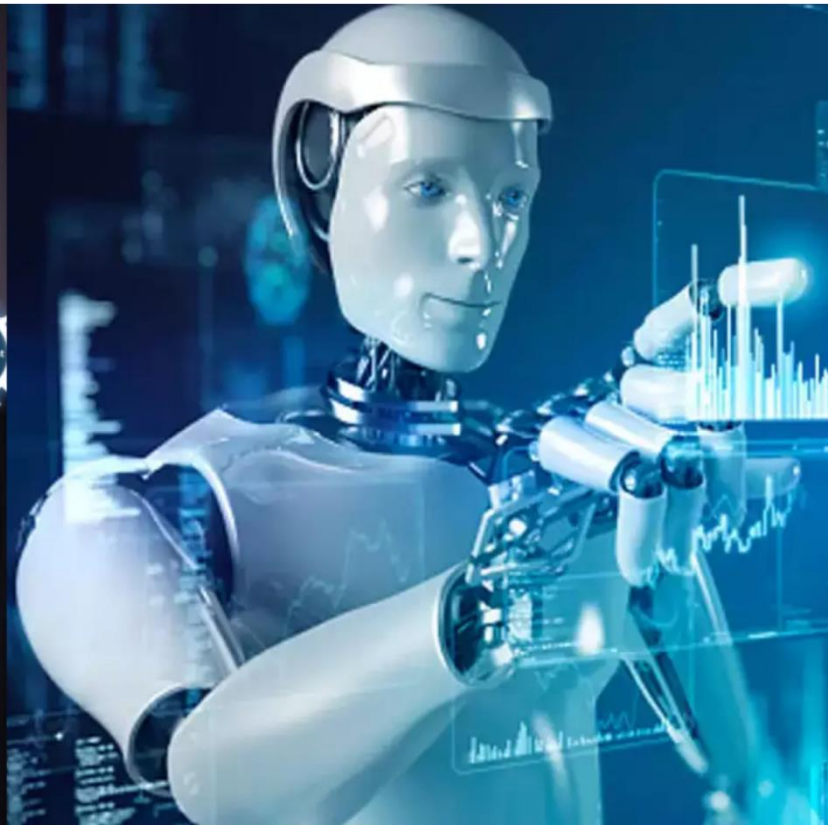
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches

S.Sowmya[1], Kaushal Badvane[2], Noudu Lakshmi Mahitha[3], Kyatham Manideep[4], Misala Manideep[5]

Assistant Professor, Dept. of CSE, Malla Reddy University, Hyderabad, Telangana, India[1]

UG Scholar, Dept. of CSE, Malla Reddy University, Hyderabad, Telangana, India[2]

UG Scholar, Dept. of CSE, Malla Reddy University, Hyderabad, Telangana, India[3]

UG Scholar, Dept. of CSE, Malla Reddy University, Hyderabad, Telangana, India[4]

UG Scholar, Dept. of CSE, Malla Reddy University, Hyderabad, Telangana, India[5]

ABSTRACT: The rise of digital platforms for job recruitment has brought with it a growing threat of fraudulent job postings, undermining the trust and safety of online hiring systems. This paper proposes an advanced fraud detection system based on the ALBERT (A Lite BERT) model to identify fraudulent job postings. The system will utilize a dataset created by merging job postings from multiple sources to better capture both legitimate and fraudulent job listings. A comprehensive pre-processing pipeline, including data cleaning and feature engineering, will be applied to prepare the data for model training. The system will incorporate various techniques, such as SMOTE (Synthetic Minority Over-sampling Technique), to address class imbalance. The ALBERT model will then be fine-tuned to classify job postings, and the system will be evaluated using standard performance metrics like accuracy, precision, recall, and F1-score. The goal is to provide an effective and scalable solution for detecting fraudulent job postings in real-time, enhancing the overall security and trustworthiness of online recruitment platforms. The accuracy of the proposed technique is 99.82% which is higher than other well-known existing techniques.

KEYWORDS: Class imbalance, deep learning, Fraud detection System, ALBERT (A Lite BERT) Model, Online Recruitment Platforms, SMOTE, Job Posting Classification.

I. INTRODUCTION

Online recruitment platforms have revolutionized the hiring process, allowing companies and candidates to connect easily[1]. However, this convenience has also opened the door to various fraudulent activities. Recruitment fraud can involve impersonation of employers, fake job offers, and other deceitful tactics aimed at exploiting job seekers[2],[3]. Detecting such fraud is crucial to maintaining the integrity and trust of online recruitment platforms[2].
Traditionally, fraud detection in online recruitment has been handled using rule-based systems or traditional machine learning models, but these methods often struggle to adapt to the evolving nature of fraud[5].
Deep learning, especially with transformer-based models like ALBERT (A Lite BERT), offers a promising solution for more effective fraud detection.

Therefore, this research aims to analyze and alarm people about rapidly growing employment scams and to detect ORF by implementing transformer-based deep learning models[2]. Consequently, people would not fall into the trap of job scams anymore. So by detecting ORF, the people wasting their time and money on those fraudulent activities can be more careful. In this research, we presented a novel dataset of fake job postings labeled as ‘‘fraudulent’’ for fake job postings and ‘‘non-fraudulent’’ for legitimate job postings[2]. The proposed data is a combination of job postings from three different sources. We use ‘‘Fake Job Postings1 as a primary dataset and add publicly available job postings to extend the dataset with the latest job postings. We have done this because the existing benchmark datasets are outdated and limited due to knowledge of specific job postings, which limits the capability of existing models in detecting fraudulent jobs. After preparing the dataset, Exploratory Data Analysis (EDA) was performed on this data[4]. Through EDA, it was identified that the dataset has an imbalanced class distribution. Imbalance class distribution can be defined as the ratio of the number of samples in the minority class to the number in the majority class[15],[19] . It may cause high predictive accuracy for frequent classes and low predictive accuracy for infrequent classes. Class imbalance problem

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

occurs in various real-world domains, including anomaly detection , face recognition , medical diagnosis , text classification , and many others. SMOTE gained extensive popularity as an oversampling technique. Almost 85 different SMOTE variants have been introduced in the literature and are recently used by various researchers to handle class imbalance problems in multiple domains[15],[19].

II. RELATED WORKS

This section reviews multiple studies related to Online Recruitment Fraud (ORF) detection. Moreover, as it is mentioned earlier that the collected dataset for this research. It has a class imbalance problem associated with it; hence, the literature related to handling class imbalance problem is also reviewed in this section[15].

Fraud Detection in Online Platforms: Recent studies have highlighted the increasing prevalence of fraud in online job markets. Research has focused on various machine learning approaches to identify fraudulent activities. Traditional models, including decision trees and random forests, have been utilized, but they often fall short in handling the nuances of natural language data[4],[6].

Natural Language Processing for Fraud Detection: Natural language processing (NLP) techniques have been applied to analyse job postings and resumes. Previous works have leveraged models like BERT and its derivatives to extract features from textual data, focusing on linguistic patterns that indicate fraudulent behaviour. These models have shown promise in improving the accuracy of fraud detection systems[3].

ALBERT: A Lite BERT for Self-supervised Learning: The ALBERT model, a lightweight version of BERT, has gained attention for its efficiency and effectiveness in NLP tasks. Studies indicate that ALBERT can achieve comparable or superior performance on various benchmark datasets while requiring fewer resources. Its capacity for fine-tuning makes it particularly suitable for domain-specific applications, such as recruitment fraud detection.

Sentiment Analysis and Fraud Detection: Research has explored the use of sentiment analysis in identifying fraudulent job postings. By assessing the sentiment expressed in job descriptions, certain studies have correlated negative sentiment with fraudulent activity. ALBERT's capabilities can be harnessed to enhance sentiment analysis by providing context-aware embeddings[2],[5].

Anomaly Detection Techniques: Existing work on anomaly detection in recruitment focuses on identifying unusual patterns in applicant behaviour or job postings. Combining ALBERT with traditional anomaly detection algorithms can create a hybrid approach that improves fraud detection accuracy[7].

User Behaviour Analysis: Studies have shown that analysing user behaviour on recruitment platforms can help detect fraudulent activities[3],[12]. By incorporating user interaction data (e.g., click patterns, application frequency), ALBERT can be used to create models that predict fraudulent intent based on behaviour.
Real-time Fraud Detection Systems: Several projects have aimed to build real-time fraud detection systems using machine learning[2],[3]. Implementing ALBERT in such systems can facilitate the immediate flagging of suspicious applications and postings, enhancing user trust in online recruitment platforms.
Challenges and Future Directions: Despite advancements, challenges remain in adapting NLP models for fraud detection. Issues such as data imbalance, evolving fraud tactics, and interpretability of model decisions are critical areas for future research. Leveraging ALBERT's efficiency can help address some of these challenges[8].

III. METHODOLOGY

3.1 Data Collection

The data for online recruitment fraud detection was collected from EMSCAD which includes job descriptions, titles, company information, and metadata like posting date and location. Label data as fraudulent or non-fraudulent[9].

3.2 Data Preprocessing

Clean the text by tokenizing job descriptions using ALBERT's built-in tokenizer, removing noise (special characters, URLs), and converting text to lowercase. Convert text into embeddings using ALBERT and prepare additional

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

features like job category and company reputation.

3.3 Model Architecture and Training

Fine-tune a pre-trained ALBERT model on the labeled dataset. The ALBERT model processes the input text using its transformer layers, followed by a dense layer for binary classification (fraudulent vs. non-fraudulent). Use binary cross-entropy loss and Adam optimizer for training.

3.4 Model Evaluation

Model evaluation is crucial to ensure the ALBERT model performs effectively in detecting fraudulent job listings. Using k-fold cross-validation helps assess the model's ability to generalize and avoid overfitting. Key evaluation metrics include accuracy, though it's less reliable in imbalanced datasets, and precision and recall, which are critical for fraud detection. Precision measures the correctness of predicted fraud, while recall evaluates the model's ability to identify all fraudulent listings, with recall being prioritized to avoid missing fraud. The F1-score provides a balanced measure of precision and recall, making it particularly useful for imbalanced datasets[15],[19]. The AUC-ROC curve assesses the model's overall ability to distinguish between fraudulent and non-fraudulent posts, with a higher AUC indicating better performance. The confusion matrix visualizes true positives, false positives, true negatives, and false negatives, offering deeper insights into the model's errors and effectiveness in detecting fraud[5].

3.5 Deployment

Deployment of the ALBERT-based fraud detection model involves integrating it into online recruitment platforms for real-time job posting analysis[2]. The model can be deployed as a web service using frameworks like Flask , allowing it to receive job postings and classify them as fraudulent or non-fraudulent automatically[3]. Once deployed, it can flag suspicious job listings, providing feedback to users (job seekers or recruiters). Continuous monitoring is essential to track model performance in real-world scenarios, and periodic retraining ensures that the model stays updated with evolving fraud tactics. This integration helps maintain an effective, automated fraud detection system[6].

3.6 Continous Improvement

Implement a feedback loop for users to validate model predictions. Periodically retrain the model with new data to adapt to evolving fraudulent tactics.

IV. DESIGN AND IMPLEMENTATION

The processes involved in the Online Recruitment Fraud (ORF) Detection using the Deep Learning approaches have been systematically represented through block diagrams, flow charts, and architectural components.
System Architecture and Components The system architecture consists of several layers, including data collection, processing, modeling, and user interfaces.
Data collection and preprocessing: This module involves gathering job postings from multiple sources and performing necessary pre-processing steps, such as cleaning the data, handling missing values, removing irrelevant content, and renaming columns for clarity and consistency. The pre-processed data will then be prepared for model training.

Model Training and Evaluation: In this module, the ALBERT model will be fine-tuned to classify job postings. The performance of the model will be evaluated using key metrics, such as accuracy, precision, recall, and F1-score, to assess its ability to correctly classify job postings as fraudulent or legitimate.

Real-Time Classification and User Interface:
This module will allow for real-time processing of job postings. New job listings will be classified as they arrive, with the ALBERT model applied to classify each posting. A user-friendly web interface, built with Flask, will allow users to submit job postings for classification and view the results. The interface will also handle user authentication and session management to ensure secure use.
Frontend: It offers an interactive, dynamic, and user-friendly interface to the users, and administrators. This enables real-time interactions and easy navigation through the application.
Backend: The server-side logic for user authentication, and role-based access control are managed using the JWT (JSON web token).

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

Database (sqlite3): SQLite3 can effectively support an "Online Recruitment Fraud Detection" system by providing a lightweight, flexible, and easy-to-use database solution. Its capabilities in managing user data, job postings, and application records, along with its integration with SQL and machine learning workflows, make it a suitable choice for such applications.

API Layer: Facilitates communication between the frontend and backend, enabling efficient data exchange.

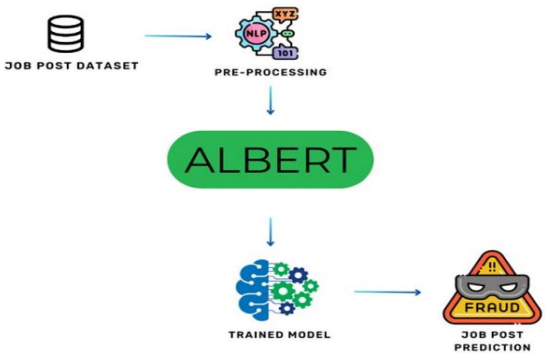


FIGURE 4.1: System Architecture

Workflow Representation
The flow chart below outlines the sequential steps for job posting submission, fraud prediction:
User Registration: Users and administrators create accounts by providing necessary information. Role-based access control ensures users have appropriate privileges.
Job Posting Description Submission: Users upload their Job Posting Description.
Data Collection: Scrape job postings from online platforms, collect labelled datasets (legitimate vs. fraudulent), and gather additional features such as job descriptions, company details, salary, and location.
Data Preprocessing: Clean text by removing stop words, special characters, and unnecessary data. Tokenize job descriptions and encode categorical labels to prepare the dataset for feature extraction.
Feature Extraction using ALBERT: Use the ALBERT model to convert job descriptions into dense vector embeddings, capturing contextual meaning and semantic relationships for fraud detection.
Model Training & Evaluation: Fine-tune ALBERT on the dataset using a classification layer. Evaluate model performance using accuracy, precision, recall, and F1-score. Optimize hyperparameters if needed.
Fraud Detection: Input a new job posting, and the trained model classifies it as legitimate or fraudulent based on learned patterns. Confidence scores help determine reliability.

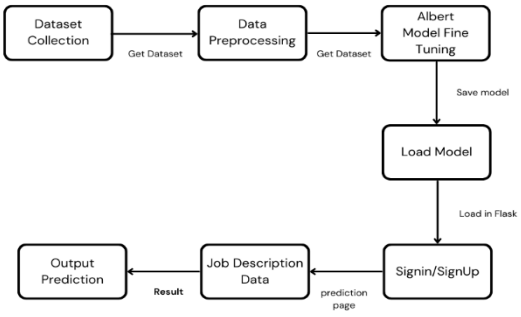


FIGURE 4.1: WORKFLOW DIAGRAM

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

Component	Technology Stack	Why This Technology?
Frontend	HTML CSS JavaScript	HTML structures the web page, CSS styles the interface, and JavaScript enables interactivity and API communication for web application.
Backend	Python Deep Learning	Python is used for implementing the ALBERT model and backend processing, while Deep Learning enables accurate fraud detection by analysing job descriptions for hidden patterns.
Database	SQLite3	SQLite3 is useful for web applications because it is lightweight, requires no server setup, and efficiently handles small to medium-sized databases with minimal configuration.
Authentication	JWT (JSON Web Tokens)	Secure user authentication for authors, reviewers, and admins.
API	RESTful APIs	Simplifies communication between frontend and backend.
Version Control	Git + GitHub	Collaboration, version management, and code tracking.
Testing	Unit testing	It ensures the ALBERT-based fraud detection model functions correctly by validating data preprocessing, model predictions, and API responses, reducing errors before deployment.

TABLE 4.1: TECHNOLOGY IMPLEMENTATION

V. RESULTS AND DISCUSSION

a. Results

The development and implementation of the **Online Recruitment Fraud Detection using Deep Learning Approaches** were carried out successfully, achieving the intended functionalities. The system was tested under various conditions to

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

evaluate its performance, usability, and security.

FIGURE 5.a.1: HOME PAGE

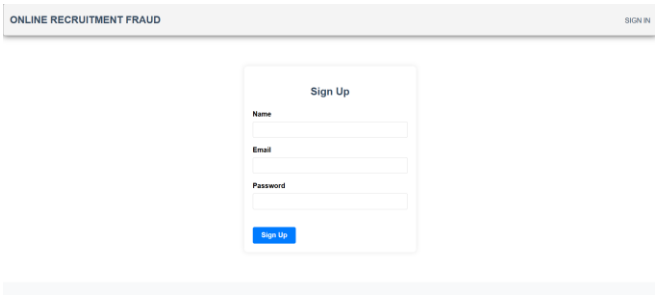


FIGURE 5.a.1: SIGN UP PAGE

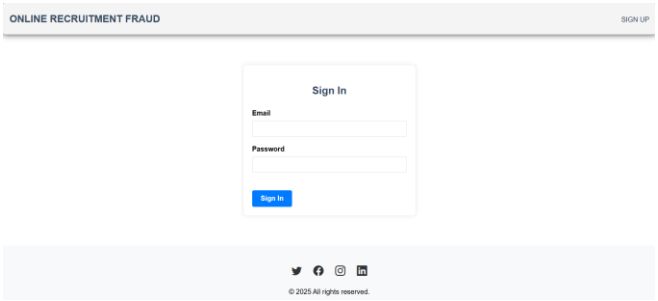


FIGURE 5.a.1: SIGN IN PAGE

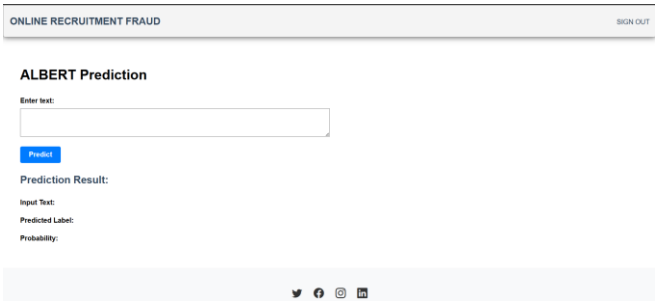


FIGURE 5.a.1: ALBERT PREDICTION PAGE

b. Discussion

The construction of the Online Recruitment Fraud Detection using Deep Learning Approaches (i.e. ALBERT model) has proven the possibility of building a contemporary, efficient, and scalable solution for detecting the fake job postings. The

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

outcomes show that the system efficiently processes job posting submissions and prediction of fake job postings by detecting the given job posting is legitimate or fraudulent using ALBERT model .
Online recruitment fraud is a growing challenge, with fraudsters exploiting job seekers through fake job postings. Traditional fraud detection methods, such as manual reviews, are ineffective due to the increasing volume of job listings and evolving fraud tactics. Implementing an AI-driven approach using the ALBERT model provides an efficient and automated solution to identify fraudulent job postings with high accuracy.

The ALBERT model, a lightweight variant of BERT, is highly effective for text classification tasks. It captures the context and semantics of job descriptions, enabling the system to distinguish between legitimate and fraudulent job posts. By leveraging deep learning-based embeddings, ALBERT minimizes dependency on manually engineered features and improves detection accuracy compared to traditional machine learning models.

Key components of the fraud detection system include data preprocessing, feature extraction, model training, and deployment. The system can be integrated into job portals, where it continuously analyzes new job postings in real time, flagging fraudulent ones and alerting users. This reduces the risk of job scams and ensures a safer job search experience. However, challenges remain, such as evolving fraud patterns, data imbalance, and deployment complexities. To enhance detection, the model can be periodically retrained with updated datasets and combined with additional fraud indicators, such as employer credibility and salary inconsistencies.
In Summary, ALBERT-based fraud detection is a scalable, intelligent, and effective solution for securing online recruitment platforms, ensuring trust and safety for job seekers and employers .

VI. CONCLUSION

Fraudulent job postings have become a significant issue on digital recruitment platforms, undermining the trust and safety of these systems. This paper proposes an advanced fraud detection system based on the ALBERT (A Lite BERT) model, aiming to accurately identify and classify fraudulent job listings. By using a diverse dataset and a robust data preprocessing pipeline, the system leverages natural language processing to improve fraud detection accuracy. The ALBERT model is fine-tuned and evaluated for optimal performance in distinguishing legitimate postings from fraudulent ones. The proposed system offers a solution to the growing threat of job posting scams, ensuring a safer and more reliable experience for both job seekers and employers. As a result, the design and testing of the project were successful.

ACKNOWLEDGEMENTS

We express our sincere gratitude to **Malla Reddy University** for providing the necessary resources and support to carry out this research. We extend our heartfelt thanks to our mentor, **S. Sowmya**, for their invaluable guidance, insightful feedback, and constant encouragement throughout the development of this project.
We would also like to acknowledge our team members **Kaushal Badvane, Noudu Lakshmi Mahitha, Kyatham Manideep and Misala Manideep** for their dedication, collaboration, and collective efforts in making this project a success. Their contributions have been instrumental in the research and development process.
Finally, we appreciate the open-source community and developers of the python and deep learning technologies, whose contributions have enabled the successful implementation of this Online Recruitment Fraud (ORF) Detection using Deep Learning Approaches.

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, Jun. 2022, [online] Available: https://www.cyber.gov.au/acsc/report.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404255|

[5] J. Howington, Survey: More millennials than seniors victims of job scams, Flexjobs, CO, USA, Sep. 2015, [online] Available: https://www.flexjobs.com/blog/post/survey-results-millen-nials-seniors-victims-job-scams.
[6] Report Cyber, Jun. 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, pp. 6, Mar. 2017.
[8] B. Alghamdi and F. Alharby, "An intelligent model for online recruitment fraud detection", J. Inf. Secur., vol. 10, no. 3, pp. 155-176, 2019.
[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:
[11] Ensemble learning based online recruitment fraud detection", Proc. 12th Int. Conf. Contemp. Comput. (IC3), pp. 1-5, Aug. 2019.
[12] I. M. Nasser, A. H. Alzaanin and A. Y. Maghari, "Online recruitment fraud detection using ANN", Proc. Palestinian Int. Conf. Inf. Commun. Technol. (PICICT), pp. 13-17, Sep. 2021.
[13] 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.
[14] O. Nindyati and I. G. Bagus Baskara Nugraha, "Detecting scam in online job vacancy using behavioral features extraction", Proc. Int. Conf. ICT Smart Soc. (ICISS), vol. 7, pp. 1-4, Nov. 2019.
[15] 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.
[16] 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.
[17] Y.-H. Liu and Y.-T. Chen, "Total margin based adaptive fuzzy support vector machines for multiview face recognition", Proc. IEEE Int. Conf. Syst. Man Cybern., pp. 1704-1711, Oct. 2005.
[18] 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 classification performance", Neural Netw., vol. 21, no. 2, pp. 427-436, Mar. 2008.
[19] Y. Li, G. Sun and Y. Zhu, "Data imbalance problem in text classification", Proc. 3rd Int. Symp. Inf. Process., pp. 301-305, Oct. 2010.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details