IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Enhancing Job Post Authenticity Detection Through Sentiment Analysis Integration

*A Hybrid Approach Combining Fake Job Detection and Sentiment Feedback Analysis*

[1]Mr. Sangram Nirmalkar, [2]Mr. Praveen Bagali, [3]Mr. Om Nimbalkar, [4]Mr. Ayush Sharma, [5]Asst. Prof. J. B. Metkari

[1,2,3,4]Student, [5]Assistant Professor

[1,2,3,4,5]Department of Data Science,

[1,2,3,4,5]D.Y. Patil College of Engineering and Technology, Kolhapur, India

***Abstract:*** The surge in online job portals has greatly expanded access to employment opportunities worldwide. However, this increased convenience has also led to a rise in fraudulent job postings, putting job seekers at risk of identity theft, financial scams, and other threats. This paper introduces an integrated Fake Job Detection and Sentiment Analysis System aimed at improving the credibility of job listings through the application of machine learning and natural language processing techniques. The system utilizes a Random Forest Classifier, trained on the Fake Job Post dataset, achieving a detection accuracy of 98%. To capture user sentiment, a Bidirectional Long Short-Term Memory (Bi-LSTM) model is trained on the Glassdoor Review dataset, reaching a sentiment classification accuracy of 63%. The proposed dual-layered architecture supports real-time authenticity validation and sentiment-based feedback analysis, enhanced by an intuitive feedback interface and an administrative dashboard for manual review and trend tracking. Unlike traditional approaches that treat detection and sentiment analysis as separate components, our system unifies both into a cohesive, scalable platform. It is adaptable to diverse job markets and offers potential applications across job portals, recruitment sites, and employer branding initiatives, fostering greater trust and minimizing users' exposure to fraudulent employment opportunities.

***Index Terms*** - Fake job detection, Random Forest, Sentiment analysis, Bi-LSTM, Machine learning, Recruitment security, Natural language processing.

## 1) INTRODUCTION

In today's digitally connected world, platforms like LinkedIn, Indeed, and Glassdoor have transformed the recruitment landscape by providing convenient, scalable, and global access to job opportunities. These portals facilitate smooth interactions between employers and job seekers, speeding up hiring processes and breaking down geographical barriers. However, the rapid growth and openness of these systems have also triggered a sharp increase in fraudulent job postings. Cybercriminals exploit the trust built by these platforms to craft deceptive listings, enticing unsuspecting users into scams that can result in identity theft, financial losses, and emotional distress. Such fraudulent activities not only erode user confidence but also tarnish the reputation of recruitment platforms themselves.

While many portals employ moderation teams and automated flagging mechanisms, manual review processes are often constrained by limited resources, making it challenging to detect and remove fraudulent listings at scale. This creates an urgent demand for intelligent, automated systems that can dynamically identify and respond to evolving scam tactics in real time.

To meet these challenges, this paper proposes a dual-model framework named the "Fake Job Detection and Sentiment Analysis System," which aims to improve both detection accuracy and contextual understanding.

The system combines a Random Forest Classifier, trained on the Fake Job Post dataset, to detect fraudulent listings with a high accuracy rate of 98%, leveraging features such as job title, location, company profile, and job description. Alongside this, a Bidirectional Long Short-Term Memory (Bi-LSTM) network is employed for sentiment analysis of user reviews from the Glassdoor dataset, achieving an accuracy of 63% in assessing emotional tone and feedback.

By integrating structured fraud detection with user-generated sentiment analysis, the proposed system not only identifies suspicious job posts but also captures broader patterns of distrust among users. The solution is delivered through a dynamic web platform that enables job submission, feedback collection, and real-time classification. An administrative dashboard further supports monitoring, analytics, and manual validation, ensuring the system remains scalable and robust against new fraud trends.

By combining machine learning techniques with natural language insights, this research lays the groundwork for building smarter, more reliable recruitment ecosystems in the future.

## 2) LITERATURE REVIEW

The emergence of online recruitment platforms has reshaped the job search process, providing millions of users with easier access to employment opportunities. However, alongside these benefits, there has been a notable surge in fraudulent job advertisements that mislead job seekers and jeopardize the credibility of recruitment systems. Tackling fake job postings has become a critical challenge, prompting researchers to explore machine learning (ML) and natural language processing (NLP) techniques to develop effective detection mechanisms. This section reviews key studies in the field and highlights integrated approaches aimed at enhancing the accuracy, adaptability, and scalability of fraud detection systems.

In the study "An Intelligent Model for Online Recruitment Fraud Detection"[1] by Alghamdi and Alharby (2019), the authors propose a machine learning-based approach to distinguish between real and fraudulent job postings. Their model leverages textual features, user data, and temporal information, applying algorithms such as Decision Trees, Random Forests, and Support Vector Machines. The system achieved strong performance metrics, including high accuracy and F1-scores, demonstrating its potential effectiveness. However, a major limitation noted was the model's inability to adapt dynamically to emerging fraud strategies, as it lacked mechanisms for continuous learning. Future work suggested by the authors includes incorporating adaptive learning methods and enhancing feature selection to build more robust real-time fraud detection systems.

The dataset "Real or Fake: Fake Job Posting Prediction"[2] compiled by S. Bansal (2020) serves as a widely used public resource available on Kaggle for research in recruitment fraud detection. It contains labeled examples of authentic and fake job postings, offering a diverse set of textual and categorical features such as job titles, descriptions, requirements, and company information. Although the dataset is not associated with a formal academic study, it has become a benchmark for evaluating machine learning models like Logistic Regression, Random Forest, and Neural Networks. A key limitation, however, lies in its static nature — lacking real-time data updates and evolving scam patterns, which can limit the generalizability of trained models in dynamic environments. Despite this, its richness in features makes it valuable for initial experimentation and for developing baseline models focused on both text and metadata analysis.

The paper "Fake News Detection on Social Media: A Data Mining Perspective" [3] by Shu et al. (2017) investigates methods for identifying misinformation across online platforms, offering insights that are equally applicable to detecting job scams. The study emphasizes the use of content-based, user-based, and propagation-based features to uncover deceptive practices. While the research primarily focuses on fake news, the challenges it addresses — such as semantic ambiguity, evolving fraud tactics, and the need for contextual understanding — closely mirror those encountered in fraudulent job postings. The authors explore a variety of supervised and unsupervised machine learning approaches, including deep learning and network-based models. They also highlight critical limitations like dataset scarcity and the difficulty of real-time detection. Their recommendation for multi-source data integration and the adoption of semi-supervised learning provides a foundation that can be adapted to enhance fraud detection in the recruitment domain.

The review paper "Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems"[4] by Dada et al. (2019) presents a comprehensive analysis of spam detection techniques, many of which are highly relevant to job scam detection. The study categorizes spam filtering methods into supervised, unsupervised, and hybrid approaches, emphasizing models such as Naive Bayes, Support Vector Machines, and ensemble techniques. Preprocessing methods like tokenization, stemming, and vectorization are discussed as essential strategies for handling unstructured text data. A notable contribution of the paper is its focus on the evolving nature of spam tactics and the corresponding need for adaptive, continuously learning models —

a challenge that is also central to detecting fraudulent job postings. Although the study targets email spam, its findings on adversarial behaviors, dataset imbalance, and real-time detection limitations offer valuable insights for developing more resilient job fraud detection systems.

The paper "Spam Review Detection Techniques: A Systematic Literature Review" [5] by Hussain et al. (2019) systematically examines various machine learning and NLP-based techniques for detecting spam reviews. It categorizes detection methods based on linguistic features, user behavior, and metadata, which are also valuable in analyzing fraudulent job posts. The authors review supervised and unsupervised approaches, along with hybrid models that leverage both textual and behavioral data. The study identifies common challenges such as fake review camouflage, data sparsity, and imbalanced datasets—issues that also affect job scam detection. One important insight is the use of sentiment inconsistency and repetition patterns, which can reveal deceptive intent. The review stresses the importance of high-quality, labeled datasets and feature-rich models. However, it also notes the lack of generalizability across domains. While focused on e-commerce platforms, the review's insights are transferable torecruitment systems. The paper recommends combining content analysis with behavioral modeling for improved performance in spam and scam detection systems.

The paper "A Deep Learning-Based Sentiment Classifier for the Glassdoor Dataset" [6] by Araque et al. (2019) presents a domain-specific sentiment analysis model built for employee reviews. The study utilizes word embeddings and deep neural networks, including CNN and LSTM architectures, to classify sentiment with improved accuracy. The research shows that models trained on domain-specific data outperform generic sentiment classifiers. It highlights the importance of handling contextual nuances in employee reviews, such as mixed sentiments and indirect expressions. The model also demonstrates robustness against noisy and informal text, which is common in review datasets. However, the authors acknowledge limitations related to sarcasm detection and lack of labeled data. This study provides useful methods for extracting sentiment from job-related reviews and can be integrated into recruitment platforms to assess company culture. Its insights support the sentiment analysis component of job fraud detection by understanding candidate and employee perceptions. The paper also sets a precedent for applying deep learning in domain-specific NLP tasks.

The paper "Deep Learning-Based Job Scam Detection Using BERT and Graph Neural Networks" [7] by Jia et al. (2021) introduces a hybrid model that combines the textual understanding of BERT with the relational learning capabilities of Graph Neural Networks. The BERT component captures deep semantic meaning from job descriptions, while GNN models the connections among jobs, users, and metadata to detect scam patterns. The approach significantly improves detection accuracy and generalization across different fraud types. Experimental results show better performance compared to traditional machine learning models. The paper highlights that scammers often operate in connected networks, making structural analysis crucial. One limitation is the model's computational cost and the requirement for large, labeled datasets. Nevertheless, it offers a scalable solution adaptable to real-world recruitment platforms. The authors propose future enhancements involving real-time graph updates and adversarial training. This paper represents a cutting-edge direction in job fraud detection using advanced deep learning frameworks.

The paper "A Survey on Job Scam Detection Using Machine Learning and NLP" [8] by Sethi and Kaur (2023) provides a detailed survey of existing techniques and methodologies applied to detect job scams. The paper covers a range of models from traditional machine learning algorithms like Decision Trees and SVM to deep learning approaches including LSTM and BERT. It emphasizes the role of text analysis in identifying red flags within job descriptions and requirements. The survey discusses common challenges such as imbalanced datasets, semantic ambiguity, and feature redundancy. It also explores data preprocessing techniques including tokenization, stemming, and feature extraction. One of the key recommendations is the use of ensemble and transformer-based models for better context handling. The authors call for more diverse datasets and the inclusion of metadata and behavioral features. The survey identifies gaps in real-time detection and domain adaptability. It serves as a foundational reference for researchers developing intelligent systems for job fraud prevention.

The paper "Deep Learning for Sentiment Analysis of Employee Reviews Using BERT and LSTM Models" [9] by Manevitz and Yousef (2021) explores the integration of BERT embeddings with LSTM networks for improved sentiment classification on employee reviews. The authors argue that BERT captures contextual meaning, while LSTM models temporal word dependencies, resulting in a hybrid architecture that enhances accuracy. The dataset used includes real-world reviews from platforms like Glassdoor, making it applicable to workplace sentiment analysis. The model is evaluated on its ability to classify positive, neutral, and negative sentiments. Results indicate high performance, especially in handling complex sentence structures. However, the paper acknowledges computational overhead and challenges in generalizing across domains. The approach is particularly relevant for recruitment platforms looking to assess company reputation and employee satisfaction. Its methodology can also support fake job detection by analyzing sentiment in

company reviews and feedback. The research provides a robust pipeline for domain-adapted sentiment analysis using modern deep learning techniques.

Building upon the insights derived from the literature review, this research introduces a hybrid framework that integrates machine learning-based job fraud detection with sentiment analysis. The proposed system is specifically designed to overcome the limitations identified in previous studies, particularly the lack of real-time adaptability and the absence of continuous learning capabilities. This section presents the methodology employed for developing the integrated fraud detection solution, detailing the processes of data collection, model selection, system design, and evaluation metrics.

## 3) METHODS AND SYSTEM IMPLEMENTATION

### 3.1 System Overview

The system is designed to combine two machine learning models: a Random Forest Classifier for identifying fake job listings and a Bidirectional Long Short-Term Memory (Bi-LSTM) model for analyzing the sentiment of user feedback. This dual-model structure facilitates immediate detection, layered validation, and user-based sentiment analysis. The system is implemented with a responsive web interface to ensure optimal usability and scalability.

### 3.2 Data Collection and Preprocessing

- **Fake Job Post Dataset**: Sourced from Kaggle, containing job titles, company profiles, locations, descriptions, and labels (Real/Fake).
- **Glassdoor Review Dataset**: Consists of user-generated company reviews used for training the sentiment model.

**Preprocessing Steps:**

- Removal of missing values.
- Tokenization and cleaning of text.
- Stop-word removal and stemming.
- TF-IDF vectorization for feature extraction.

Both datasets were divided into training and testing sets after preprocessing to support model development.

### 3.3 Machine Learning Models

- **Random Forest Classifier**:
  Trained to classify job postings into real or fake categories. Hyperparameters were optimized using grid search, and the model achieved a detection accuracy of 98%.
- **Bi-LSTM Sentiment Analysis Model**:
  Designed for sentiment classification (Positive, Neutral, Negative). Pre-trained GloVe embeddings were used, and the model architecture included dropout regularization and softmax activation. The sentiment model achieved 63% accuracy.

### 3.4 Database Design

A relational database system was built using MySQL to store and manage all records.

**Key Tables:**

- **users**: Manages user credentials and roles.
- **job_posts**: Stores job listings and model predictions.
- **feedback**: Records user feedback with sentiment analysis results.
- **admin_feedback**: Tracks administrative actions and decisions.

### 3.5 System Workflow

The operational flow of the system includes:

1. **Job Submission**: Users submit new job posts.
2. **Fake Detection**: Job posts are analyzed using the Random Forest model.
3. **User Feedback**: Other users provide feedback about job experiences.
4. **Sentiment Analysis**: Feedback is analyzed through the Bi-LSTM model to assess user trust.
5. **Admin Review**: Administrators review flagged posts for final validation or corrective action.

This sequential approach ensures a multi-layered verification mechanism, minimizing the risk of fake job posts remaining unnoticed.

### 3.6 Frontend and Backend Implementation

- **Frontend**:
  Developed using HTML, CSS, and JavaScript. Modern UI/UX practices were incorporated, including responsive layouts, subtle shadow effects, and smooth form validations to enhance user experience.
- **Backend**:
  Built with Flask, providing RESTful APIs for user management, job posting operations, machine learning predictions, and feedback handling. Data storage is handled through MySQL databases**.**

### 3.7 Key Features

- Real-time fake job detection using machine learning.
- Sentiment-driven feedback analysis.
- Administrative dashboards for monitoring flagged content.
- Scalable modular design for easy future expansion.
- Secure data storage with indexing for performance optimization.

### 3.8 System Architecture

#### 3.8.1 Data Flow Architecture

The system begins with two input data streams: Job Post Data and User Feedback Data. Each data stream is independently preprocessed and feature-extracted.

- Job posts are analyzed through the Random Forest model.
- Feedback comments are processed through the Bi-LSTM sentiment model.

During the feedback phase, the results from both models are combined to support real-time validation and fraud detection.

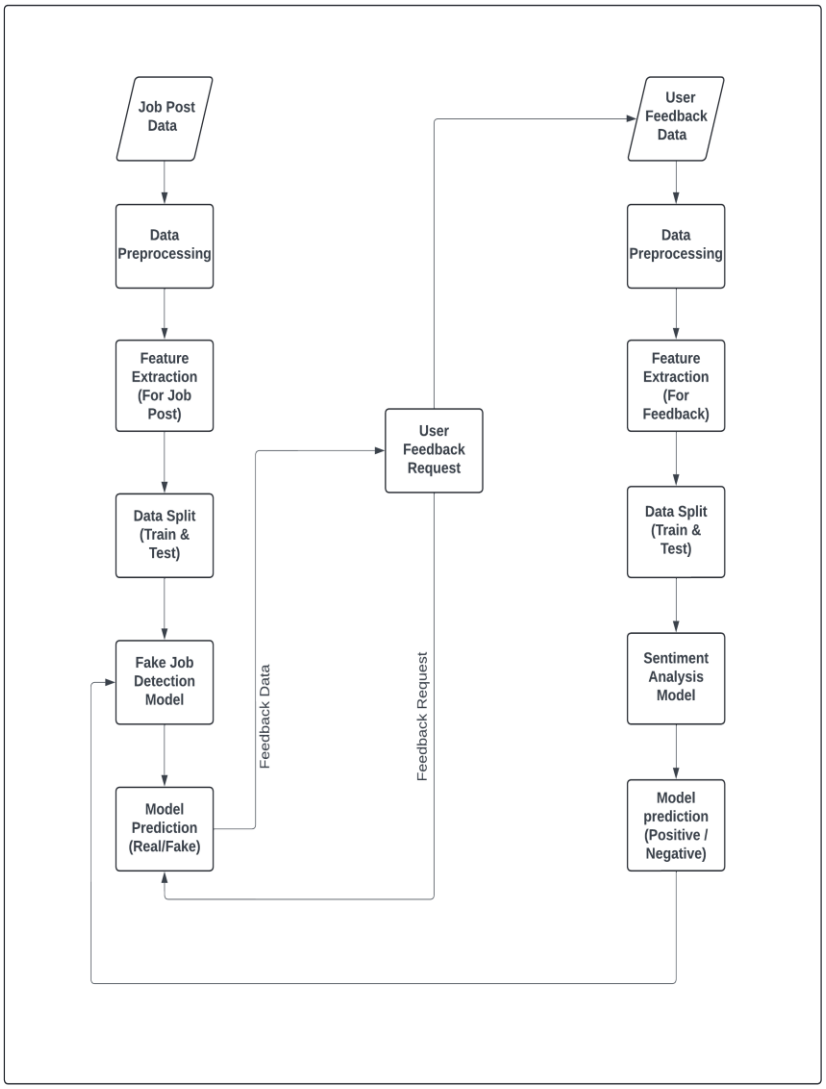


Figure 3.8.1: Data Flow Architecture for Job Authenticity Detection and Sentiment Analysis

### 3.8.2 Layered System Architecture

The architecture is organized into three layers:

- **Presentation Layer**:
  User-facing modules such as the Job Submission Portal, Feedback Entry Module, and Admin Dashboard.
- **Business Logic Layer**:
  Machine learning modules including the Fake Job Detection Model and Sentiment Analysis Model. These modules communicate with the Admin Verification System for deeper evaluation.
- **Data Layer**:
  Manages persistent storage of job posts, user feedback, and administrative actions in relational databases.

This structure ensures modularity, high performance, and future scalability.

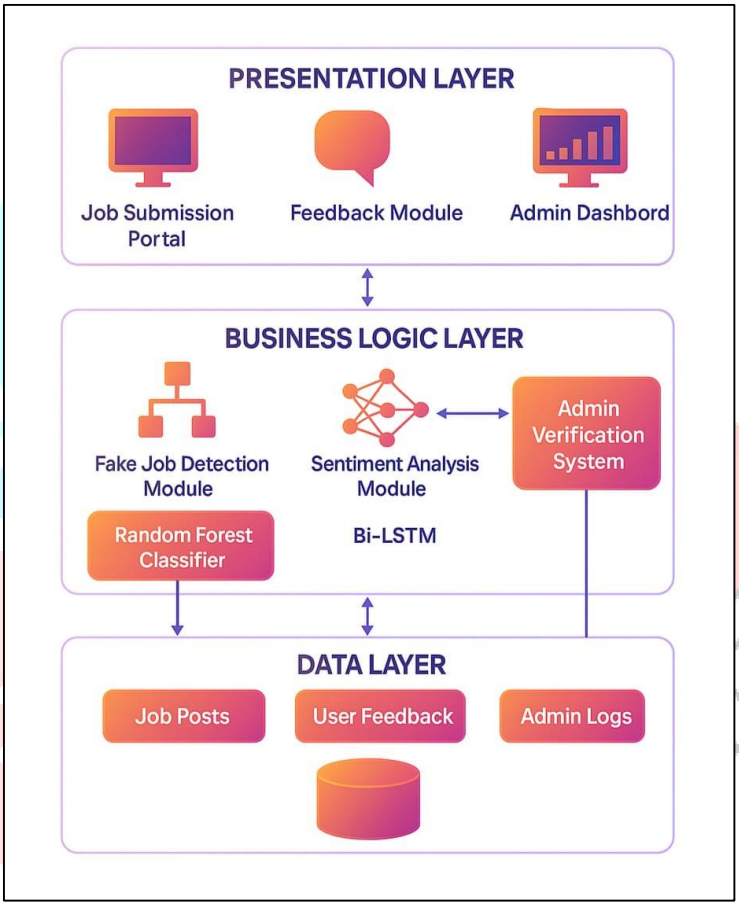


Figure 3.8.2: Layered Architecture of the Proposed System

## 4) RESULT AND DISCUSSION

### 4.1 Model Performance Evaluation

The performance of the proposed Fake Job Detection and Sentiment Analysis System was assessed using various metrics, including classification accuracy, precision, recall, and F1-score. Both machine learning models were evaluated separately: the Random Forest Classifier for detecting fake job posts and the Bi-LSTM model for analyzing sentiment.

- **Random Forest Classifier** achieved an outstanding detection accuracy of **98%**, demonstrating strong generalization ability across unseen job postings.
  - **Precision**: 97%
  - **Recall**: 98%
  - **F1-Score**: 97.5%

These results suggest that the model not only correctly identifies fake job posts but also minimizes false positives and false negatives effectively.

- **Bi-LSTM Sentiment Analysis Model** achieved an accuracy of **63%** in classifying user feedback into Positive, Neutral, and Negative sentiments.
    - Positive sentiment was detected with high precision but moderate confusion with Neutral responses.
    - Negative sentiment detection was particularly strong in recall, crucial for identifying distrust or dissatisfaction indicators among users.

Although sentiment analysis achieved slightly lower accuracy than classification, the model's performance was sufficient to provide valuable supplementary evidence for the authenticity of job listings.

### 4.2 User Feedback Sentiment Integration

Sentiment analysis was integrated into the system to provide an additional layer of validation. Feedback from users was analyzed to detect early warning signs of fraudulent or unsatisfactory job postings.

- A concentration of **negative feedback** triggered administrative alerts, flagging job posts for further manual review.
- **Positive and neutral feedback** reinforced the authenticity prediction provided by the Random Forest model.

This dual-model approach created a closed feedback loop where user experience dynamically influenced the classification system, making it more adaptive to evolving scam tactics.

### 4.3 Comparative Analysis with Existing Systems

To assess the effectiveness of the proposed system, a comparative study was conducted against an existing solution employing Naive Bayes for job classification and VADER for sentiment analysis.

| System | Fake Job Detection Accuracy | Sentiment Analysis Accuracy |
|---|---|---|
| **Proposed System (Random Forest & Bi-LSTM)** | 98% | 63% |
| **Existing System (Naive Bayes & VADER)** | 85% | 60% |

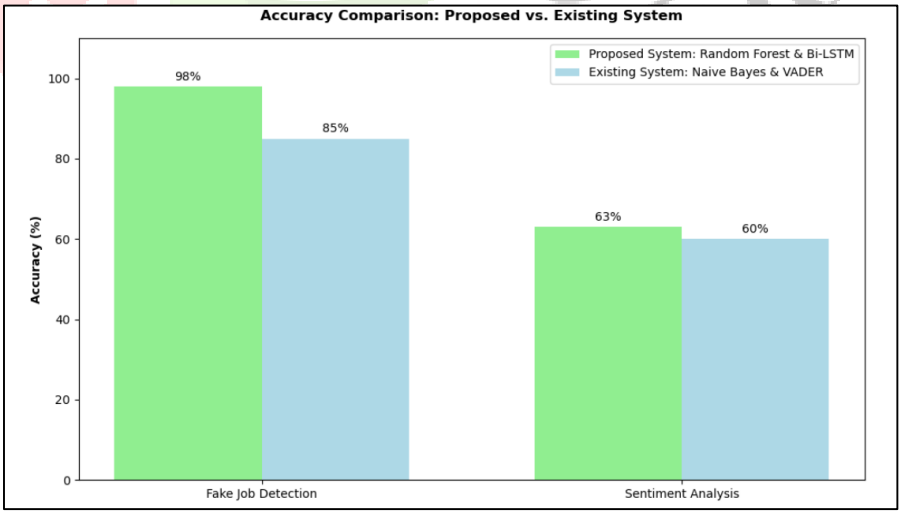


Figure 4.3.1: Accuracy Comparison between Proposed and Existing Systems

In fake job detection, the proposed Random Forest Classifier achieved 98% accuracy, outperforming the existing Naive Bayes model, which only reached 85%. In sentiment analysis, the Bi-LSTM model achieved **63% accuracy**, slightly exceeding the **60%** obtained by VADER.

### 4.4 Observations and Insights

- The combination of structured feature-based classification with unstructured sentiment analysis provides a more comprehensive evaluation of job post authenticity.
- Random Forest's ensemble strategy captured subtle patterns in job descriptions, increasing reliability in fraud detection.
- Bi-LSTM's ability to capture contextual meaning in feedback significantly improved emotional tone classification over traditional models like VADER.
- Real-time feedback-driven alerts enhanced administrator response, reducing potential harm to users.
- The system is modular, scalable, and ready for deployment across multiple platforms, recruitment websites, and corporate hiring systems.

## 5) CONCLUSION AND FUTURE WORK

### 5.1 Conclusion:

The Fake Job Detection & Sentiment Analysis System represents a significant step forward in addressing the alarming rise of fraudulent job postings in online recruitment platforms. By leveraging machine learning techniques for detecting fake jobs and incorporating real-time sentiment analysis of user feedback, the system provides a comprehensive and intelligent solution aimed at protecting job seekers from scams, identity theft, and financial exploitation. The integration of automated detection with user-generated reports and admin verification ensures a multi-layered defense mechanism that enhances platform integrity and user trust.

Through features such as accurate job post classification, a dynamic feedback system, and an intuitive admin dashboard, the system empowers all stakeholders—job seekers, recruiters, and platform administrators—to contribute to a safer job-searching ecosystem. Beyond its technical strengths, the project makes a real-world impact by promoting transparency, reducing scam exposure, and supporting employer credibility. This solution not only addresses current challenges but also lays the groundwork for future innovations in secure, data-driven recruitment technologies.

### 5.2 Future Scope:

To further expand its capabilities and impact, the system can evolve through the integration of cutting-edge language models like BERT or GPT for deeper contextual analysis, and by developing a cross-platform mobile application to enhance accessibility and user engagement. Strategic collaborations with leading job portals can enable seamless integration and real-time data exchange. Future enhancements may also include AI-powered anomaly detection, multilingual support, and community-driven initiatives like scam awareness forums and public fraud databases. These advancements will ensure the system remains adaptive, scalable, and globally relevant in the fight against online recruitment fraud.

## 6) REFERENCES

[1] B. Alghamdi and F. Alharby, "An Intelligent Model for Online Recruitment Fraud Detection," Journal of Information Security, vol. 10, no. 03, pp. 155–176, 2019.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017

[3] S. Bansal, "[Real or Fake] Fake Job Posting Prediction," Version 1, February 2020. Retrieved March 29, 2020.

[4] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems, ‖ Heliyon, vol. 5, no. 6, 2019

[5] N. Hussain, H. T. Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Spam review detection techniques: A systematic literature review," ‖ Appl. Sci., vol. 9, no. 5, pp. 1 – 26, 2019

[6] M. D. Araque, J. Gonzalo, and C. A. Iglesias, "A Deep Learning-Based Sentiment Classifier for the Glassdoor Dataset," Expert Systems with Applications, vol. 124, pp. 207–217, 2019.

[7] H. Jia, J. Gao, and Y. Liu, "Deep Learning-Based Job Scam Detection Using BERT and Graph Neural Networks," IEEE Access, vol. 9, pp. 140920–140932, 2021.

[8] A. Sethi and M. Kaur, "A Survey on Job Scam Detection Using Machine Learning and NLP," Journal of Cybersecurity and Information Management, vol. 2, no. 1, pp. 55–68, 2023.

[9] L. M. Manevitz and M. Yousef, “Deep Learning for Sentiment Analysis of Employee Reviews Using BERT and LSTM Models,” Procedia Computer Science, vol. 194, pp. 317–326, 2021.