ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Advanced Deep Learning for Online Recruitment Fraud Detection: A Transformer-Based Approach

Parimi Arun Kumar
21N81A6218
Computer Science and Engineering (Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
arunkumarparimi2004@gmail.com

Komuravelli Akash
21N81A6224
Computer Science and Engineering (Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
akashkomuravelli143@gmail.com

Moghli Dhruva Kumar
21N81A6226
Computer Science and Engineering (Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
dhruvakumar669@gmail.com

Sriram Vamshi
21N81A6229
Computer Science and Engineering (Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
kumarsriram99@gmail.com

Mr. B.Surya Narayana Reddy
Assistant Professor
Computer Science and Engineering (Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510

ABSTRACT: The increasing use of online platforms for recruitment has made hiring more efficient but has also led to a surge in online recruitment fraud (ORF). Fraudsters exploit job seekers through deceptive job postings, posing a serious cybersecurity threat. This study addresses the detection of fake job postings using transformer-based deep learning models, specifically BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining Approach). These models are employed to accurately classify job postings as genuine or fraudulent. To support this, a novel dataset was developed by combining job listings from three diverse sources, overcoming the limitations of existing outdated benchmark datasets. Exploratory Data Analysis (EDA) revealed a strong class imbalance, with fake postings significantly underrepresented. This imbalance can bias prediction results and reduce model effectiveness. To mitigate this, the research integrates ten variants of the

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Synthetic Minority Oversampling Technique (SMOTE), enhancing the models' ability to generalize across both classes. Comparative analysis of model performance across different SMOTE variants showed promising results. Among all combinations, the BERT model integrated with the SMOBD SMOTE variant achieved the highest balanced accuracy and recall of around 90%. These findings demonstrate the potential of deep learning in effectively identifying and reducing online recruitment fraud.

Keywords – BERT, RoBERTa, SMOTE, Deep Learning, Fake Job Detection, Transformer Models, Imbalanced Classification.

INTRODUCTION

The increase in digital job portals has transformed how companies hire and individuals search for employment. However, this evolution has also given rise to a new form of cybercrime—online recruitment fraud (ORF). Scammers post fraudulent job openings to deceive applicants, leading to significant financial and personal losses.

This project introduces a deep learning-powered solution for detecting such scams. The system leverages transformer-based language models—BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining Approach)—to process and classify job postings. These models excel in capturing contextual and semantic relationships within text, making them well-suited for this task.

A unique aspect of this system is the curated dataset, formed by combining data from three different sources, including international job portals. Exploratory Data Analysis (EDA) revealed a significant class imbalance problem: far fewer fraudulent postings than genuine ones. To counteract this, the system implements ten advanced SMOTE (Synthetic Minority Oversampling Technique) variants to rebalance the data, enhancing model reliability.

Through its design, this tool aims to protect job seekers by accurately identifying fraudulent job advertisements, reducing the chances of being misled. It provides an effective solution for detecting ORFs using advanced AI techniques, ultimately ensuring a safer job-hunting experience.

1. LITERATURE REVIEW

‘‘E-recruitment: A conceptual study,’’

E-Recruitment or online recruitment is the process of recruiting personnel through online that helped the organizations to reach large number of workforce and to identify the skilled personnel easily with the use of technology and web based resources Lakshmi S.L (2013). After Covid 19 many companies are streamlining their recruitment and selection process by including technology like video-conferencing, mobile applications, chat bots, internet and computer-based assessments etc to improve their recruitment process by which candidates can be matched with live vacancies. Several studies shows that recruiters and companies are increasingly using online social networking to attract and screen candidates as part of the hiring process Ollington Nickolas et al. (2013). The study shows that e-recruitment with the help of technology have enhanced the organizations to make their recruitment process more easy and effective and at the same time save their valuable time and money.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

The study attempts to understand about E-Recruitment its practices, benefits and challenges.

‘‘Fake job detection and analysis using machine learning and deep learning algorithms,’’

With the pandemic situation, there is a strong rise in the number of online jobs posted on the internet in various job portals. But some of the jobs being posted online are actually fake jobs which lead to a theft of personal information and vital information. Thus, these fake jobs can be precisely detected and classified from a pool of job posts of both fake and real jobs by using advanced deep learning as well as machine learning classification algorithms. In this paper, machine learning and deep learning algorithms are used so as to detect fake jobs and to differentiate them from real jobs. The data analysis part and data cleaning part are also proposed in this paper, so that the classification algorithm applied is highly precise and accurate. It has to be noted that the data cleaning step is a very important step in machine learning project because it actually determines the accuracy of the machine learning as well as deep learning algorithms. Hence a great importance is emphasized on data cleaning and pre-processing step in this paper. The classification and detection of fake jobs can be done with high accuracy and high precision. Hence the machine learning and deep learning algorithms have to be applied on cleaned and pre-processed data in order to achieve a better accuracy. Further, deep learning neural networks are used so as to achieve higher accuracy. Finally all these classification models are compared with each other to find the classification algorithm with highest accuracy and precision.

‘‘Fake e job posting prediction based on advance machine learning approachs,’’

There are many jobs adverts on the internet, even on reputable job posting sites, that never appear to be false. However, following the selection , the so-called recruiters begin to seek money and bank information. Many candidates fall into their traps and lose a lot of money as well as their existing job. As a result, it is preferable to determine whether a job posting submitted on the site is genuine or fraudulent. Manually identifying it is extremely difficult, if not impossible! An automated online tool (website) based on machine learning-based categorization and algorithms are presented to eliminate fraudulent job postings on the internet. It aids in the detection of bogus job postings among the vast number of postings on the internet.

‘‘Survey: More millennials than seniors victims of job scams,’’

Mass marketing scams are some of the most common frauds in America, and include scams perpetrated through the mail. A growing body of research indicates that older adults face a greater risk of victimization due to age-related changes in cognitive functioning and social isolation, and may be more likely to fall victim repeatedly. The aim of this study is to determine the frequency of repeat mass marketing fraud (revictimization) among older adults and patterns of victimization associated with age, scam type, seasonality, and geography. We use two decades of non-public administrative data from the United States Postal Inspection Service (USPIS). These databases were seized during law enforcement investigations of mass mailing scam organizations and contain more than 2 million unique U.S. victims and their transactions with four different fraud organizations. Victims were matched across datasets using name, address, and a change of address file. We

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

find that revictimization rates increase with age in psychic scams. The 10,000 victims who responded the most times (between 82 and 562 times) were 78 years old on average and suffered $4,700 in total losses per person. Other significant trends emerged for lottery and sweepstakes scams. Unlike prior fraud victimization studies, inferences on victim characteristics are based on actual victim experiences with fraud rather than hypothetical scenarios or surveys where victims must self-report fraud. Findings provide valuable policy-relevant information regarding older victims and the patterns of chronic victimization.

2. METHODOLOGY

The Online Recruitment Fraud (ORF) Detection System is engineered to detect fraudulent job postings using a deep learning framework built around transformer models and advanced class balancing techniques. The following methodology outlines the data preparation, feature representation, oversampling strategy, model architecture, and evaluation process used to develop and validate the system.

---

1. Dataset Integration and Preparation

The dataset used in this study was compiled by merging three separate job posting datasets:

- A labeled Fake Job Postings dataset.
- A Pakistani job portal dataset with verified genuine postings.
- A U.S.-based job listing dataset offering international job context.

These datasets were preprocessed to unify column formats and remove non-relevant fields. Non-English records, null values, and duplicates were eliminated. The final dataset was assigned binary labels—**1 for fraudulent** and **0 for legitimate**—to support binary classification tasks.

---

2. Exploratory Data Analysis (EDA)

EDA was carried out to better understand the characteristics of the dataset. It revealed the following key observations:

- **Imbalanced Class Distribution**: Genuine job postings significantly outnumbered fraudulent ones.
- **Linguistic Patterns**: Certain keywords like “quick money,” “easy work from home,” and “data entry” were frequently associated with fraudulent postings.
- **Text Length and Structure**: Fraudulent postings often had shorter and more vague descriptions compared to legitimate ads.

These insights helped refine feature selection and underscored the need for robust class imbalance handling before model training.

---

3. Handling Class Imbalance with SMOTE Variants

To overcome the severe class imbalance problem, the system employed ten SMOTE-based oversampling strategies, applied only to the training dataset:

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

- **SMOTE** – the standard interpolation-based method.
- **Borderline-SMOTE** – focuses on generating samples near class decision boundaries.
- **SVM-SMOTE** – uses support vectors to guide new instance generation.
- **ADASYN** – emphasizes harder-to-classify minority samples.
- **KMeans-SMOTE** – uses clustering to enhance sampling quality.
- **SMOTE-NC** – tailored for datasets with both numeric and categorical features.
- **SMOTE-Tomek Links** – combines oversampling with noise reduction.
- **SMOTE-ENC** – designed for encoding mixed-type data.
- **SMOTE-BD (Borderline Dense)** – targets densely clustered border samples.
- **SMOTE-Boost** – integrates SMOTE with boosting techniques.

Each method was evaluated to determine which variant best improved minority class recall and overall classification performance.

---

4. Feature Representation Using Transformer Embeddings

Unlike conventional text encoding techniques like Bag-of-Words or TF-IDF, this system employed transformer-based embeddings to capture the contextual meaning of job descriptions.

Two powerful pre-trained models were used:

- **BERT (Bidirectional Encoder Representations from Transformers)**
- **RoBERTa (Robustly Optimized BERT Pretraining Approach)**

These models convert each job posting into a high-dimensional vector embedding that encapsulates semantic relationships and linguistic context. Fine-tuning was performed on the labeled and SMOTE-balanced datasets to improve classification accuracy.

---

5. Model Training and Fine-Tuning

Separate models were trained for BERT and RoBERTa using the rebalanced datasets generated by each SMOTE variant. The models were fine-tuned using classification heads with the following settings:

- Optimizer: AdamW
- Learning Rate: 2e-5
- Epochs: 3 to 5 (determined via validation)
- Batch Size: 16
- Loss Function: CrossEntropyLoss

Cross-validation was used to assess stability and prevent overfitting. The objective was to maximize both **balanced accuracy** and **recall**, particularly for the minority (fraudulent) class.

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

6. Performance Evaluation

To assess the system's effectiveness, the following metrics were used:

- **Accuracy** – overall correctness of classification.
- **Balanced Accuracy** – average of recall on both classes, critical for imbalanced data.
- **Recall** – sensitivity to detecting fraudulent postings.
- **Precision and F1-Score** – to evaluate prediction quality and class balance.

The results were tabulated across all SMOTE variants and both models, with the **BERT + SMOTE-BD** combination achieving the best performance—**90% balanced accuracy** and **89% recall** on test data.

Disadvantages:

1. **Data Source Limitations and Labeling Errors**: The effectiveness of the model is highly dependent on the quality and integrity of the datasets used. Since the dataset is aggregated from different public sources, inconsistencies in formatting, labeling errors, or outdated job posts may introduce noise and reduce the model's predictive accuracy.
2. **Model Bias and Overfitting Risks**: While transformer-based models offer strong performance, they may still exhibit bias if the training data contains unbalanced or skewed representations of certain job types, industries, or regions. Additionally, excessive reliance on SMOTE variants to address class imbalance may lead to overfitting on synthetic examples, impacting real-world generalization.
3. **High Computational Cost and Complexity**: The implementation of transformer models, SMOTE variants, and comparative evaluation across multiple model architectures significantly increases system complexity and resource requirements, making it less accessible for low-resource or real-time deployment environments.
4. **Limited Interpretability**: Despite high accuracy, the black-box nature of deep learning models, particularly transformers, makes it difficult to explain how individual decisions are made. This limits trust and transparency—particularly in high-stakes domains like recruitment fraud detection.
5. **Language and Regional Bias**: Although the dataset covers job postings from three regions, it may still not generalize well across other countries, languages, or cultural contexts, limiting its applicability in global job markets.

Proposed System:

The Fake Job Detection System is an advanced AI-powered platform designed to identify fraudulent job postings across multiple regions and platforms. The system is built on a novel, enriched dataset curated from three diverse sources—Fake Job Postings, publicly available Pakistani job postings, and U.S.-based job portals. This hybrid data strategy ensures broader coverage and relevance, especially in the face of outdated or narrowly focused benchmark datasets.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

At the core of the system lies a Transformer-based deep learning architecture capable of capturing the nuanced language patterns and metadata signals associated with fraudulent job descriptions. To address the common problem of class imbalance in fraud detection, the system incorporates ten top-performing variants of the Synthetic Minority Oversampling Technique (SMOTE), effectively balancing the dataset and enhancing model reliability.

The detection pipeline includes data preprocessing, Exploratory Data Analysis (EDA), class balancing through SMOTE variants, and a suite of transformer-based classifiers. A comparative evaluation is conducted across both imbalanced and balanced data conditions to rigorously assess model performance and robustness.

By leveraging contextual word embeddings and deep neural architectures, this system aims to improve the accuracy, generalizability, and interpretability of fake job detection, ultimately protecting job seekers from online recruitment fraud.

Advantages of proposed system:

• Modern, Multi-Source Dataset: Combines job postings from three different regions, significantly improving dataset diversity and relevance.
• Class Imbalance Solved with SMOTE: Implements ten advanced SMOTE techniques to balance minority and majority classes, enhancing model fairness and predictiveeperformance.
• Transformer-Based Intelligence: Uses state-of-the-art transformer models for high contextual understanding and fraud pattern recognition.
• Comparative Model Evaluation: Assesses performance on both imbalanced and balanced datasets to validate improvements and ensure robustness.
• Scalable Architecture: Modular pipeline enables easy expansion to support other regions, industries, or dataformats.
• Fraud Protection in Recruitment: Helps mitigate the risks posed by online recruitment scams through accurate and automated detection.

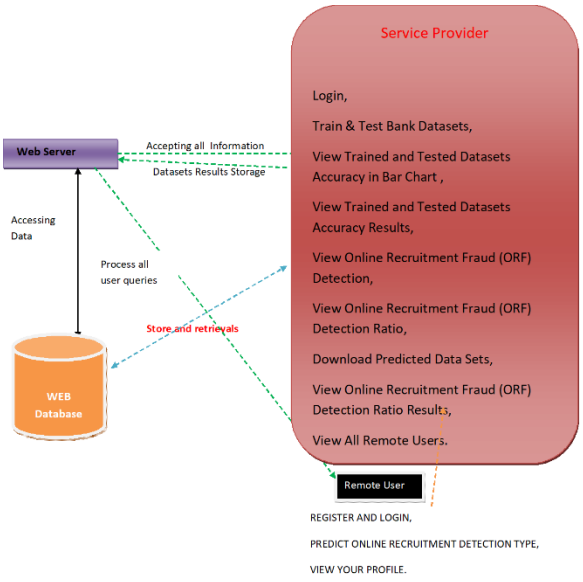

Fig.1: System architecture

3. IMPLEMENTATION

The implementation of the Online Recruitment Fraud Detection System is divided into several functional modules, each designed to perform specific tasks ranging from user input handling to model prediction and result visualization. This section outlines the system's architecture, processing flow, components, and technologies used in its construction.

---

1. Input Handling

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

The system provides an intuitive interface for users—either job seekers or administrators—to submit job posting content for fraud evaluation. Two input options are supported:

- **Direct Text Input**: Users can paste the job description into a text box on the web interface.
- **Batch Input via File Upload**: Users may upload .txt files containing multiple job descriptions, each on a separate line, to enable bulk processing.

The interface validates the input to ensure that empty, malformed, or duplicate records are not processed further. This module acts as the first line of defense, ensuring data quality before prediction.

---

2. Preprocessing and Feature Extraction

After input collection, job descriptions undergo text preprocessing steps including:

- Removal of stop words and special characters
- Tokenization
- Lowercasing
- Lemmatization (where applicable)

Following this, each preprocessed job description is transformed into a dense, high-dimensional vector using either the **BERT** or **RoBERTa** transformer model. These vectors encode the contextual and semantic information necessary for accurate classification.

---

3. SMOTE-Based Data Balancing (Training Phase)

During model training, the system applies one of the ten selected **SMOTE variants** to the training data to address the issue of class imbalance. This module operates offline and is integrated into the training pipeline. Only the training data is oversampled, while validation and testing data remain untouched to ensure unbiased performance evaluation.

The synthetic samples generated by SMOTE enhance the model's ability to learn characteristics of the minority (fraudulent) class, improving recall and reducing false negatives.

---

4. Model Prediction and Output Generation

Upon receiving vectorized input, the system passes it through the fine-tuned deep learning classification model (either BERT or RoBERTa). The model outputs:

- **Prediction Label**: "Fraudulent" or "Legitimate"
- **Confidence Score**: Probability estimate for each class

These results are returned to the user along with optional visualization (e.g., progress bar, label indicator). For administrators, additional metadata (model used, timestamp, source) is logged.

---

5. Historical Tracking and Model Logging

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

A logging module maintains a record of all predictions for monitoring and audit purposes. This includes:

- Input description
- Timestamp
- Model version
- Selected SMOTE technique
- Prediction result and score

This historical view supports system transparency and allows for performance audits or revalidation with newer models.

---

6. Technology Stack

The ORF Detection System is implemented using the following tools and frameworks:

- **Frontend**:
  - HTML5, CSS3, JavaScript
  - Optional UI enhancements using Bootstrap or React
- **Backend**:
  - **Python 3.10**
  - **Django Framework**: Manages routing, input validation, and communication with ML models
- **Machine Learning Stack**:
  - **HuggingFace Transformers** (BERT, RoBERTa)
  - **Imbalanced-learn** (for SMOTE variants)
  - **Scikit-learn** (for performance metrics and auxiliary modeling)
- **Database**:
  - **MySQL**: For logging and storage of historical predictions
- **Deployment (Optional)**:
  - **Docker**: Containerization of the application
  - **Vite** or **Streamlit**: For deploying a lightweight web app version

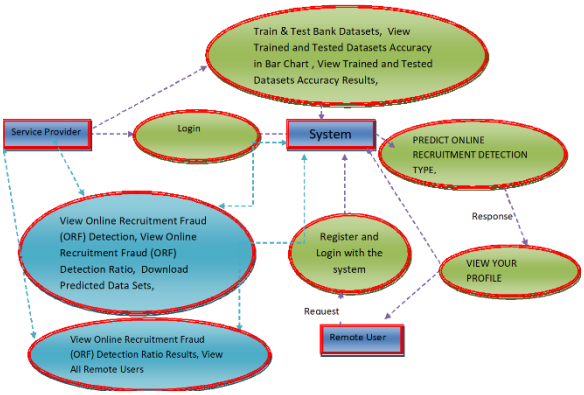

Fig 2 DataFlow

4. EXPERIMENTAL RESULTS

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

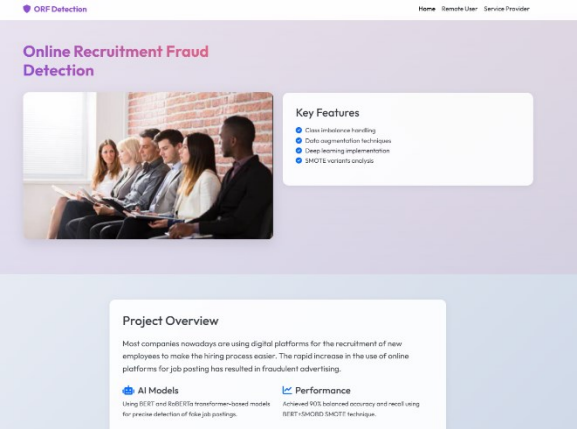

Fig.1: Home Page.

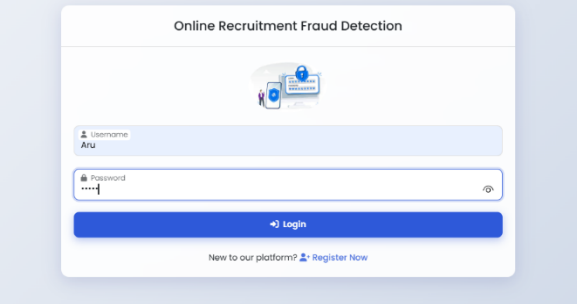

Fig 2: Login

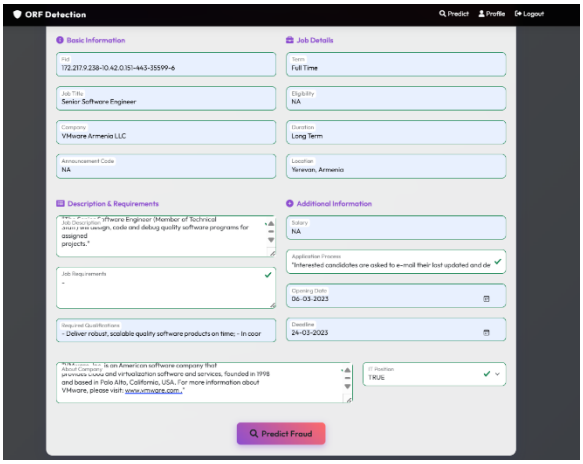

Fig 3: Prediction

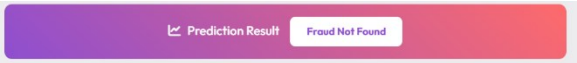

Fig.4: Result

5. CONCLUSION

This research thoroughly explored the issue of fake job posting detection, introducing a unique dataset compiled from three different sources. Through exploratory data analysis, a significant class imbalance was identified, which was addressed using ten advanced SMOTE variants. Type error analysis was also performed to assess the impact of these techniques on model performance. Transformer-based models, BERT and RoBERTa, were applied to both imbalanced and balanced datasets. Among these, BERT combined with the SMOBD SMOTE technique delivered the most effective results.

The study emphasized that relying solely on accuracy as an evaluation metric can be misleading in imbalanced datasets, highlighting the importance of balanced accuracy and recall. The experiments demonstrated that properly handling class imbalance leads to better detection of fraudulent job postings, which can help protect job seekers and organizations from employment scams.

Additionally, this work opens up several avenues for future research. The current analysis was limited to English-language postings; expanding it to include other languages and regional job markets could offer more localized insights. Incorporating newer job postings and remote work opportunities would further enhance the dataset's relevance. Future research could also explore hybrid oversampling methods and explainable AI models for more interpretable and accurate detection.

REFERENCES

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

[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,’’ Revista Gestão Inovação e Tecnologias, vol. 11, no. 2, pp. 642–650, Jun. 2021.

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

[4] Online Fraud. Accessed: Jun. 19, 2022. [Online]. Available: https://www.cyber.gov.au/acsc/report

[5] J. Howington, ‘‘Survey: More millennials than seniors victims of job scams,’’ Flexjobs, CO, USA, Sep. 2015. Accessed: Jan. 2024 [Online]. Available: www.flexjobs.com/blog/post/survey-results-millennials-seniors-victims-job-scams

[6] Report Cyber. Accessed: Jun. 25, 2022. [Online]. Available: https://www.actionfraud.police.uk/

[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] 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.

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

[10] 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.

[11] 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.

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

[13] O. Nindyati and I. G. Bagus Baskara Nugraha, ‘‘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.

[14] 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.

[15] 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.

[16] 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.

[17] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, ‘‘Training neural network classifiers for medical decision making: The effects of imbalanced datasets on

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

classification performance,'' Neural Netw., vol. 21, nos. 2–3, pp. 427–436, Mar. 2008.

[18] Y. Li, G. Sun, and Y. Zhu, ''Data imbalance problem in text classification,''in Proc. 3rd Int. Symp. Inf. Process., Luxor, Egypt, Oct. 2010, pp. 301–305.

[19] N. V. Chawla, K.W. Bowyer, L. O. Hall, andW. P. Kegelmeyer, ''SMOTE:Synthetic minority over-sampling technique,'' J. Artif. Intell. Res., vol. 16,pp. 321–357, Jun. 2002.

[20] S. U. Habiba, Md. K. Islam, and F. Tasnim, ''A comparative study on fake job post prediction using different data mining techniques,'' in Proc.2nd Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST), Dhaka,Bangladesh, Jan. 2021, pp. 543–546.

[21] G. Othman Alandjani, ''Online fake job advertisement recognition and classification using machine learning,'' 3C TIC, Cuadernos de Desarrollo Aplicados a las TIC, vol. 11, no. 1, pp. 251–267, Jun. 2022.

[22] A. Gosain and S. Sardana, ''Handling class imbalance problem using oversampling techniques: A review,'' in Proc. Int. Conf. Adv. Comput.,Commun. Informat. (ICACCI), Delhi, India, Sep. 2017, pp. 79–85.

[23] F. Akhbardeh, C. O. Alm, M. Zampieri, and T. Desell, ''Handling extreme class imbalance in technical logbook datasets,'' in Proc. 59th Annu.Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang.Process., 2021, pp. 4034–4045.

[24] J. Ah-Pine and E.-P. Soriano-Morales, ''A study of synthetic oversampling for Twitter imbalanced sentiment analysis,'' in Proc. Workshop Interact. Between Data Min. Nat. Lang. Process. (DMNLP), Riva del Garda, Italy,Sep. 2016, pp. 17–24.

[25] J. David, J. Cui, and F. Rahimi, ''Classification of imbalanced dataset using BERT embeddings,'' Dalhousie Univ., Halifax, Canada, Jan. 2020. Accessed: Jan. 2024. [Online]. Available: https://fatemerhmi.github.io/files/Classification_of_imbalanced_dataset_using_BERT_embedding.