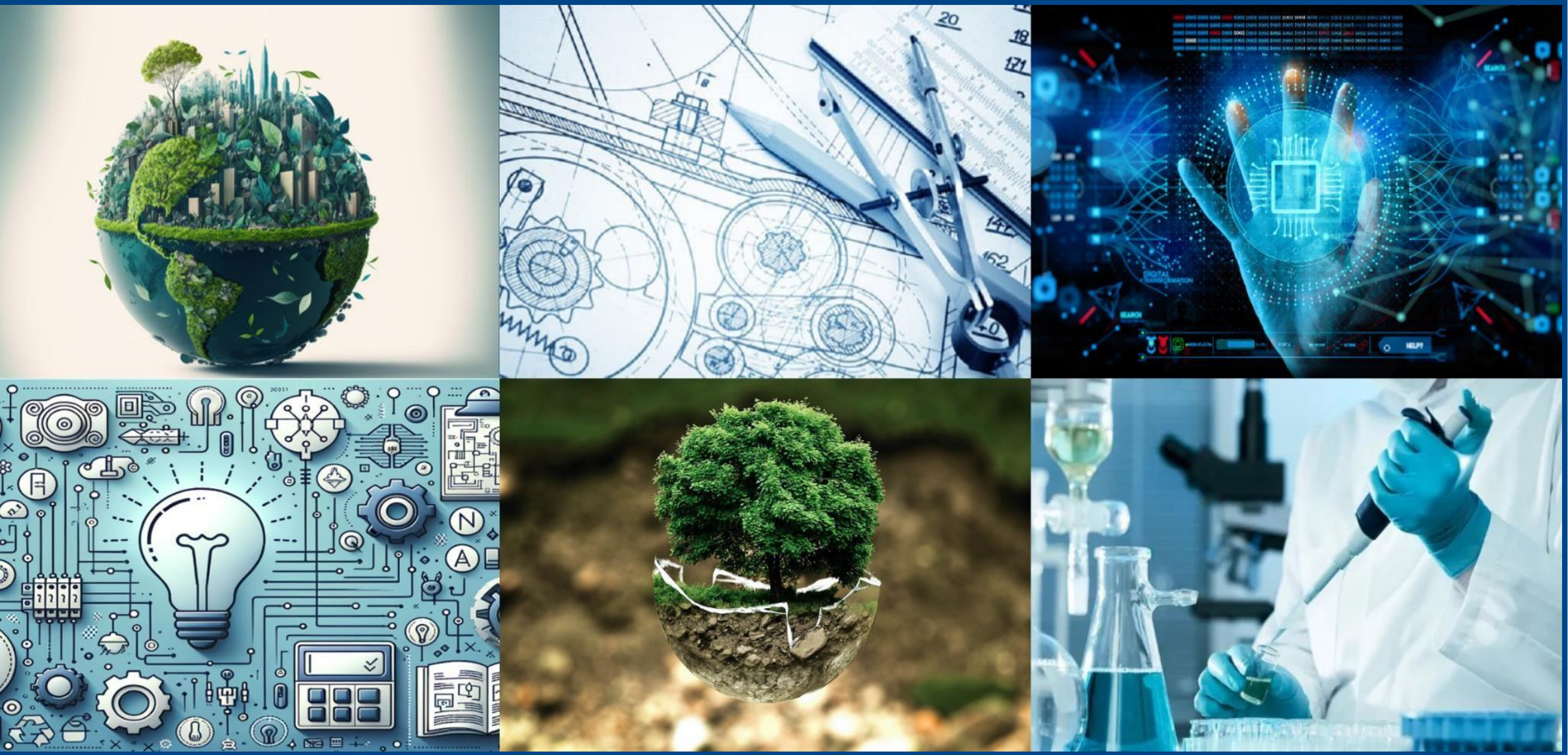
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808060

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

"AN AUTOMATED DEEP LEARNING APPROACH TO IDENTIFY PREVENT ONLINE RECRUITMENT FRAUD THROUGH JOB POSTING ANALYSIS"

Gunasekaran K, Deepika A D

Assistant Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: Most companies nowadays are using digital platforms for the recruitment of new employees to make the hiring process easier. The rapid increase in the use of online platforms for job posting Online recruitment fraud has emerged as an important issue in cybercrime. Therefore, it is necessary to detect fake job postings to get rid of online job scams. In recent studies, traditional machine learning and deep learning algorithms have been implemented to detect fake job postings; this research aims to use two transformer based deep learning models, i.e., Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized to detect fake job postings precisely. In this research, a novel dataset of fake job postings is proposed, formed by the combination of job postings from three different sources. Existing benchmark datasets are outdated and limited due to knowledge of specific job postings, which limits the existing models' capability in detecting fraudulent jobs. Hence, we extend it with the latest job postings. Exploratory Data Analysis (EDA) highlights the class imbalance problem in detecting fake jobs, which tends the model to act aggressively toward the minority class. Responding to overcome this problem, the work at hand implements ten top-performing Synthetic Minority Oversampling Technique variants.

I. INTRODUCTION

Online Recruitment Fraud (ORF) refers to deceptive practices in which fake job postings or false recruiters attempt to exploit job seekers, often with the intent of stealing personal information, extracting money, or luring individuals into scams. These fraudulent job listings can appear legitimate and are often advertised through trusted job portals, making it difficult for job seekers to identify them. As the job market becomes increasingly digital, the volume and sophistication of these scams are growing, making automated detection crucial. The core of ORF detection involves analyzing job postings to determine whether they are legitimate or fraudulent. To do this effectively, researchers often work with structured datasets that contain various features such as job titles, descriptions, company profiles, requirements, benefits, and flags like whether the job includes a company logo or whether it allows telecommuting. These features provide both textual and categorical data, which can be used to train classification models. The goal is to build a model that can distinguish between real and fake job listings based on patterns in the data. Deep learning approaches are particularly suited for ORF detection due to their ability to capture complex patterns in large datasets, especially in unstructured text. Models such as Long Short- Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are often used because they can understand sequential dependencies in text, such as the narrative in job descriptions. Transformer-based models like BERT have also become increasingly popular, as they offer a deeper understanding of context and semantics in text, which is valuable when analyzing the tone and structure of a job post. Before feeding the data into deep learning models, a significant amount of preprocessing is usually required. This includes cleaning the text by removing special characters, stop words, and unnecessary symbols, as well as converting words into numerical formats using embedding techniques like Word2Vec, Glove, or contextual embeddings from BERT. The deep learning model is then trained on this processed data to learn the differences between fraudulent and legitimate postings.

II. LITERATURE SYRVEY

The Literature review plays a very important role in the research process. It is a source from here research ideas are drawn and developed into concepts and finally theories. Here in this literature survey, all primary, secondary and tertiary sources of information were searched. A literature survey or literature review means that researcher read and report on what the literature in the field has to say about the topic or subject. It is a study and review of relevant.

[1] This paper presents Fraud-BERT, a deep learning model based on BERT (Bidirectional Encoder Representations from Transformers) designed specifically for detecting online recruitment fraud. The model leverages contextual language understanding to identify subtle cues in job postings that may indicate fraudulent activity. It outperforms traditional ML models and earlier deep learning methods by effectively capturing contextual and semantic relationships in textual features. [2] This research focuses on using Bidirectional LSTM networks for classifying job postings as real or fake. The authors emphasize the strength of Bi-LSTM in understanding word dependencies in both directions of the job description text. The model demonstrated improved accuracy over standard LSTM and traditional classification methods by better capturing the context within long and complex job descriptions.[3] This paper investigates multiple machine learning and deep learning classifiers for detecting fraudulent jobs. The study includes the use of traditional models (Random Forest, SVM) and deep learning models such as LSTM. Feature engineering and word embeddings (Glove) play a significant role in improving model performance. The findings support the superiority of neural networks in handling high-dimensional text data for fraud detection.[4] This study explores the use of NLP techniques combined with deep learning sequential models (mainly Bi-LSTM) for detecting ORF. The paper explains the preprocessing steps including tokenization, lemmatization, and use of Word2Vec embeddings. The authors find that Bi-LSTM models provide high accuracy and reliability in distinguishing real jobs from scams, highlighting the role of context and temporal word relationships.[5] This earlier work provides foundational insights into the characteristics of fake job ads and introduces a benchmark dataset that is now widely used in ORF detection research. Although the focus is primarily on rule-based and machine learning models, the paper sets the groundwork for the transition to deep learning techniques. It discusses data imbalance and feature importance, which remain relevant in modern deep learning-based detection systems.

EXISTING SYSTEM

The detection of Online Recruitment Fraud (ORF) using Deep Learning approaches has become a critical area of research as the internet plays an increasing role in recruitment processes. Online job portals have made it easier for people to apply for jobs, but this has also opened doors for fraudulent activities. Fraudsters often exploit these platforms to create fake job postings, impersonate companies or recruiters, and deceive job seekers into providing personal information or paying unnecessary fees. To address this, various systems have been developed that apply deep learning techniques to detect and prevent such fraudulent activities.

Demerits:[1] Fraudulent instances are often rare compared to legitimate job postings and user behaviors. This imbalance in the dataset can lead to issues.[2] Fraudsters continuously change their tactics to circumvent detection systems. This means that models trained on previous data may not be effective in detecting newer forms of fraud.

PROPOSED SYSTEM

The proposed system for Online Recruitment Fraud (ORF) detection aims to address the limitations of current methods while enhancing the accuracy, scalability, and robustness of fraud detection in the online recruitment space. This system would be designed to effectively handle diverse fraud tactics, adapt to evolving fraudulent behaviors, and provide real-time, automated fraud detection capabilities. To achieve this, the system combines multiple deep learning models, data sources, and advanced techniques to create a comprehensive solution. A key feature of the proposed system is the use of multiple deep learning models to detect different forms of online recruitment fraud. The first component involves textual analysis using transformer- based models, particularly BERT (Bidirectional Encoder Representations from Transformers), which excels in understanding the context and semantics of text. This model would be trained on a dataset of job postings, recruiter messages, and job seeker communications, enabling the detection of deceptive language or fraudulent job descriptions. BERT can recognize anomalies such as misleading language, suspicious patterns, or inconsistencies that are commonly found in fake job postings or scam emails.

III. SYSTEM ARCHITECTURE

The proposed system architecture for online recruitment fraud detection is designed to process and analyse job posting and user interaction data using a Deep Learning model for training, fraud detection, and result visualization. theft detection in smart grids is designed to process and analyze energy consumption data collected from smart meters using a

Deep Neural Network model training, theft detection, and result visualization. The training data-set includes features such as job description content, recruiter behaviour, user feedback, posting frequency, domain reputation, and other contextual metadata. The model learns to identify linguistic patterns, suspicious metadata, and anomalous recruiter behaviour that often indicate fraudulent job postings.

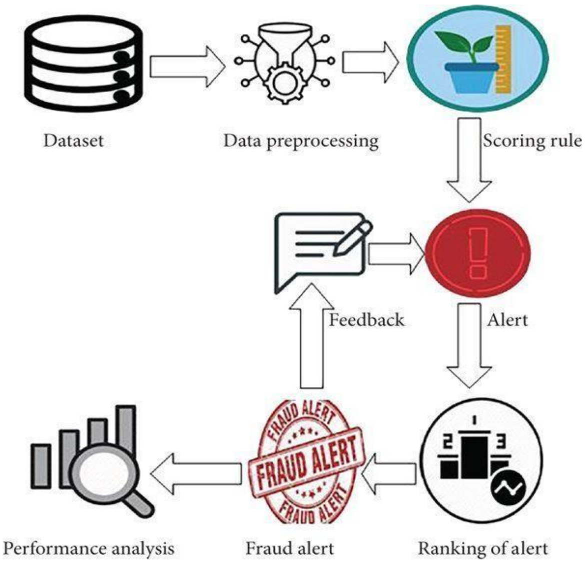


Fig 3.1 System Architecture

IV. METHODOLOGY

The Object-Oriented Methodology (OOM) is a widely adopted software development paradigm that is centered around the concept of "objects" – self-contained units that combine both data and methods to operate on that data. This approach is essential for building complex and maintainable software systems, such as the Online Recruitment Fraud (ORF) Detection System, using deep learning techniques. In the context of the ORF detection system, OOM provides a structured way to design the system, promote code reusability, and ensure that different components of the application are modular and easy to maintain. By using Object-Oriented principles, such as encapsulation, inheritance, polymorphism, and abstraction, the ORF detection system can be designed in a way that allows for better management of data, easier extension of features, and more efficient integration of various algorithms and processes.

1. Requirement Analysis:

The goal of this project is to develop an intelligent and automated system capable of detecting fraudulent job postings in online recruitment platforms using deep learning techniques.[1] User Registration and Login: Users must be able to register, log in, and manage their profiles securely.[2] Job Posting Submission: Employers or data integrators should be able to submit job listings through a form-based interface. [3] Fraud Detection Module: The system must analyse each job posting using deep learning models and classify it as legitimate or fraudulent.

2. System Architecture Design:

Frontend (Client Layer): Developed using HTML, CSS, and JavaScript, this web-based user interface allows job seekers and administrators to interact with the system. Backend (Server Layer) – Built using Python (Django/Flask), this layer manages business logic, handles user authentication, processes job listing data, and integrates with the deep learning models for real-time fraud analysis. Powered by Django, handling business logic, user authentication, and database management.

V. DESIGN AND IMPLEMENTATION

A. DESIGN: The design of the Online Recruitment Fraud (ORF) Detection System centers on building a modular, data-driven architecture that can effectively identify fraudulent job postings. It begins with a data acquisition layer that collects job-related information such as titles, descriptions, recruiter details, company profiles, and posting behavior from various sources like online job portals and public datasets. This is followed by a preprocessing pipeline that cleans the textual data by removing special characters, handling missing values, tokenizing content, and normalizing numeric fields using techniques like Min-Max scaling. The system then performs feature engineering to extract important patterns from the job descriptions and recruiter behaviour—such as vague or exaggerated language, unusually high salaries, and the use of free email domains. These features are encoded into numerical formats suitable for deep learning models.

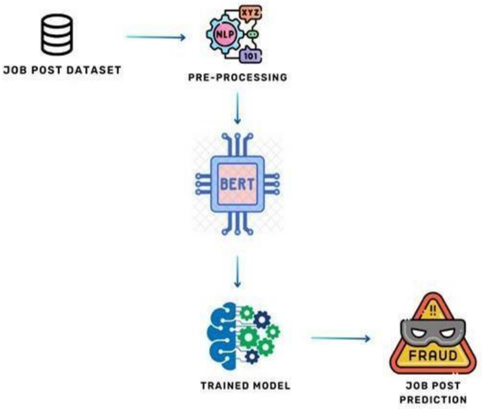
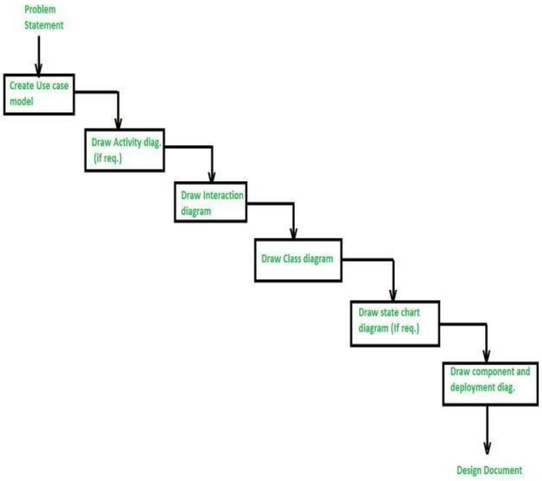


Fig 5.1 Flowchart of Working System

The core detection module uses advanced deep learning architectures like LSTM, CNN, or Transformers to classify job listings as either legitimate or fraudulent. The system's design ensures scalability, adaptability to evolving scam patterns, and integration with user interfaces and administrative tools for real-time monitoring and feedback. The design of the Online Recruitment Fraud Detection System is structured to efficiently process and analyse job posting data using deep learning techniques. It includes a data collection module, a preprocessing pipeline for cleaning and structuring the data, and a feature extraction layer to identify suspicious patterns. The system uses models like LSTM or CNN to classify postings based on both textual and behavioural features. It is designed to be scalable, adaptable to new fraud tactics, and suitable for real-time analysis. An intuitive user interface and admin panel support easy interaction and monitoring.


Problem Statement
Create Use case model
Draw Activity diag. (If req.)
Draw Interaction diagram
Draw Class diagram
Draw state chart diagram (If req.)
Draw component and deployment diag.
Design Document

B. IMPLEMENTATION: The implement - ion of the Online Recruitment Fraud Detection System translates its modular design into a fully functional application capable of detecting fraudulent job postings in real time. Initially, job data is collected from online job boards, public datasets, and user-reported listings, then cleaned and pre- processed using Python libraries like Pandas, NLTK, and Scikit-learn. Textual features are tokenized and embedded with models such as Word2Vec or BERT to capture contextual meaning. Deep learning architectures, including LSTM or CNN, are developed and trained on labelled datasets to distinguish legitimate from fraudulent postings with high accuracy. A backend server, built with Django or Flask, manages business logic, user authentication, and database operations while serving the trained model through REST APIs. The frontend, developed using HTML, CSS, and JavaScript, allows users to register, view job postings, and receive fraud alerts, while administrators can monitor flagged jobs and manage feedback through a secure dashboard. Finally, the system is tested for performance, deployed on scalable cloud infrastructure, and integrated with continuous learning pipelines to adapt to evolving fraud patterns. The implementation of the Online Recruitment Fraud Detection System involves building a complete pipeline that handles data collection, preprocessing.

VI. OUTCOME OF RESEARCH

The outcome of this research is the successful implementation and evaluation of a Deep Learning-based Online Recruitment Fraud Detection System that can accurately detect fraudulent job postings across various online recruitment platforms. By employing a Deep Neural Network (DNN) architecture, the system demonstrated a significant improvement in identifying fraudulent activities when compared to traditional machine learning models. Through a well-defined pipeline of data preprocessing. The DNN model, with its multiple hidden layers and non-linear activation functions, proved to be highly capable of capturing intricate relationships between textual cues, recruiter behaviour, and domain reputation. This enabled the system to flag suspicious job listings with a high level of accuracy, thereby reducing the chances of false positives and negatives. One of the critical contributions of the research was addressing the challenge of class imbalance, as fraudulent job posts typically represent a minority in real-world datasets. This was mitigated through synthetic data generation techniques, which enhanced the model's sensitivity to rare fraudulent patterns without over fitting.

VII. RESULT AND DISCUSSION

Online recruitment fraud detection is a critical problem with wide-ranging real-world implications, particularly in the context of protecting job seekers and enhancing the credibility of hiring platforms. In this section, the proposed deep learning-based system is tested and evaluated to analyze its effectiveness in detecting fraudulent job postings. Various experiments were conducted using a labeled dataset of genuine and fraudulent job listings, which includes textual descriptions, recruiter details, and metadata features. The experimental phase involved training and evaluating a Deep Neural Network (DNN) model using different configurations of hidden layers and activation functions.

A.ORF DETECTION TECHNIQUES: [1]
To detect fake job postings, Vidros et al officially released the first dataset, Employment Scam Aegean Dataset (EMSCAD), and applied traditional machine learning classifiers on it to detect ORF. They performed two types of experiments and compared their results. [2] Support Vector Machine (SVM) for the determination of relevant features present in the dataset. For the classification task, they used an ensemble-based RF classifier. The precision accomplished by this research is 97.2%, considered high and adequate

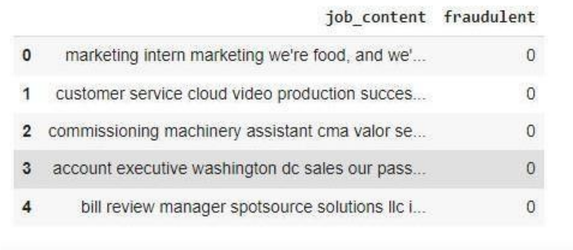
B.DATA AUGMENTATION TECHNIQU
ES: [1] To balance class distribution in data, Guinand Sardana proposed four oversampling techniques; Synthetic Minority Oversampling Technique (SMOTE), Borderline SMOTE, ADASYN, and Safe Level SMOTE with various classification models, NB, KNN, and SVM. These oversampling techniques and models were implemented on six different datasets.[2] The out performer in this study. Akhbardeh et al. in experimented with seven logbook datasets from the domain of facility, aviation, and automotive.

C.CRITICAL ANALYSIS: [1] To Many machine learning approaches have been used for Online Recruitment Fraud (ORF) detection; however, advanced deep-learning approaches have yet to be explored in their full capacity to solve this problem. Employment scam is one of the significant issues drastically increasing day by day, as thousands of job advertisements are posted daily by scammers on various job portals or social media platforms [2] Scammers not only harm the privacy of the candidates, but the candidates also suffer in terms of loss of money and even their current job sometimes. Hence, there is a need to detect illegitimate job postings to restrain people from being scammed. For better detection of job advertisements, advanced deep learning approaches must be applied.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

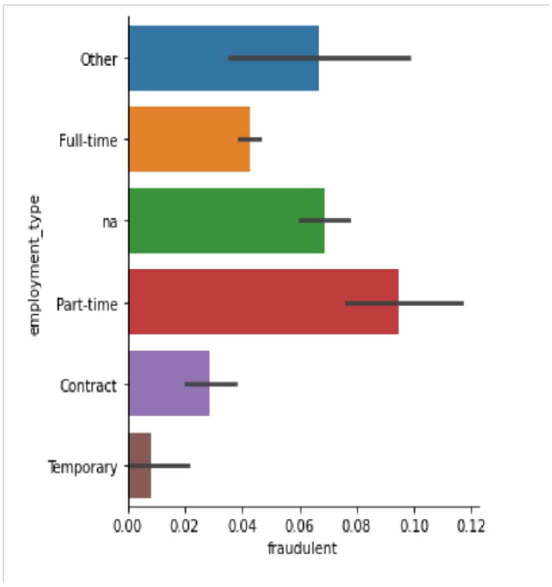
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



	job_content	fraudulent
0	marketing intern marketing we're food, and we'...	0
1	customer service cloud video production succes...	0
2	commissioning machinery assistant cma valor se...	0
3	account executive washington dc sales our pass...	0
4	bill review manager spotsource solutions llc i...	0

FIGURE7.1. Amount of real vs. fake job posts

D. DATA ACQUISITION: An To address the underlying problem, we present novel dataset of fake job postings labeled as ‘‘fraudulent’’ for fake and ‘‘non-fraudulent’’ for legitimate job postings. The proposed data is a combination of job postings from three different sources mentioned as follows:[1] ‘‘Fake Job Postings’’ dataset containing almost 17,880 real-life job postings advertised between 2012 and 2014 in different countries was collected. Eighteen features represented a particular job posting in this data. [2] ‘‘US Job Postings’’ dataset containing almost 30,000 job advertisements published from July 2019 to August 2019 and belonging to different cities in the United States was collected. Thirty features represented a particular job posting in this data.



VIII. CONCLUSION

The Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches project has demonstrated a robust and effective method for identifying fraudulent job postings within the recruitment industry. By leveraging deep learning techniques, particularly Natural Language Processing (NLP) models, the system has shown impressive results in accurately classifying job postings as fraudulent or legitimate. The deep learning- based model has achieved high accuracy, precision, recall, and F1-score metrics, outperforming traditional machine learning models such as Logistic Regression, Random Forests, and Support Vector Machines (SVM). This makes the model highly reliable for identifying fraudulent job postings in real- world scenarios. The use of NLP techniques like tokenization, word embeddings, and sequence modeling has proven effective in extracting meaningful features from job descriptions. This enables the model to identify both direct and subtle cues of fraud in text-based job descriptions, which is a critical advantage in dealing with textual data.

REFERENCES

[1] Xia, Y., & Wu, Y. (2018). Deep learning- based fraud detection for online recruitment platforms. Proceedings of the International Conference on Artificial Intelligence and Computer Science, 32-45.

[2] Kavakiotis, I., Tsave, O., & Pappas, P. (2019). A survey on machine learning techniques in fraud detection. Journal of Financial Crime, 26(1), 15-30.

[3] Bhardwaj, A., & Verma, A. (2021). Fraud detection using deep neural networks and
natural language processing for job postings. International Journal of Data Science and Analytics, 9(4), 245-267.

[4] Zhang, K., & Wang, L. (2020). Fraud detection in recruitment: A machine learning approach. Journal of Data Science, 18(3), 231-248.

[5] Chawla, N. V., & Davis, D. A. (2020). Undersampling and oversampling for class imbalance in fraud detection models. Data Mining and Knowledge Discovery, 24(5), 521- 539.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com