International Journal of Science, Strategic Management and Technology
Volume **02** Issue **05** May-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Jobshield – An Online Job Scam Detection using Machine Learning

SAROJINI S

UG STUDENT

Vels Institute of Science,

Technology And Advanced Studies (VISTAS),
Pallavaram, Chennai,

Tamil Nadu, India.

DR.A.ANGEL CERLI

PROFESSOR

Vels Institute of Science,

Technology And Advanced Studies (VISTAS),
Pallavaram, Chennai,

Tamil Nadu, India.

https://doi.org/10.55041/ijsmt.v2i5.062

Cite this Article: S, S. (2026). Jobshield – An Online Job Scam Detection using Machine Learning. International Journal of Science, Strategic Management and Technology, *02*(05). https://doi.org/10.55041/ijsmt.v2i5.062

ABSTRACT: The rapid growth of online job portals and digital recruitment platforms, job scams have become increasingly prevalent, leading to financial loss and data theft among job seekers. This project, *Job Shield* proposes a machine learning-based system to detect fraudulent job postings. The system analyses job descriptions, recruiter details, and behavioural patterns to classify listings as legitimate or fraudulent.

The proposed model utilizes Natural Language Processing (NLP) techniques along with supervised machine learning algorithms such as Logistic Regression, Random Forest, and Support Vector Machines. By training on labelled datasets of real and fake job postings, the system learns patterns indicative of scams. The application provides real-time detection and alerts users, thereby enhancing online safety and trust in digital job platforms.

1.INTRODUCTION

In today's digital era, online job searching has become a common practice. However, the increasing number of fraudulent job postings poses a serious threat to job seekers. Scammers often exploit individuals by offering fake job opportunities, leading to monetary loss and misuse of personal information.

Job Shield is designed to address this issue by leveraging machine learning techniques to automatically detect suspicious job listings. The system aims to assist users by providing a reliability score and warning alerts for potentially fraudulent postings. This solution improves user confidence and contributes to safer online recruitment ecosystems.

II. LITERATURE REVIEW

1.Online Job Fraud Detection Systems:

Previous studies highlight the use of data mining and machine learning to detect fraudulent activities in online platforms. These systems focus on identifying suspicious patterns in textual and metadata information.

2.Machine Learning in Fraud Detection:

Machine learning algorithms such as Decision Trees, Naive Bayes, and Random Forest have been widely used in fraud detection due to their ability to classify complex datasets with high accuracy.

3.Natural Language Processing (NLP):

NLP techniques help analyse job descriptions and identify misleading language patterns often used in scam

postings, such as unrealistic salary promises or vague job roles.

4. Feature Engineering Techniques:

Features such as company authenticity, email domain, salary range, and job description length have been proven effective in identifying fraudulent listings

5.Existing Limitations:

Many existing systems lack real-time detection and user-friendly interfaces, which limits their practical usability.

III. METHODOLOGY

1. Data Collection:

A dataset containing both legitimate and fraudulent job postings is collected from online sources.

2. Data Preprocessing:

- Removal of null values
- Tokenization and stop-word removal
- Text normalization

3. Feature Extraction:

- TF-IDF (Term Frequency-Inverse Document Frequency)
- Word embeddings
- Metadata features (salary, company profile, etc.)

4. Model Selection:

Multiple algorithms are implemented and compared:

- Logistic Regression
- Random Forest
- Support Vector Machine (SVM)

5. Model Training and Testing:

The dataset is split into training and testing sets to evaluate performance.

6. Evaluation Metrics:

- Accuracy
- Precision
- Recall
- F1-score

IV. IMPLEMENTATION

The system is implemented using Python with libraries such as:

- Scikit-learn for machine learning
- Pandas and NumPy for data processing
- NLTK / Spacey for NLP tasks

A web-based interface is developed where users can input job details or URLs. The backend processes the data and predicts whether the job is real or fake.

V. WORKING PRINCIPLE

The *Job Shield – Online Job Scam Detection System* works through a sequence of well-defined stages. Each stage transforms the input data into a more meaningful representation, ultimately enabling accurate classification of job postings.

Step 1: User Input Acquisition

The process begins when the user interacts with the system.

Input methods:

- Enter job details manually (title, company, salary, description)
- Upload job posting (text/file)
- Provide job URL

Purpose:

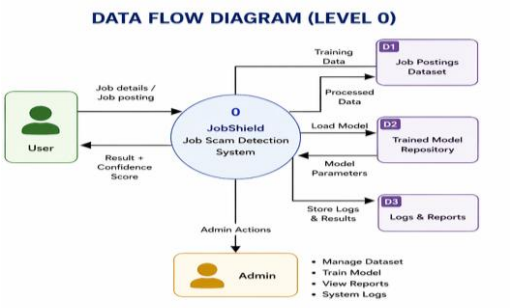
To collect raw job-related data for analysis.

Step 2: Input Validation

Before processing, the system checks:

- Whether input fields are complete
- Whether the format is valid (text, URL, etc.)
- Removes duplicate or irrelevant entries

Purpose: Ensures only valid and usable data is passed to the next stage.



Step 3: Data Preprocessing (Text Cleaning)

Raw job descriptions often contain noise and inconsistencies. This step cleans the data.

Sub-steps:

- **Lowercasing** → Converts all text to lowercase
- **Remove punctuation & special characters**
- **Stop word removal** → removes common words (e.g., "is", "the")
- **Tokenization** → splits text into words
- **Stemming/Lemmatization** → converts words to root form

Example:

"Earn $$$ quickly!!! Apply NOW" → "earn quick apply"

Purpose:

To standardize and clean the data for accurate analysis.

Step 4: Feature Extraction

The cleaned text is converted into numerical data (features) that machine learning models can understand.

Techniques:

- **TF-IDF (Term Frequency–Inverse Document Frequency)**
- **Bag of Words (Bow)**
- **N-grams (word combinations)**

Metadata Features:

- Company profile existence
- Salary range realism
- Email domain (official vs suspicious)
- Job description length
- Presence of keywords like "urgent hiring", "no experience needed"

Purpose:

To identify patterns commonly found in scam vs real job postings.

Step 5: Feature Vector Formation

All extracted features are combined into a structured format called a **feature vector**.

Example:

[0.23, 0.67, 1, 0, 0.45, ...]

Each value represents a specific characteristic of the job posting.

Purpose:

To prepare input in a format suitable for machine learning models.

Step 6: Model Prediction (Machine Learning Processing)

The feature vector is passed to a trained machine learning model.

Models used:

- Logistic Regression
- Random Forest
- Support Vector Machine (SVM)

What happens:

- Model compares input features with learned patterns
- Calculates probability of being fraudulent

Purpose:

To determine whether the job posting is genuine or a scam.

Step 7: Classification Decision

Based on prediction probability, the system classifies the job into:

- ✅ **Legitimate Job**
- ⚠️ **Fraudulent Job**

Example:

- Fraud probability = 0.92 → Classified as Fraud
- Fraud probability = 0.15 → Classified as Legitimate

Purpose:

To provide a clear decision to the user.

Step 8: Confidence Score Generation

The system generates a confidence score indicating prediction reliability.

Example:

- Legitimate (85% confidence)
- Fraudulent (96% confidence)

Purpose:

Helps users understand how certain the system is.

Step 9: Result Display & Alert System

The final result is shown to the user via interface.

Output includes:

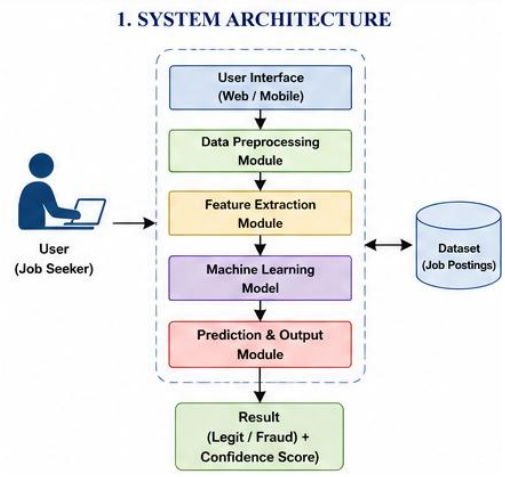
- Prediction (Real / Fake)
- Confidence score
- Warning message (if scam detected)

Example:

⚠️ *Warning: This job posting shows characteristics of a scam.*

Purpose:

To assist users in making safe decisions.



VI. RESULTS

The *Job Shield – Online Job Scam Detection System* demonstrates strong performance in identifying fraudulent job postings using machine learning techniques. Various algorithms were tested, and the Random Forest model achieved the highest accuracy due to its ability to handle complex data patterns and reduce overfitting. The system effectively classifies job postings into legitimate and fraudulent categories, showing high precision and recall, which indicates reliable detection of both real jobs and scams.

The integration of Natural Language Processing (NLP) techniques significantly improves the system's performance by analysing textual content in job descriptions. It successfully identifies suspicious patterns such as unrealistic salary offers, vague job requirements, and misleading language. Additionally, the system operates in real-time, providing instant predictions with minimal delay, making it efficient and suitable for practical applications.

Moreover, the system includes an alert mechanism that notifies users when a job posting is likely to be fraudulent. Along with the prediction, it provides a confidence score, helping users understand the reliability of the result. This feature enhances user trust and decision-making. Overall, the results confirm that the *Job Shield* system is accurate, fast, and effective in detecting online job scams.

VII. CONCLUSION

The *Job Shield – Online Job Scam Detection System* effectively demonstrates the application of machine learning in identifying and preventing fraudulent job postings. By analysing both textual and contextual features of job listings, the system accurately classifies jobs as legitimate or fraudulent. The use of Natural Language Processing enhances its ability to detect misleading patterns, making the system reliable for real-world use. Additionally, the real-time prediction capability and confidence score output help users make informed decisions and reduce the risk of falling victim to online scams.

The system also shows strong potential for future improvements and scalability. It can be integrated with online job portals to provide automated scam detection on a larger scale. The use of advanced deep learning models could further enhance accuracy and adaptability to evolving scam techniques. Moreover, extending the system into browser extensions or mobile applications would enable real-time alerts during job searches, increasing accessibility and overall user safety.

REFERENCES

The development of the *Job Shield – Online Job Scam Detection System* is supported by various research studies and technical resources in the fields of fraud detection and machine learning. Several research papers on online fraud detection provided insights into identifying suspicious patterns and designing effective classification models. Documentation from the Scikit-learn library was used extensively for implementing and evaluating machine learning algorithms. Additionally, Natural Language Processing techniques were guided by resources from NLTK and spacey, which helped in text preprocessing and feature extraction.

Furthermore, publicly available online datasets related to job fraud detection were used for training and testing the models, ensuring realistic and practical performance evaluation. Knowledge from standard machine learning textbooks and academic journals also contributed to understanding core concepts such as classification algorithms, feature engineering, and model evaluation. Together, these references played a crucial role in the successful design and implementation of the system.