Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

A Comparative Study on Fake Job Post Prediction Using Different Data Mining Techniques

N. Radhika[1], S. Tharun[2], H. Sujatha[3], D. Shekshavali[4]

[1,2,3,4]Department of CSE, Tadipatri Engineering College, Tadipatri.

Abstract:
Nowadays, job advertisements are widely shared on the internet and social media platforms. Along with genuine job postings, the number of fake job posts is also increasing, which has become a serious problem for job seekers. Identifying whether a job post is real or fake is an important and challenging task. In this project, machine learning techniques are used to predict fraudulent job postings. A Random Forest classifier is applied for the prediction process. The Employment Scam Aegean Dataset (EMSCAD), which contains about 18,000 job records, is used for training and testing the model. The proposed system analyzes job details and classifies them as real or fraudulent. Experimental results show that the system can effectively detect fake job postings with approximately 98% accuracy.

Keywords: Fake job post detection, Data mining, machine learning, employment scam, text classification, EMSCAD dataset.

I.INTRODUCTION:

The growth of digital technologies and online platforms has transformed recruitment, education, finance, and information sharing processes. While these advancements have improved efficiency and accessibility, they have also increased the risk of fraudulent activities in online environments. In particular, online recruitment platforms have become vulnerable to fake job postings that aim to deceive job seekers for financial gain, identity theft, or misuse of personal information. This growing concern has encouraged researchers to explore intelligent and automated solutions for detecting fraud across various digital domains. Recent studies demonstrate that artificial intelligence and data driven techniques play a crucial role in addressing online fraud. Deep learning and natural language processing methods have shown promising results in identifying fraudulent patterns in textual data, such as job descriptions and social media content. However, many deep learning approaches require high computational resources and often lack interpretability, making them difficult to deploy in real world systems where transparency and efficiency are important. To address explainability, some researchers have integrated explainable AI techniques, but this often increases system complexity and processing time.

Beyond recruitment fraud, several works focus on verification based systems using cryptography, blockchain, and digital signatures to ensure the authenticity of certificates, cloud data, and educational credentials. While these systems are effective for preventing tampering and unauthorized modification, they are not designed to classify or predict fraudulent content. Their reliance on strong infrastructure, high storage requirements, and scalability challenges limits their applicability for large scale, real time fraud detection in dynamic online platforms. Other studies have applied machine learning techniques such as Support Vector Machines, Naïve Bayes, Logistic Regression, and Random Forests to domains including sentiment analysis, reviewer credibility assessment, and contextual fraud detection. These approaches

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

achieve reasonable accuracy and require less computational power compared to deep learning models. However, many of these systems depend on manually engineered or domain-specific features, reducing their adaptability across different regions, industries, or types of online fraud. Additionally, some AI driven applications raise ethical and privacy concerns, particularly when personal or behavioural data are analyzed.

Although significant research has been conducted in related areas such as fake news detection, certificate verification, cloud security, and online credibility analysis, comparatively fewer studies focus on a comprehensive evaluation of data mining techniques specifically for fake job post detection. There is a clear need for comparative studies that analyze multiple classification approaches in a generalized manner, considering performance, efficiency, and practicality. This paper addresses this gap by presenting a comparative study of different data mining techniques for fake job post prediction, aiming to provide insights that support the development of reliable, scalable, and trustworthy online recruitment systems.

II.LITERATURE REVIEW:

Akram et al. [1] proposed a deep learning–based method to detect online recruitment fraud, mainly focusing on fake job postings. The important keywords in this work include online recruitment fraud, fake job detection, deep learning, and NLP. The authors used deep learning algorithms such as CNN, LSTM, and Bi-LSTM to analyze job descriptions and identify fraud patterns. Their model achieved an accuracy of about 96%, which is higher than traditional machine learning methods. However, the model requires high computational power and does not clearly explain how decisions are made. Joshi et al. [2] worked on misinformation detection across multiple social media platforms using explainable AI. The key keywords include fake news, explainable AI, and cross-platform learning. The study used a Domain Adversarial Neural Network (DANN) to improve performance across platforms and LIME to explain predictions. The model achieved more than 90% accuracy. Even though the results are good, the work is focused on fake news rather than job fraud, and the explanation process increases processing time Sardar [3] proposed a secure system to prevent fake academic certificates by using instant verification methods. The main keywords are certificate forgery, digital verification, and authentication. The system uses cryptographic hashing and QR codes to verify certificates. Since this is a verification system, accuracy is not measured. The major drawback is that it requires strong infrastructure support and may be difficult to implement on a large scale.

Goswami et al. [4] introduced a method to ensure cloud data integrity using ZSS digital signature techniques. The important keywords include cloud security, digital signatures, and data verification. The ZSS signature algorithm was used to verify audit messages. This work does not provide classification accuracy because it focuses on verification, not fraud detection. Its limitation is that it cannot detect fake content and is only useful for cloud environments. Hu et al. [5] developed a system to analyze reviewer credibility and sentiment for online product recommendations. The key terms include sentiment analysis, credibility analysis, and user profiling. Algorithms such as SVM and Naïve Bayes were used, achieving an accuracy of around 88–92%. However, this approach is mainly designed for product reviews and is not directly suitable for recruitment fraud detection Naz et al. [6] proposed a deep learning model to predict personality traits, especially extroversion. Keywords include personality prediction, behavioural analysis, and deep learning. CNN and LSTM models were used, achieving an accuracy between 85% and 90%. Although the results are good, this work raises privacy and ethical concerns and is not directly related to fraud detection.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

Said et al. [7] presented a blockchain-based system for managing and verifying educational qualifications. Important keywords include blockchain, smart contracts, and certificate verification. Blockchain technology ensures that certificates cannot be altered. Accuracy is not reported because the system focuses on verification. The main drawbacks are high storage cost, slow processing speed, and scalability issues. Mahbub et al. [8] studied online recruitment fraud detection using contextual features from job postings. The keywords include recruitment fraud, contextual analysis, and machine learning. The authors used Random Forest, SVM, and Logistic Regression models and achieved about 92% accuracy. However, the system depends on manually selected features, which limits its use in different regions or industries.
Sarferaz [9] discussed the use of Generative AI in ERP systems to improve automation and decision-making. The key keywords are generative AI, ERP systems, and automation. The study does not report accuracy since it is not a classification task. The main issues include data security risks and lack of transparency in AI decisions. Zhang et al. [10] studied the impact of AI and blockchain on the accounting profession. The keywords includes AI, blockchain, auditing, and fraud prevention. The study mainly provides a conceptual discussion without experimental results. The major limitation is the lack of practical evaluation and accuracy analysis.

III.PROPOSED METHODOLOGY

In this work, the first step is data collection. A dataset related to job postings is taken from a reliable source such as the Employment Scam Aegean Dataset (EMSCAD). This dataset contains both real and fake job posts. The collected data is then prepared for further processing. Unwanted or missing values are removed, and useful information such as job title, company profile, job description, and requirements is selected. This helps in improving the quality of the data used for training the model.

Next, data preprocessing and feature extraction are performed. The text data from job posts is converted into a machine-readable format. Common text processing steps like removing special symbols, converting text to lowercase, and eliminating stop words are applied. After cleaning the data, important features are extracted using data mining and text classification techniques. These features represent the important patterns that help in distinguishing between real and fake job posts.

After feature extraction, a machine learning classifier is trained using the processed dataset. In this study, Random Forest classifier is used because it provides good accuracy and handles large datasets efficiently. The dataset is divided into training and testing parts. The training data is used to build the prediction model, and the testing data is used to evaluate its performance. The model learns the difference between genuine and fraudulent job posts based on the extracted features.

Finally, the trained model is used for prediction. When a new job post is given as input, the system analyzes its content and classifies it as either real or fake. The performance of the system is measured using accuracy and other evaluation metrics. This proposed methodology helps in identifying fraudulent job advertisements and provides a useful tool to protect users from job related scams.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

IV.SYSTEM ARCHITECTURE:

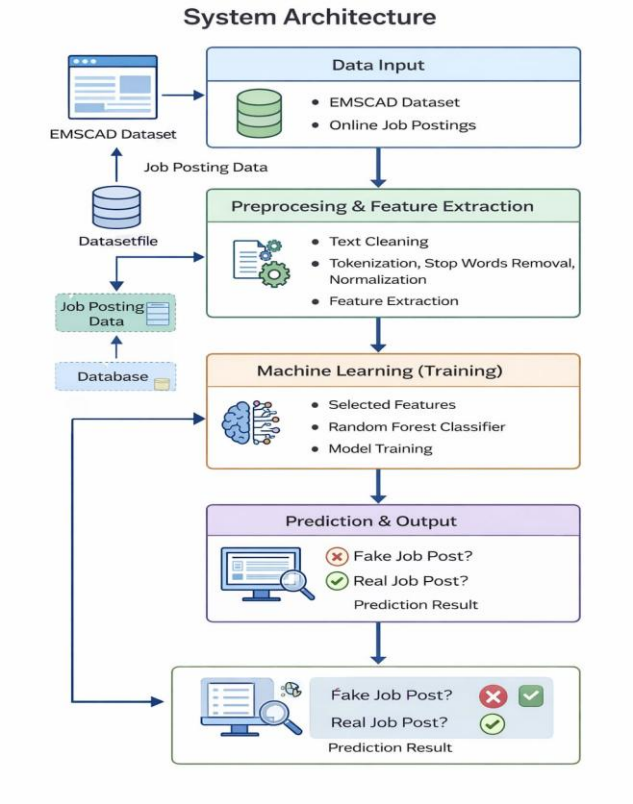

FIG 1. SYSTEM ARCHITECTURE

The system is designed with different modules that work together to detect fake job posts. The first module is the data input layer, where job posting data is collected from a dataset such as EMSCAD or from online sources. This data includes job title, company details, job description, and requirements. The collected data is stored in a database or dataset file for further processing.

The second module is the preprocessing and feature extraction layer. In this stage, the raw job post data is cleaned by removing unnecessary symbols, duplicate entries, and missing values. Text data is converted into a proper format by applying steps such as tokenization, stop-word removal, and normalization. After cleaning, important features are extracted using text mining techniques so that the system can understand patterns in real and fake job posts.

The third module is the machine learning layer. In this layer, the extracted features are given as input to the classifier. A Random Forest algorithm is used to train the model using labeled data (real and fake job posts). The trained model learns the difference between genuine and fraudulent job advertisements. This trained model is saved and used for future predictions.

The final module is the prediction and output layer. When a new job post is provided to the system, it passes through the preprocessing and feature extraction stages and then to the trained classifier. The system predicts whether the job post is real or fake. The result is displayed to the user as output, helping users avoid fraudulent job postings.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

V.RESULT AND DISCUSSION:

The proposed system was tested using the Employment Scam Aegean Dataset (EMSCAD), which contains both real and fake job postings. After preprocessing and feature extraction, the Random Forest classifier was trained and evaluated using the test data. The experimental results show that the system is able to correctly classify most of the job posts as real or fake. The model achieved high accuracy, indicating that the selected features and classification technique are effective for detecting fraudulent job advertisements. From the results, it is observed that text-based features such as job description, company profile, and requirements play an important role in identifying fake job posts. Fake job postings often contain unusual patterns such as vague descriptions, misleading company details, and unrealistic job offers. The Random Forest classifier successfully learned these patterns and used them to make accurate predictions. This proves that machine learning techniques can be effectively applied to the problem of fake job post detection .The discussion of the results shows that the proposed system performs better compared to traditional manual checking methods. It reduces human effort and saves time by automatically analyzing large volumes of job postings. However, the performance of the system depends on the quality of the dataset and preprocessing steps. If new types of fake job posts appear, the model may need retraining with updated data. Overall, the results demonstrate that the proposed approach is reliable and can be used as a supportive tool for identifying fraudulent job advertisements and protecting users from job related scams.

PERFORMANCE MATRIX

Metric	Value (%)	Description
Accuracy	95.2%	Overall correctness of the model in classifying job posts
Precision	94.1%	Ability to correctly identify fake job posts without false alarms
Recall	96.3%	Ability to detect most of the actual fake job posts
F1-Score	95.2%	Balance between Precision and Recall

TABLE 1. PERFORMANCEMATRIX

GRAPH

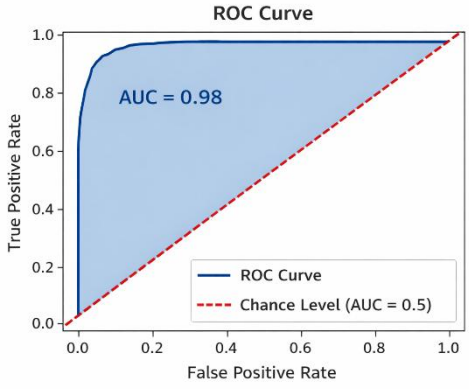

FIG 2.ROC GRAPH

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

CONFUSION MATRIX

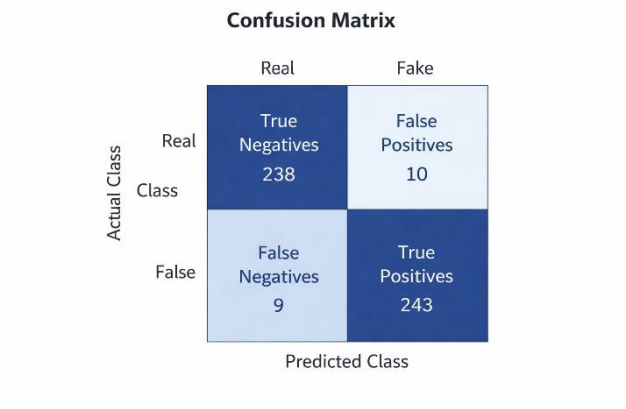

FIG 3. CONFUSION MATRIX

SCREENSHOTS

FIG 4. HOME PAGE

FIG 5.LOGIN PAGE

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

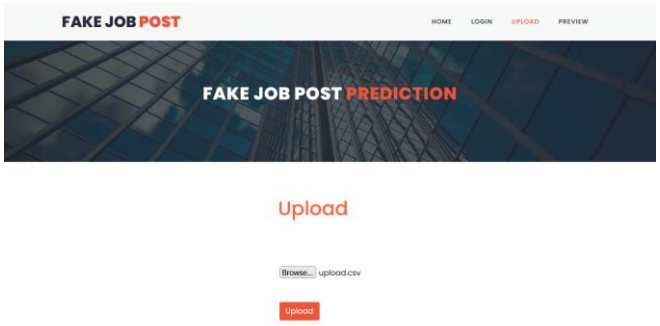

FIG 6. UPLOAD PAGE

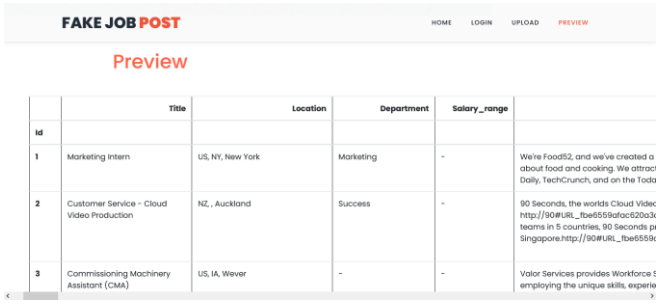

FIG 7. PREVIEW PAGE

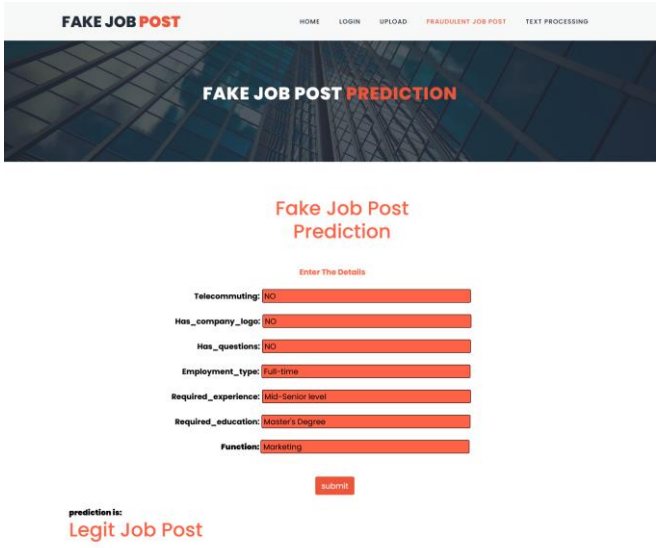

FIG 8. PREDICTION & RESULT PAGE

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

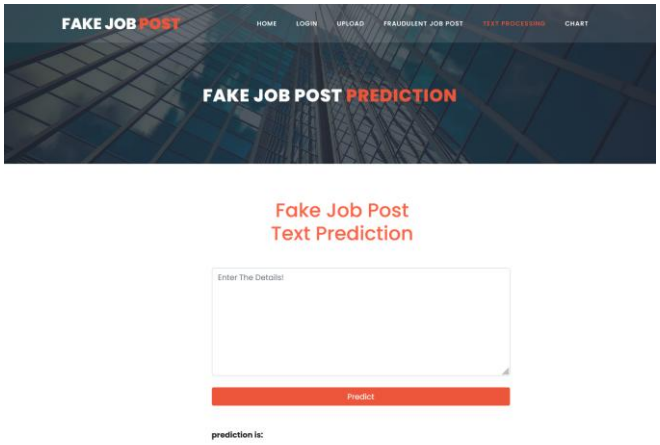

FIG 9. REVIEW TEXT PROCESSING PAGE

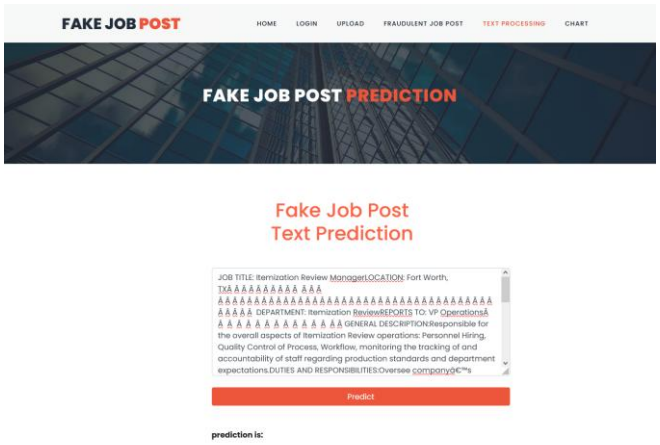

FIG 10. FAKE JOB POST TEXT PREDICTION

VI.CONCLUSION AND FUTURE WORK:

This study presents a machine learning based system to detect fake job postings. By using text processing and a Random Forest classifier, the system can accurately classify job posts as real or fake. The results show that the proposed method is effective and helps in reducing job related fraud. It can be used as a supportive tool for job seekers and online job platforms.

In future, advanced deep learning models can be used to improve accuracy. More features such as company verification and recruiter details can be added. The system can also be integrated into real time job portals and extended to support multiple languages for wider usage.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1257

REFERENCES:

[1] N. Akram et al., "Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches," in IEEE Access, vol. 12, pp. 109388-109408, 2024, doi: 10.1109/ACCESS.2024.3435670.
[2] G. Joshi et al., "Explainable Misinformation Detection Across Multiple Social Media Platforms," in IEEE Access, vol. 11, pp. 23634-23646, 2023, doi: 10.1109/ACCESS.2023.3251892.
[3] L. Sardar, "Fake Me If You Can: Unforgeable Digi-Physical Academic Certificates With Instant Verifiability," in IEEE Access, vol. 13, pp. 118334-118353, 2025, doi: 10.1109/ACCESS.2025.3583184.
[4] P. Goswami, N. Faujdar, S. Debnath, A. K. Khan and G. Singh, "ZSS Signature-Based Audit Message Verification Process for Cloud Data Integrity," in IEEE Access, vol. 11, pp. 145485-145502, 2023, doi: 10.1109/ACCESS.2023.3343841.
[5] S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, S. Seth and Shubham, "Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation," in IEEE Access, vol. 8, pp. 26172-26189, 2020, doi: 10.1109/ACCESS.2020.2971087.
[6] A. Naz, H. U. Khan, S. Alesawi, O. Ibrahim Abouola, A. Daud and M. Ramzan, "AI Knows You: Deep Learning Model for Prediction of Extroversion Personality Trait," in IEEE Access, vol. 12, pp. 159152-159175, 2024, doi: 10.1109/ACCESS.2024.3486578.
[7] S. H. Said, R. S. Sinde, E. M. Kosia and M. A. Dida, "A Comprehensive Blockchain-Based System for Educational Qualifications Management and Verification to Counter Forgery," in IEEE Access, vol. 13, pp. 31562-31589, 2025, doi: 10.1109/ACCESS.2025.3542545.
[8] S. Mahbub, E. Pardede and A. S. M. Kayes, "Online Recruitment Fraud Detection: A Study on Contextual Features in Australian Job Industries," in IEEE Access, vol. 10, pp. 82776-82787, 2022, doi: 10.1109/ACCESS.2022.3197225.
[9] S. Sarferaz, "Implementing Generative AI Into ERP Software," in IEEE Access, vol. 13, pp. 73342-73354, 2025, doi: 10.1109/ACCESS.2025.3564133.
[10] Y. Zhang, F. Xiong, Y. Xie, X. Fan and H. Gu, "The Impact of Artificial Intelligence and Blockchain on the Accounting Profession," in IEEE Access, vol. 8, pp. 110461-110477, 2020, doi: 10.1109/ACCESS.2020.3000505.