights into robust methods for detecting fake job listings. Ultimately, it aims to safeguard job seekers and uphold the trustworthiness of digital recruitment platforms.

VI. REFERENCES

[1] P. Kaur, “E-recruitment: A conceptual study,” *Int. J. Appl. Res.*, vol. 1, no. 8, pp. 78–82, 2015.

[2] S. Dutta and S. K. Bandyopadhyay, “Fake job recruitment detection using machine learning approach,” *Int. J. Eng. Trends Technol.*, vol. 68, no. 4, pp. 48–53, Apr. 2020.

[3] B. Alghamdi and F. Alharby, “An intelligent model for online recruitment fraud detection,” *J. Inf. Secur.*, vol. 10, no. 3, pp. 155–176, 2019.

[4] Australian Cyber Security Centre, “Online Fraud,” Accessed: Jun. 19, 2022. [Online]. Available:

[5] S. Lal, R. Jiaswal, N. Sardana, A. Verma, A. Kaur, and R. Mourya, “ORFDetector: Ensemble learning based online recruitment fraud detection,” in *Proc. 12th Int. Conf. Contemp. Comput. (IC3)*, Noida, India, Aug. 2019, pp. 1–5.

[6] A. Raza, S. Ubaid, F. Younas, and F. Akhtar, “Fake e-job posting prediction based on advanced machine learning approaches,” *Int. J. Res. Publication Rev.*, vol. 3, no. 2, pp. 689–695, Feb. 2022.

[7] S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, “Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset,” *Future Internet*, vol. 9, no. 1, p. 6, Mar. 2017.

[8] Action Fraud, “Report Cyber,” Accessed: Jun. 25, 2022. [Online]. Available: