# FAKE JOB POSTING DETECTION

[1] P.Anusha, [2] Nukala Kavya, [3] Thumma Amulya
[1]AssistantProfessor, [23]Students
Department of Computer Engineering(Software Engineering)
Siddhartha Institute of Technology & Sciences, Narapally
anushaparvathagiri@siddhartha.org.in, 23tq1a5631@siddhartha.co.in, 23tq1a5626@siddhartha.co.in,

## Abstract

Fake job postings have become a significant issue on online job portals and social media platforms, where fraudulent recruiters exploit job seekers by collecting personal information, demanding registration fees, or conducting financial scams. These activities not only result in financial loss but also reduce trust in online recruitment systems. Therefore, detecting fraudulent job advertisements is essential to ensure a safe and reliable job search environment.

This project focuses on developing a machine learning-based system to identify whether a job posting is genuine or fake. The model analyzes various attributes of job listings, including job title, company profile, job description, location, salary details, and other relevant features. These data points are processed and used to train classification models capable of distinguishing between legitimate and fraudulent postings.

Data preprocessing techniques such as cleaning, feature extraction, and text analysis are applied to improve model performance. Machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, and Naïve Bayes are utilized to build and evaluate the predictive model. The system learns patterns and characteristics commonly associated with fake job postings and uses them to detect suspicious or misleading content.

## I. Introduction

In recent years, online job portals and social media platforms have become the primary sources for job searching and recruitment. While these platforms provide convenience and accessibility, they have also created opportunities for fraudulent activities. Fake job postings are increasingly being used by scammers to deceive job seekers by offering attractive job opportunities that do not actually exist.

These fraudulent postings often aim to collect sensitive personal information, demand registration or processing fees, or carry out financial scams. As a result, job seekers may face financial losses, identity theft, and emotional distress. Moreover, the presence of fake job advertisements reduces the credibility and trustworthiness of online recruitment platforms.

Traditional methods of detecting fake job postings rely on manual verification or simple rule-based systems, which are time-consuming and not efficient for handling large volumes of data. These methods also fail to identify sophisticated fraudulent patterns used by scammers.

To address this issue, Machine Learning (ML) and Natural Language Processing (NLP) techniques offer an effective solution. By analyzing patterns, textual content, and various features of job postings, ML models can automatically classify job advertisements as genuine or fraudulent.

## II. Literature Survey

In recent years, online recruitment platforms have become one of the most popular ways for job seekers to find employment opportunities. However, the rapid growth of these platforms has also led to a significant increase in fraudulent job advertisements. To address this issue, several researchers have explored the use of data analysis and machine learning techniques to detect fake job postings effectively.

Many studies have focused on analyzing large job posting datasets to identify patterns and characteristics associated with fraudulent advertisements. Researchers have found that fake job postings often include incomplete or missing company information, unrealistic salary offers, vague or poorly written job descriptions, and requests for sensitive personal or financial details. These patterns serve as key indicators for detecting suspicious job listings.

Various machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, and Naïve Bayes have been widely used for classification tasks in this domain. These models are trained on labeled datasets to distinguish between genuine and fake job postings based on extracted features. Additionally, Natural Language Processing (NLP) techniques are applied to analyze textual content and improve model accuracy by capturing semantic meaning.

Recent research also explores advanced approaches such as ensemble methods and deep learning models to enhance detection performance. These methods provide better accuracy and robustness by combining multiple models or learning complex patterns from data.

## III. System Analysis

The system analysis focuses on understanding how fake job postings are currently identified and how machine learning can improve detection. Traditional systems rely on manual verification or simple rule-based filtering, which are not effective for large-scale data. These systems fail to detect complex fraud patterns and cannot handle the growing number of job postings. To overcome these limitations, the proposed system uses Machine Learning (ML) and Natural Language Processing (NLP) techniques. It analyzes various features such as job title, description, salary, and company details to classify postings. The system processes unstructured text data and converts it into meaningful numerical features. It then applies classification algorithms to identify fake or genuine jobs. The model continuously improves with more data, making it adaptable.

### Existing System

The existing system for detecting fake job postings mainly depends on manual verification and basic rule-based approaches. Job portals often rely on moderators to

review job listings, which is time-consuming and inefficient. Some systems use keyword-based filtering to detect suspicious content, but these methods lack accuracy. They cannot understand the context or meaning of the job description. Additionally, existing systems use fixed rules that do not adapt to new fraud patterns. As the number of job postings increases, manual monitoring becomes difficult to manage. These systems also struggle to analyze large datasets in real time. Many fraudulent postings bypass detection due to weak filtering mechanisms. Overall, the existing system is limited, less accurate, and not suitable for modern large-scale online recruitment platforms.

**Disadvantages of Existing System**

- Time-consuming manual verification process
- Low accuracy in detecting fake job postings
- Cannot handle large volumes of data efficiently
- Lack of adaptability to new fraud techniques
- Poor understanding of text context and semantics

**Proposed System**

The proposed system uses Machine Learning and NLP techniques to automatically detect fake job postings. It collects job data including job title, company profile, description, salary, and location. The system performs data preprocessing such as cleaning, tokenization, and feature extraction to prepare the data. Textual information is converted into numerical features using techniques like TF-IDF. Machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Naïve Bayes are used to train the model. The system learns patterns and characteristics of fraudulent job postings. It can classify new job advertisements as real or fake with high accuracy. The system is scalable and can handle large datasets efficiently. It can also be integrated into job portals for real-time detection. Overall, it provides a reliable and automated solution.

**Advantages of Proposed System**

- High accuracy in detecting fake job postings
- Automated and time-efficient process
- Handles large datasets effectively
- Learns and adapts to new fraud patterns
- Better understanding of text using NLP
- Reduces dependency on manual verification

## IV. Methodology

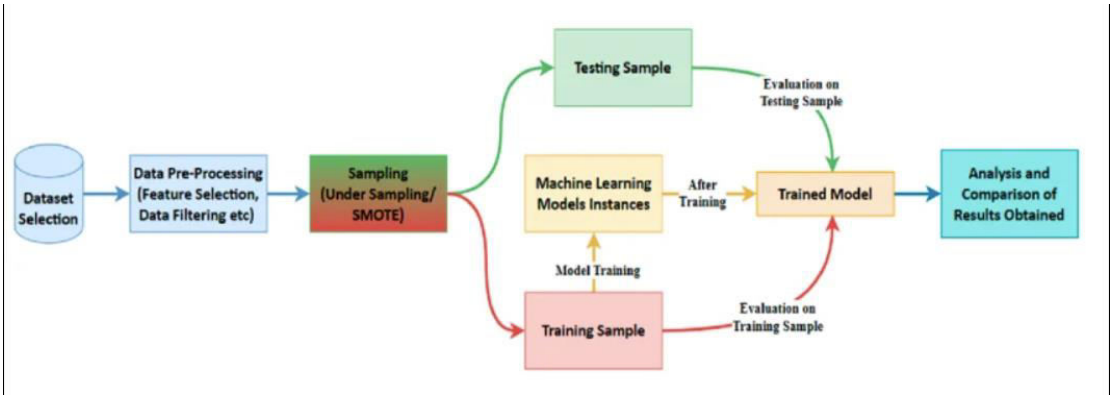
The fake job posting detection system follows a structured machine learning workflow. Initially, data collection is performed by obtaining a dataset containing job postings with labels indicating whether they are real or fake. The dataset includes features such as job title, company profile, job description, salary, location, and other relevant attributes.

Next, data preprocessing is carried out to clean the data by removing missing values, duplicates, and irrelevant information. Textual data is processed using Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and normalization. The cleaned text is then converted into numerical form using methods like TF-IDF or Count Vectorizer.

After preprocessing, feature extraction and selection are performed to identify important attributes that contribute to classification. The dataset is then split into training and testing sets.

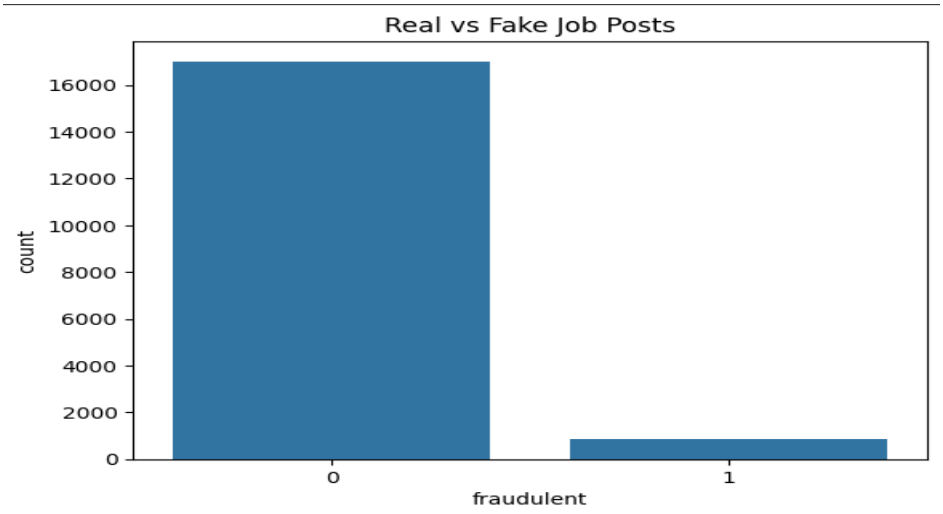
**System Architecture**

- Input Layer: Job posting data (title, description, salary, company info)
- Data Preprocessing: Cleaning, tokenization, NLP processing
- Feature Extraction: TF-IDF / Count Vectorizer
- Model Training: ML models (LR, RF, NB, DT)
- Evaluation Module: Accuracy, precision, recall
- Prediction Module: Classifies job postings
- Output Layer: Displays result (Real / Fake)

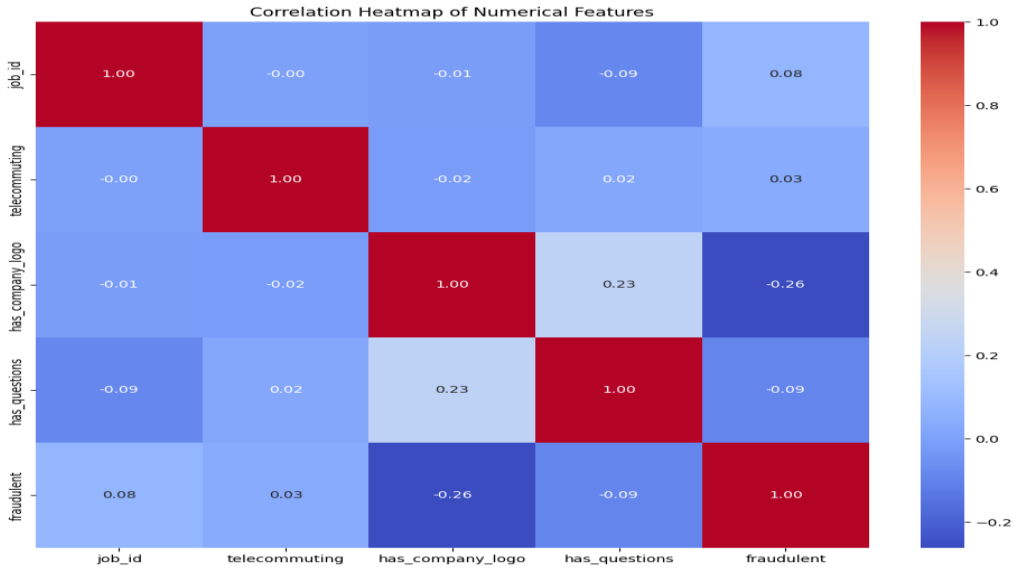



## V. Result and Output

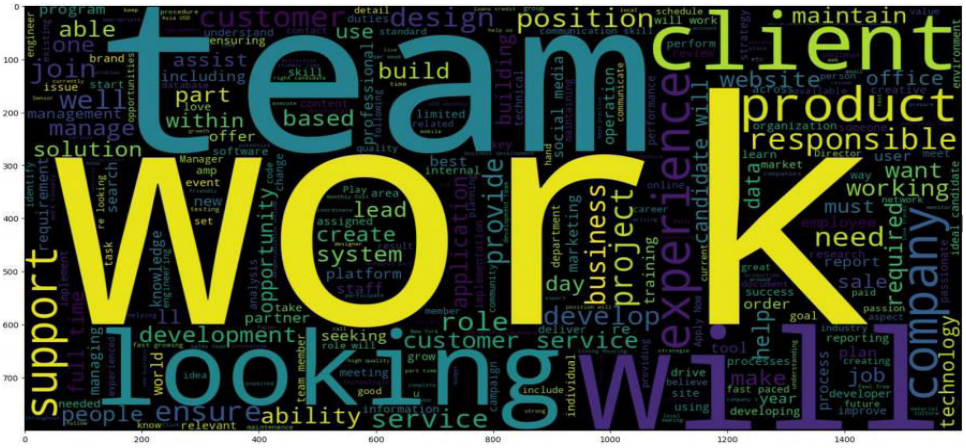
Exploratory Data Analysis Result

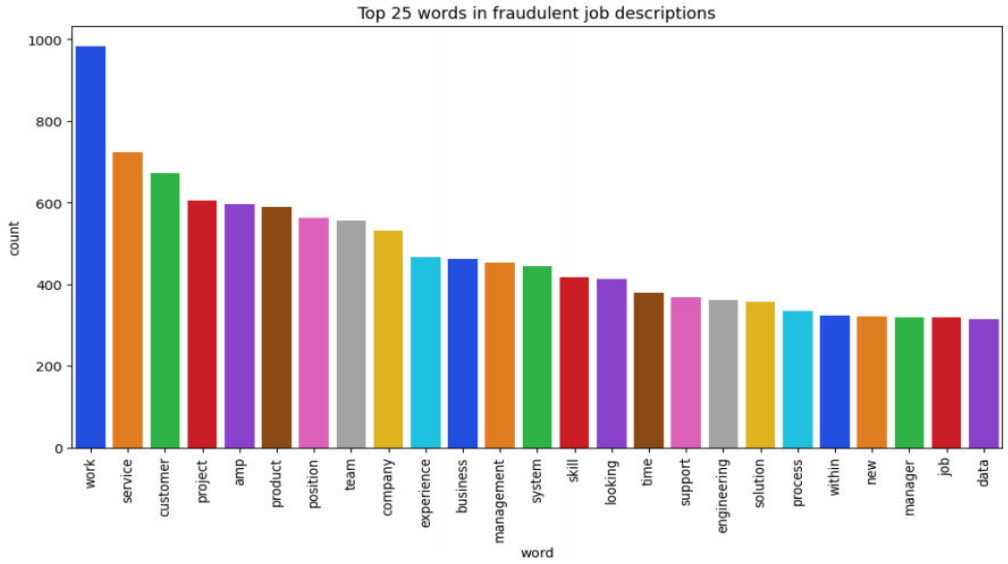


Correlation Heatmap of Numerical Features




Word Cloud for Non-Fraudulent Job Description



## VI. Conclusion

This project successfully demonstrates the application of machine learning and natural language processing (NLP) techniques to address the critical issue of fake job postings. Through comprehensive exploratory data analysis on a publicly available dataset, important challenges such as class imbalance were identified, along with key insights into the patterns and characteristics of both genuine and fraudulent job advertisements.

The structured methodology, including data preprocessing, feature engineering, and model evaluation, provides a strong foundation for building an accurate and reliable classification system. Visual analysis techniques such as word clouds and bar charts revealed meaningful differences in the language used by scammers, highlighting the importance of textual analysis in fraud detection.

The developed system proves to be an effective automated solution capable of identifying fake job postings with improved accuracy. Although certain limitations exist, the project establishes a strong proof-of-concept for real-world implementation. Overall, this work contributes to enhancing trust, safety, and reliability in online recruitment platforms, helping protect job seekers from potential fraud.

## References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring
COVID-19 through intensive investigation with supervised machine learning algorithm. In
Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International
Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer
Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on
colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve
1Professor, Department of computer Science & engineering, Anurag University, TS, India.
2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time
Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT,
Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6,
doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis
of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial
Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications
in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer,
2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting
fake documents from land records using blockchain technology," in Blockchain for Smart
Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
**[8]** Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie
recommender using collaborative filtering and sentiment analysis," *International Research
Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025,
doi: 10.56726/IRJMETS81618.
**[9] Ravi Kumar Banoth, Ramana Murthy B V**, "Automatic crop recommendation system
using LightGBM and decision tree machine learning models," *Journal of Machine and
Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
**[10] Ravi Kumar Banoth**, **Dr. B.V. Ramana Murthy**, "Smart agriculture through IoT and
machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and
Communication Engineering (ICCSCE)*, Apr. 2025.**[11]** Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer
learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199,
2024, doi: 10.1007/s42979-023-02500-x.