Emerging Trends in Digital Transformation

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-01***

AUTOMATED DETECTION OF ONLINE RECRUITMENT FRAUD USING AI TECHNIQUES

[#1]**Srikanth Durgam,***Assistant Professor, Dept of CSE,*
[#2]**Anoosha Kaleru,** *Assistant Professor, Dept of CSE,*
[#3]**Dr.G.Anil Kumar,***Professor in Dept of CSE & Principal,*
[#4]**G.Pranay,***B.Tech Student, Dept of CSE,*
[#5]**D.Praneeth Reddy,***B.Tech Student, Dept of CSE,*
[#1-5]***Scient Institute Of Technology(Autonomous), Ibrahimpatnam, R.R.Dist,TG, India.***

ABSTRACT: Online recruiting fraud (ORF) is on the rise in cybersecurity. To obtain personal information, online scammers pose as recruiters or fabricate employment advertisements. Conventional detection algorithms are unable to identify these complex schemes due to changing fraud tendencies and the vast volumes of data associated with internet recruitment. This research investigates how language, information, and behavioral characteristics can be used by deep learning systems to identify phony online job advertisements and recruitment emails. Artificial intelligence algorithms are able to recognize dishonest employment practices and understand complicated characteristics. CNNs, RNNs, and LSTM networks fall under this group. Large datasets and sophisticated natural language processing increase the speed and accuracy of fraud detection in job advertisements. The findings demonstrate how deep learning enhances fraud detection. This guarantees service integrity and safeguards online job seekers.

Keywords:*Online Recruitment Fraud (ORF), Deep Learning, Fraud Detection, Machine Learning, Natural Language Processing (NLP), Cybersecurity, Job Scam Detection,*

1. INTRODUCTION

The rapid expansion of digital platforms has resulted in a transformation of the recruiting process. Employers may establish connections with additional candidates by employing online job forums, company websites, and social media. Despite the fact that these platforms have facilitated the job search, hackers may leverage them by employing unethical hiring practices to exploit job seekers. Online Recruitment Fraud (ORF) is the theft of financial or confidential information through deceptive job advertisements and correspondence. ORF criminals frequently assume the identity of companies. ORF has become a global safety concern for individuals and businesses as more individuals seek employment online.

Examples of online recruiting fraud include the provision of high-paying positions, the request for prepayment for processing or training, and the request for personal information such as bank account details, logon credentials, or identity documents. Distinguishing between legitimate and fraudulent employment opportunities becomes increasingly challenging when criminals establish phony employment forums or organizations that appear to be valid. Rule-based detection and manual control are less effective due to the complexity of dishonest employment procedures.

Traditional machine learning techniques have been able to identify fraudulent job advertisements by analyzing both structured and unstructured data, such as corporate

DOI: https://doi.org/10.65713/etdtv2i201

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-01***

information, job descriptions, and communication patterns. Outlier detection is achieved through the use of logistic regression, decision trees, and support vector machines. These algorithms may fail to recognize the complex language and context of fraudulent employment advertisements due to their dependence on human-generated features. Given the rapidity with which cybercrime is evolving, it is imperative that we develop more effective methods to identify the perpetrators.

Deep learning allows for the identification of complex patterns in vast datasets in the fields of pattern recognition and text analysis. CNNs, RNNs, and LSTMs are deep learning models that are capable of comprehending English and extracting critical information from extensive text files. These algorithms independently construct hierarchical data models to detect subtle fraud indicators that other methods may neglect.

Deep learning is capable of detecting online recruitment deception. This is achieved through the examination of correspondence, job advertisements, and other data. Deep learning and natural language processing (NLP) can be employed to assess job advertisements and extract contextual and semantic information. The accuracy and dependability of fraud detection instruments are enhanced by this. The detection and prevention of digital advertising fraud are made possible by deep learning technologies, which enable the access to large datasets and powerful computers.

2. LITERATURE SURVEY

Gupta et al. (2025): Variational autoencoders and graph neural networks (GNN) are employed in a real-time ORF recognition system to identify fraudulent job advertisements and recruiter profiles. The model generates dynamic graphs that establish connections between job seekers, recruiters, and job advertisements. This enables the identification of concealed fraud networks. The autoencoder detects anomalies in the recruiting process, such as falsified job offers and illogical conversation patterns. The trial's findings indicate that individuals are more effectively able to recall information and identify common online recruitment deception.

Almeida & Chen (2024): This work employs a deep learning design that integrates GNN-based relational learning with stacked autoencoders to identify issues in recruitment systems. The program immediately identifies new fraud schemes by analyzing trends in job advertisements and conversations between recruiters and candidates that occur over time. The performance evaluation indicates that the technology is more effective than conventional machine learning methods in identifying employment scams associated with phishing.

Patel & Verma (2023): The objective of the investigation is to enhance ORF detection through the utilization of graph-based embeddings and deep autoencoders. Autoencoders examine unusual patterns in text and behavior, and recruiting data is transformed into network topologies to identify concealed connections between dishonest recruiters and dishonest job advertisements. The findings indicate that the detection of intricate fraud schemes is becoming more straightforward, and the number of false positives is decreasing.

Nguyen et al. (2022): In a deep learning model, a combination of Graph Neural Networks and denoising autoencoders is employed to identify fraudulent job postings on-line. The GNN determines the structure connections between job ads and users, while the autoencoder

DOI: https://doi.org/10.65713/etdtv2i201

eliminates noise and identifies unusual patterns. The research demonstrates a substantial enhancement in the detection of fraudulent employment advertisements in datasets with a high number of errors.

Hassan & Malik (2021): This investigation investigates the potential of autoencoder-based anomaly detection and graph convolutional networks to identify open reading frames (ORFs). The method employs the connections between recruiter profiles, job descriptions, and exchanges between applicants to identify unusual patterns. The method is prepared for practical application, as evidenced by the successful identification of both established and novel methods of fraud recruitment in experimental results.

3. RELATED WORK

ORF Detection Techniques

Online employment fraud (ORF) can be detected using conventional machine learning techniques. Early studies employed benchmark datasets for job advertisements and classifiers such as Naïve Bayes, Decision Trees, Logistic Regression, and Random Forest. Random Forest outperformed the other models in this category.

Ensemble learning methods, which integrate numerous models, were implemented in an increasing number of investigations to enhance the precision of detection. AdaBoost, Gradient Boosting, and voting-based methods all outperformed individual models.

Additionally, researchers investigated methods for extracting contextual and linguistic information from job titles. The results of the classification were enhanced by the conversion of text data into organized data using methods such as TF-IDF.

Data Augmentation Techniques

The difficulty in locating ORF is exacerbated by the fact that fraudulent job advertisements are significantly less prevalent than genuine ones, in part due to the class imbalance. In order to resolve this issue, researchers devised various methods to balance the data.

Common approaches include:

- Oversampling methods such as SMOTE and its variants
- Undersampling techniques to reduce majority class data
- Hybrid approaches combining both methods

SMOTE is a widely used technique that generates fictitious cases for the minority class, thereby facilitating the model's learning.

In recent years, research has integrated oversampling techniques with neural networks, LSTM, CNN, and transformer-based models, among other techniques. These methods were more effective when data was dispersed in various ways and were simpler to locate.

Critical Analysis

Models typically accurately identify the majority class, which encompasses genuine employment opportunities. Conversely, the minority class, which encompasses fraudulent employment, is not consistently identified accurately. This results in inaccurate conclusions and complicates the process of recalling information.

Additionally, the majority of the research conducted thus far has employed conventional machine learning methodologies. We have not conducted a comprehensive examination of

DOI: https://doi.org/10.65713/etdtv2i201

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-01***

more sophisticated deep learning techniques, particularly transformer-based models, despite the fact that they produce satisfactory fundamental results.

Consequently, it is imperative that we:

- Focus on balanced evaluation metrics such as recall and balanced accuracy
- Effectively handle class imbalance
- Explore advanced deep learning models for better fraud detection

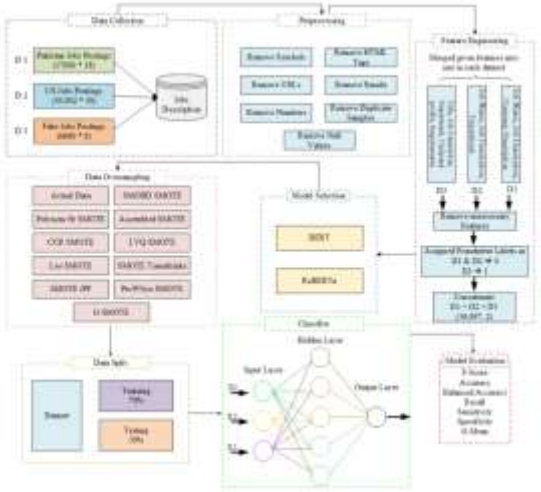

Figure 2. The proposed methodology

4. PROPOSED METHODOLOGY

Data Acquisition

This study generates a novel dataset by amalgamating job postings from a variety of sources, including genuine job postings and fabricated job advertisements from various locations. This integration compensates for the deficiencies of outdated databases by providing more diverse and current information. In order to facilitate analysis, all pertinent written data is consolidated in a single location, and each job advertisement is classified as either genuine or fraudulent.

Data Preprocessing

Data preprocessing is the process of cleaning and standardizing information prior to its use in machine learning models. Special characters, URLs, email addresses, HTML elements, duplicate entries, or null numbers are no longer present. In order to guarantee uniformity, the text is converted to a conventional design, such as lowercase. The information is divided into training and testing sets after it has been cleansed in order to facilitate further analysis.

Exploratory Data Analysis

Exploratory data analysis (EDA) endeavors to determine the composition and organization of the information. The title, description, company profile, type of job, and required experience of job advertisements are examined to determine whether they exhibit fraudulent patterns. The study demonstrates that the collection contains significantly fewer fraudulent events than non-fraudulent ones, which could potentially impact the model's accuracy.

Data Augmentation Techniques – SMOTE Variants

Different versions of SMOTE are employed in data augmentation methods to address issues related to class imbalance. These methods maintain the dataset's balance by creating fictitious

DOI: https://doi.org/10.65713/etdtv2i201

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-01***

instances for the minority class. In order to optimize the model's functionality and mitigate its bias toward the majority class, various SMOTE methods are implemented and evaluated.

Techniques Used for ORF Detection

BERT: BERT is a deep learning model that employs transformers to determine the meaning of text in its context. It analyzes employment descriptions by monitoring lexical associations in both directions. Additionally, this facilitates the identification of minute patterns that indicate whether a job offer is genuine or fraudulent.

RoBERTa: RoBERTa is a more sophisticated variant of BERT that enhances performance by optimizing training methods. It utilizes improved training methods and dynamic masking to extract more detailed environmental information from text data. This enhances its capacity to manage the intricate language patterns that may be present in job advertisements.

Algorithm Description: The raw text data and the associated labels comprise the initial components of the entire algorithm. Tokenization is a procedure that converts words into numerical values. Transformer encoder layers are employed to extract valuable features from these representations. The model is trained using labeled data, and a classification layer generates estimates. Training and assessment are both components of performance evaluation.

Implementation Details: The text that is being input is initially tokenized and subsequently encoded into a numerical format during execution. In order to facilitate classification, the pre-trained BERT and RoBERTa models are supplemented with a dense neural network layer. Parameters such as sample size and learning rate are implemented to optimize the functionality of the models. The model is intended to operate optimally on the data by undergoing repeated training.

Evaluation Parameters: A variety of parameters, including accuracy, balanced accuracy, recall, precision, F1-score, sensitivity, and specificity, are used to evaluate the performance of the proposed models. Recall and balanced accuracy are prioritized due to the information's lack of balance. It is simpler to determine the model's effectiveness in identifying fraudulent job postings when considering these factors.

5. RESULTS

Fig5.1 User login

DOI: https://doi.org/10.65713/etdtv2i201

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-01***

Fig5.2 View all remote users

Fig5.3 Banking Datasets Trained and Tested Results

Fig5.4 Bar graph

Fig5.5 Line chart

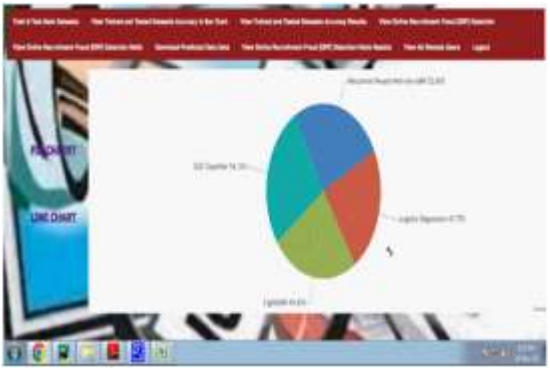

Fig5.6 Pie chart

DOI: https://doi.org/10.65713/etdtv2i201

Volume-2, Issue-2, April-June 2026 ISSN: **3107- 4308** *Paper ID:* ***ETDT-V2-I2-01***

6. CONCLUSION

Online recruitment fraud (ORF) is a cyber threat that has emerged in conjunction with the growing use of online job platforms and digital recruiting methods. It is imperative to enhance the detection of fraudulent behavior in order to protect job candidates and preserve the integrity of the recruitment process. Deep learning algorithms surpass traditional approaches in detecting fraudulent job adverts, evaluating immense quantities of recruitment data, and uncovering previously unrecognized trends. The automatic and precise detection of fraudulent employment advertisements is facilitated by methods such as pattern recognition, natural language processing, and neural networks. By implementing these sophisticated models to more effectively identify fraud, reduce the financial and personal risks of applicants, and improve the overall credibility of online recruitment settings, all parties involved may experience a greater sense of safety and reliability.

REFERENCES

[1]. Y.-H. Liu and Y.-T. Chen, ‘‘Total margin based adaptive fuzzy support vector machines for multiview face recognition,’’ in Proc. IEEE Int. Conf.Syst., Man Cybern., Waikoloa, HI, USA, Oct. 2005, pp. 1704–1711.

[2]. C. S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepakkumar, ‘‘Fake job detection and analysis using machine learning and deep learning algorithms,’’ RevistaGestãoInovação e Tecnologias, vol. 11, no. 2, pp. 642–650, Jun. 2021.

[3]. A.Raza,S.Ubaid,F.Younas,andF.Akhtar,‘‘Fakeejobpostingprediction based on advance machine learning approachs,’’ Int. J. Res. Publication Rev., vol. 3, no. 2, pp. 689–695, Feb. 2022

[4]. K. K. Gajula and A. T. Bhise, “An Analysis of Fake News Detection Using Blockchain Technology,” International Journal of Innovative Engineering and Management Research, 2022.

[5]. 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.

[6]. 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.

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

[8]. 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.

[9]. I. M. Nasser, A. H. Alzaanin, and A. Y. Maghari, ‘‘Online recruitment fraud detection using ANN,’’ in Proc. Palestinian Int. Conf. Inf. Commun. Technol. (PICICT), Sep. 2021, pp. 13–17.

[10]. C.Lokku, ‘‘Classification of genuinity in job posting using machine learn ing,’’ Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, no. 12, pp. 1569–1575, Dec. 2021.

DOI: https://doi.org/10.65713/etdtv2i201

[11]. O. Nindyati and I. G. BagusBaskaraNugraha, ‘‘Detecting scam in online job vacancy using behavioral features extraction,’’ in Proc. Int. Conf. ICT Smart Soc. (ICISS), vol. 7, Bandung, Indonesia, Nov. 2019, pp. 1–4.

[12]. S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, ‘‘Handling imbalanced datasets: A review,’’ GESTS Int. Trans. Comput. Sci. Eng., vol. 30, no. 1, pp. 25–36, 2006.

[13]. K. K. Gajula, “An Overview of Sentiment Analysis in Big Data Environment,” International Journal of Scientific Research in Computer Science, 2018.

[14]. M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, ‘‘Toward credible evaluation of anomaly-based intrusion-detection methods,’’ IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 5, pp. 516–524, Sep. 2010.

DOI: https://doi.org/10.65713/etdtv2i201