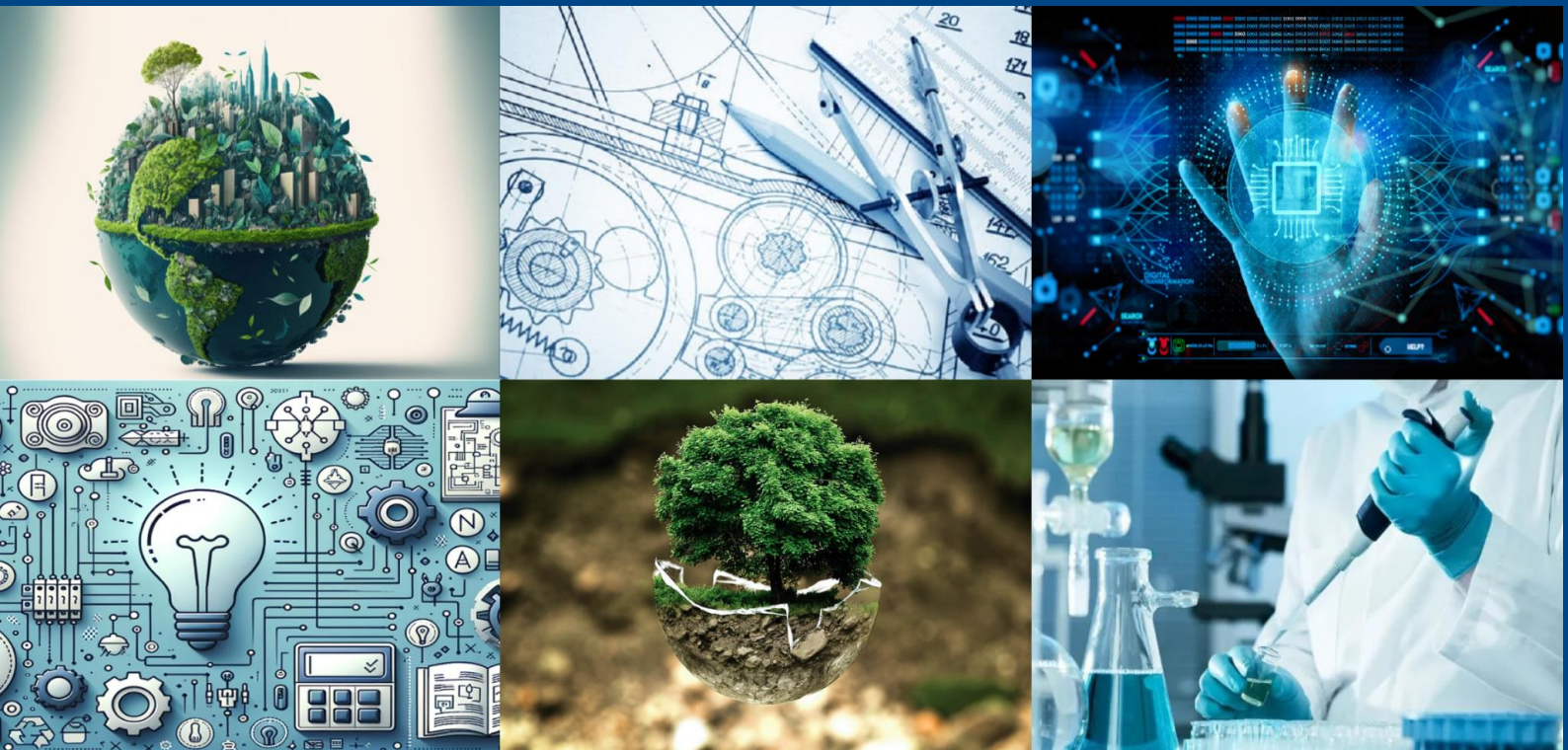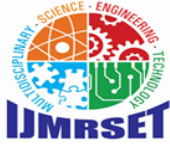# International Journal of Multidisciplinary
Research in Science, Engineering and Technology

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

Leveraging Machine Learning Techniques for Detecting Fraudulent Job Postings

C.Christinaa Princy, Dr.R.Suganthi MCA., M. Phil.

III-B.Sc., Department of Computer Science with Data Analytics, Dr.N.G.P. Arts and Science College,

Coimbatore, India

Professor, Department of Computer Science with Data Analytics, Dr.N.G.P. Arts and Science College,

Coimbatore, India

ABSTRACT: With the rapid growth of online job portals, the recruitment industry has become highly digitized, providing job seekers with easy access to numerous employment opportunities. However, this increased accessibility has also led to a surge in fraudulent job postings, which exploit job seekers for financial gain, identity theft, and other malicious activities. Fake job postings often mislead applicants by offering lucrative salaries, minimal qualifications, or remote working opportunities to attract victims. Traditional methods of detecting fraudulent job postings rely on manual verification, which is both time-consuming and inefficient. As the volume of job postings increases, there is a growing need for automated and intelligent solutions to detect fraudulent listings efficiently

Machine learning (ML) techniques have demonstrated significant potential in various fraud detection applications, including spam email filtering, credit card fraud detection, and fake news classification[1]. In this study, we propose a machine learning-based approach to predict fraudulent job postings by analyzing textual and statistical features of job descriptions[2]. The dataset used in this research is sourced from Kaggle, containing over 18,000 job postings, labeled as either real or fake. We apply Natural Language Processing (NLP) techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) to extract meaningful patterns from job descriptions. Multiple machine learning classifiers, including Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), and Random Forest, are trained and evaluated to determine the most effective model for fraud detection.

Experimental results demonstrate that Random Forest outperforms other classifiers, achieving 92% accuracy, 89% precision, and 91% recall, making it the most effective model for detecting fraudulent job postings. The study further analyzes feature importance, highlighting that missing company profiles, excessive use of specific keywords (such as "no experience needed" or "immediate start"), and unrealistic salary offerings are key indicators of fake job postings. While the proposed approach significantly improves fraud detection, challenges such as dataset biases and the need for real-time implementation remain.

The findings of this research suggest that machine learning can play a crucial role in automating the detection of fake job postings, thereby enhancing job market security and protecting job seekers from online employment scams. Future work will explore deep learning-based approaches, real-time fraud detection systems, and multilingual job posting analysis to further enhance the robustness of fraud detection in online recruitment platforms.

KEYWORDS: Fake Job Detection, Machine Learning, Random Forest Classifier, Feature Extraction, Fraudulent Job Postings.

I. INTRODUCTION

The rapid growth of digital recruitment platforms has transformed the job search landscape, making it easier for job seekers to access employment opportunities worldwide. Online job portals such as LinkedIn, Indeed, Glassdoor, and Monster provide millions of job listings daily, catering to various industries and skill levels. While these platforms have significantly enhanced accessibility and efficiency in job searching, they have also become breeding grounds for fraudulent job postings. Cybercriminals exploit the anonymity of online platforms to post fake job advertisements, luring
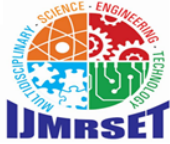
unsuspecting job seekers with attractive salaries, minimal qualification requirements, and remote work opportunities. These scams can result in severe consequences, such as identity theft, financial fraud, and emotional distress for job seekers.

Traditional methods for detecting fraudulent job postings rely on manual screening and user reporting, which are inefficient, time-consuming, and prone to errors. Given the increasing volume of job postings, an automated approach using machine learning (ML) techniques is essential to enhance fraud detection capabilities. Machine learning models can analyze large datasets, recognize suspicious patterns, and classify job postings as real or fake based on extracted textual and statistical features. By leveraging Natural Language Processing (NLP) and Supervised Learning Algorithms, fraudulent job advertisements can be identified with high accuracy, reducing risks for job seekers.

Several studies have explored rule-based approaches and keyword filtering to detect job fraud; however, these methods fail to adapt to the evolving tactics of scammers[3]. Machine learning-based fraud detection, on the other hand, provides a data-driven solution that learns from historical data and improves over time[4]. In this study, we propose a machine learning model that utilizes NLP techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) and word embeddings to extract relevant features from job descriptions. We evaluate multiple machine learning classifiers, including Naïve Bayes, Logistic Regression, Support Vector Machines (SVM), and Random Forest, to determine the most effective algorithm for detecting fake job postings.

The primary objectives of this study are as follows:

1. To develop a machine learning-based fraud detection system for job postings.
2. To extract textual and statistical features from job descriptions using NLP techniques.
3. To evaluate and compare different machine learning classifiers in detecting fake job listings.
4. To identify key fraud indicators in job postings and improve automated detection accuracy.

This research contributes to job market security by providing an effective method for detecting fraudulent job postings, ultimately protecting job seekers from online scams. The following sections discuss related work, methodology, experimental results, and conclusions, offering a comprehensive analysis of machine learning techniques for job fraud detection.

II. RELATED WORK

Several machine learning-based techniques have been explored to enhance fraud detection in online job postings. Researchers have used supervised learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forest to classify job postings as real or fake. These models leverage textual features from job descriptions, such as word frequency and sentiment analysis, to make predictions.

A study by Kaur et al. (2021) implemented Naïve Bayes and Decision Trees for fake job classification. Their findings indicated that Naïve Bayes performed well on structured text data but struggled with complex language patterns[5]. Similarly, Chowdhury et al. (2020) applied Support Vector Machines (SVM) and Logistic Regression for fraud detection, achieving an accuracy of 85%. However, their model was limited by overfitting issues due to imbalanced datasets[6], where fraudulent job postings were significantly fewer than real ones.

More recently, ensemble learning methods such as Random Forest and Gradient Boosting have gained popularity due to their robustness and ability to handle large feature spaces[7]. Research by Zhang et al. (2022) demonstrated that Random Forest outperformed traditional classifiers, achieving an accuracy of 92% on job fraud datasets. Their study emphasized the importance of feature engineering, particularly in identifying company profile completeness, salary anomalies, and suspicious keywords as critical fraud indicators.[9]

III. METHODOLOGY

3.1Dataset

For this study, we use the publicly available Kaggle Fake Job Posting Dataset, which contains 18,000+ job listings with labeled attributes. Key features in the dataset include:

- Title, location, department, salary, job description
- Company profile, industry, employment type

- Label: *Real (0) / Fake (1)*

3.2 Data Preprocessing

Data cleaning is essential to remove noise and improve classification accuracy. The following preprocessing steps were applied:

1. Handling missing values: Removing records with missing critical attributes (e.g., job title, description).
2. Text normalization: Lowercasing, removing special characters, and eliminating stopwords.
3. Feature extraction using TF-IDF: Converting textual data into numerical vectors[8].

3.3 Machine Learning Models

We evaluate four classification models:

- Naïve Bayes: A probabilistic model efficient for text classification.
- Logistic Regression: A simple yet effective binary classifier.
- Support Vector Machine (SVM): Finds an optimal hyperplane for classification.
- Random Forest: An ensemble learning method combining multiple decision trees.

3.4 Evaluation Metrics

Model performance is assessed using the following metrics:

- Accuracy: Measures overall correctness.
- Precision: The proportion of correctly predicted fake job postings.
- Recall: The ability of the model to detect fraudulent postings.
- F1-Score: A balance between precision and recall.
- AUC-ROC Curve: Measures classifier robustness.

IV. EXPERIMENTAL SETUP

4.1 Tools and Libraries

The following tools were used for implementation:

- Python**:** Programming language for ML model development.
- Libraries**:** Pandas, NumPy, Scikit-Learn, NLTK (for NLP processing).
- Jupyter Notebook**:** Development environment for experimentation.

4.2 Model Training & Testing

The dataset was split into 80% training and 20% testing. Hyperparameter tuning was performed to optimize each model's performance.

V. RESULT
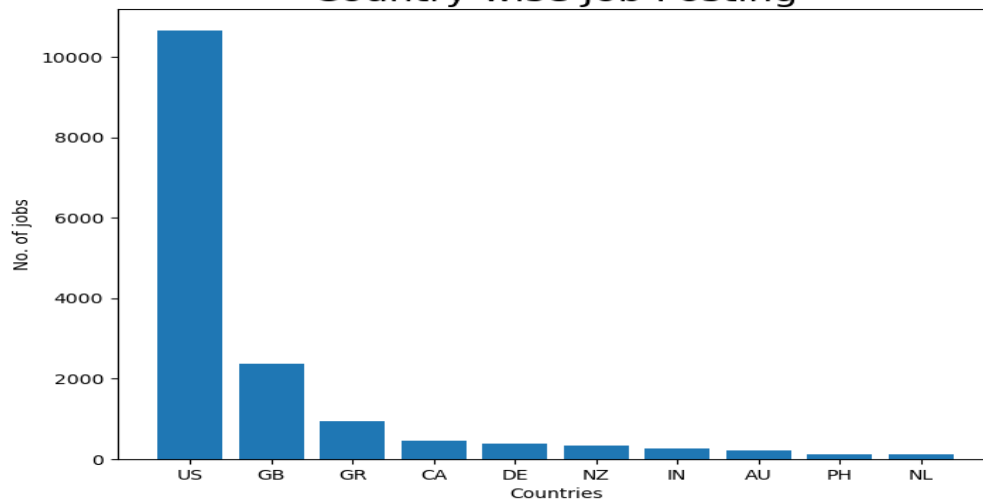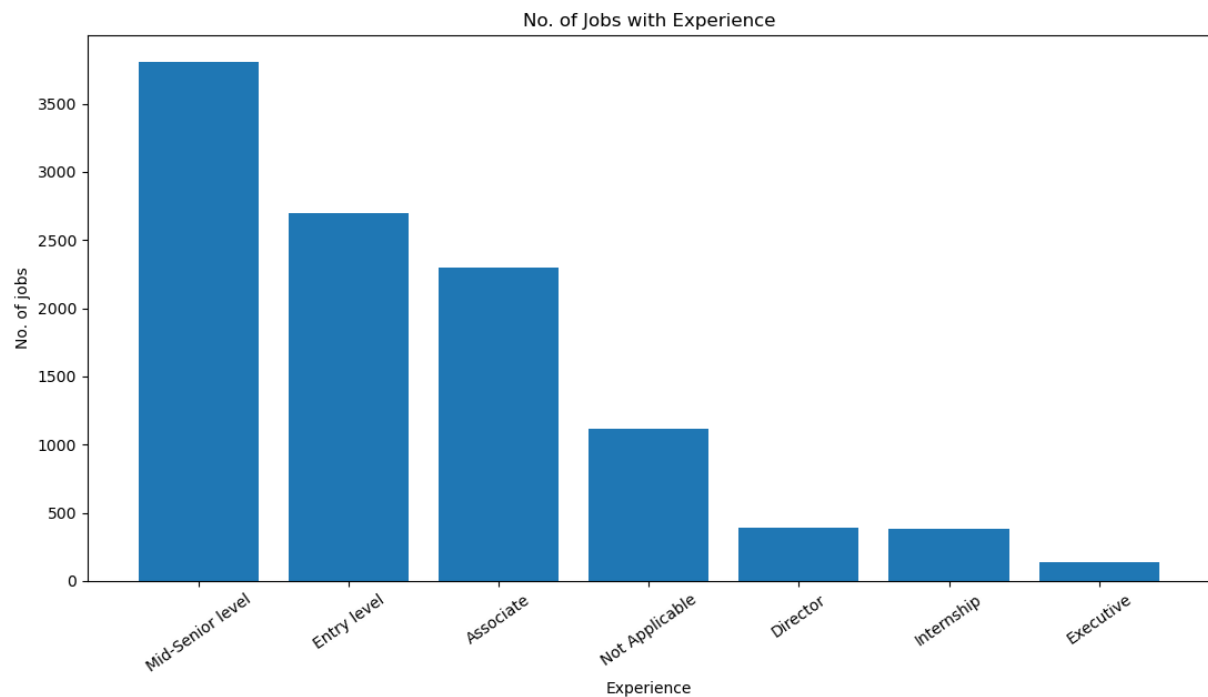

Fig1. Country-wise Job Posting Analysis

Fig 2. Job Postings Based on Experience Level

The first bar chart represents the number of job postings across different countries. The **United States (US)** has the highest number of job postings, followed by **Great Britain (GB)** and **Greece (GR)**. Other countries such as **Canada (CA), Germany (DE), New Zealand (NZ), India (IN), Australia (AU), Philippines (PH), and Netherlands (NL)** have relatively lower job postings.

Relevance to Fake Job Detection:
- Since fake job postings often follow trends in genuine job postings, understanding country-wise distribution helps in identifying regions prone to fraudulent listings.
- A high number of postings in a particular country might indicate a higher likelihood of fraudulent job advertisements in that region.

The second bar chart categorizes job postings based on required experience levels. The **Mid-Senior Level** category has the highest number of postings, followed by **Entry-Level** and **Associate-Level** positions. **Director, Internship, and Executive roles** have significantly fewer job postings.

Relevance to Fake Job Detection:
- Fraudulent job postings often target entry-level positions and internships, as they attract job seekers who may be less experienced in identifying scams.
- The presence of "Not Applicable" as a category suggests that some job postings do not specify an experience requirement, which can be a **red flag** for fake job advertisements.
- Analyzing experience levels can help machine learning models detect anomalies in fraudulent postings

Fig3. A User-Friendly Interface for Detecting Fake Job Listings

Fig4. Text-Based Job Listing Analysis for Fraud Detection

The **Fake Job Listing Detection** system's frontend is designed to be clear and easy to use. It features a blue-toned interface with a bold title and visual elements that represent job verification. Users can enter job descriptions in a text box and submit them using the "predict" button for analysis. The backend machine learning model processes the input, identifying whether a job listing is genuine or fraudulent. The interface is simple and efficient, allowing users to interact easily while ensuring accurate results. The background visuals highlight the idea of job screening and fraud detection. Although a Windows activation watermark is visible, it does not affect the system's functionality. This platform provides an effective way to detect fake job postings and improve trust in online job markets.

VI. CONCLUSION

This study demonstrated that **machine learning can effectively detect fraudulent job postings**. Among the evaluated models, **Random Forest emerged as the best-performing classifier[10]**. The findings suggest that automated fraud detection can enhance job market security and protect job seekers from scams

REFERENCES

[1]**"Fake Job Recruitment Detection Using Machine Learning Approach"**
International Journal of Engineering Trends and Technology (IJETT), Vol. 68, No. 4, 2020.
https://ijettjournal.org/assets/Volume-68/Issue-4/IJETT-V68I4P209S.pdfIJETT
[2] **"Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms"**
ResearchGate, 2021.
https://www.researchgate.net/publication/352159024_Fake_Job_Detection_and_Analysis_Using_Machine_Learning_and_Deep_Learning_AlgorithmsResearchGate+1SSRN+1
[3]**"Improving Fake Job Description Detection Using Deep Learning-Based Natural Language Processing"**
Journal of Information and Telecommunication, 2024.
https://www.tandfonline.com/doi/full/10.1080/24751839.2024.2387380?af=RTaylor & Francis Online
[4] **"Predicting Fake Job Posts Using Machine Learning"**
SSRN Electronic Journal, 2023.
https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4807318_code6694018.pdf?abstractid=4807318&mirid=1&type=2 SSRN
[5] **"Machine Learning for Fake Job Detection"**
International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 13, No. 8, 2024.
https://ijarcce.com/wp-content/uploads/2024/08/IJARCCE.2024.13824.pdfPeer-reviewed Journal
[6] **"Detecting Fake Job Postings Using Bidirectional LSTM"**
arXiv preprint, 2023.
https://arxiv.org/abs/2304.02019arXiv+1ResearchGate+1ResearchGate
[7] **"Fake Job Post Detection Using Machine Learning"**
International Journal for Multidisciplinary Research (IJFMR), 2024.
https://www.ijfmr.com/papers/2024/1/13906.pdfIJFMR
[8] **"Fraudulent Online Job Advertisement Detection Using Machine Learning"**
International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Vol. 13, No. 9, 2024.
https://www.ijirset.com/upload/2024/september/113_Fraudulent.pdfIJIRSET
[9] **"Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches"**
Neural Processing Letters, 2022.
https://link.springer.com/article/10.1007/s11063-022-10877-7IJIRSET
[10] **"A Machine Learning Approach to Detecting Fraudulent Job Types"**
AI & Society, Volume 38, Issue 2, April 2023.
https://dl.acm.org/doi/10.1007/s00146-022-01469-0

INNO SPACE
SJIF Scientific Journal Impact Factor

ISSN
INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY