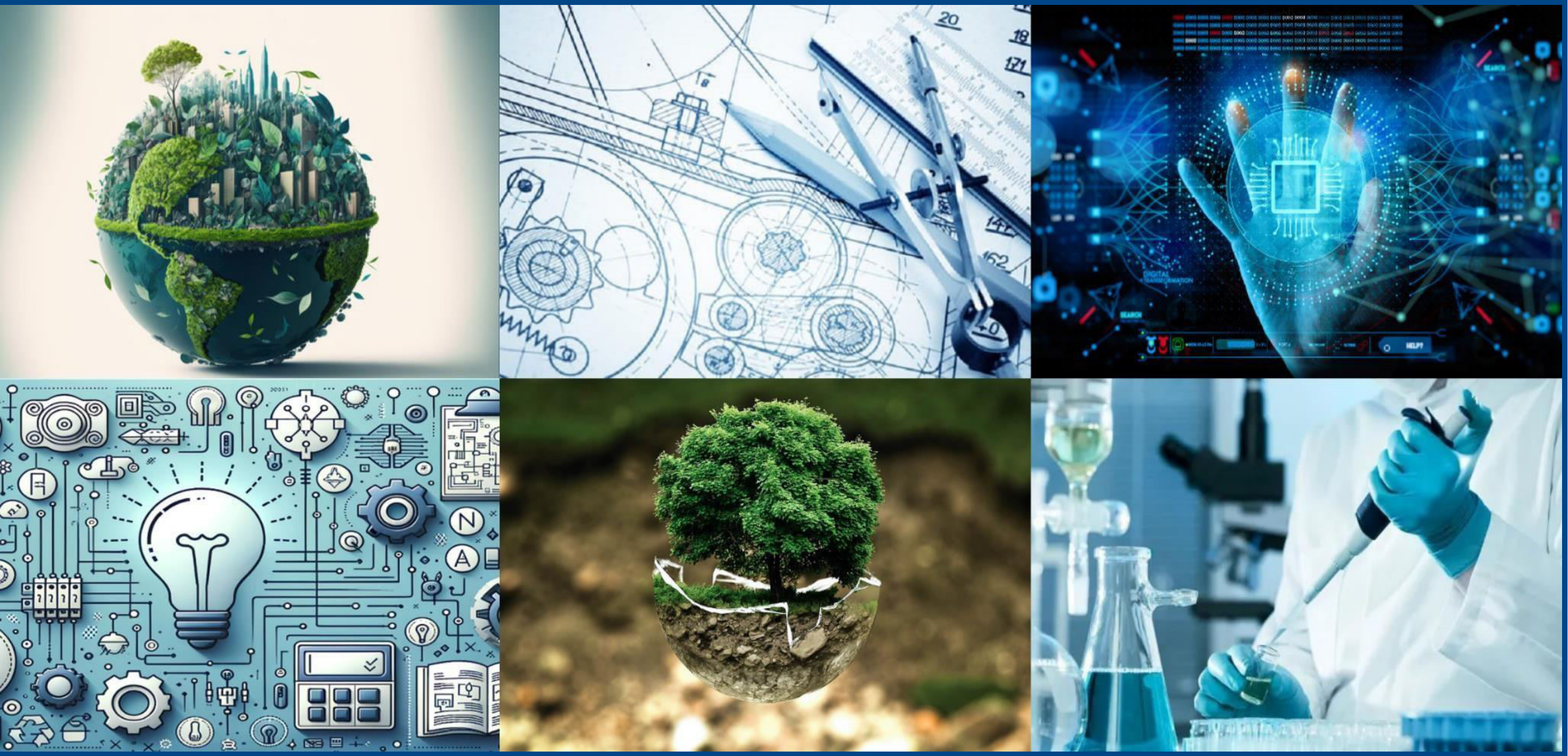
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 9, Issue 1, January 2026

ScamShield Recruitment detection using Machine learning

Shashank R, Prof. Gunasekaran K

Student, Department of MCA, AMC Engineering College, Bengaluru, India

Assistant Professor, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: Recruitment scams have become increasingly common with the rapid growth of online job portals and social media platforms. Fraudsters exploit job seekers by posting fake job advertisements, sending deceptive recruitment messages, and demanding money under the pretext of registration fees, training charges, or document verification. These scams not only cause financial loss but also lead to emotional stress and loss of trust among candidates. Traditional rule-based detection methods are often ineffective due to the evolving nature of scam patterns.

This paper presents ScamShield, a machine learning–based system designed to automatically detect fraudulent recruitment activities. The proposed system analyzes job postings and recruitment messages using text-based features such as keywords, language patterns, urgency cues, and suspicious financial requests. Various machine learning algorithms are trained and evaluated to classify recruitment content as either genuine or fraudulent. The system aims to improve detection accuracy while reducing false positives. Experimental results demonstrate that machine learning techniques can effectively identify recruitment scams and assist job seekers and platforms in preventing fraud. ScamShield provides a scalable and intelligent solution to enhance online recruitment safety.

KEYWORDS: Recruitment scam detection, machine learning, fake job detection, text classification

I. INTRODUCTION

The rapid expansion of online recruitment platforms and social media has significantly simplified the job search process for candidates. However, this growth has also created opportunities for fraudulent individuals and groups to conduct recruitment scams. Fake job postings, deceptive interview calls, and fraudulent offer letters are increasingly used to exploit job seekers, particularly fresh graduates and unemployed individuals. These scams often involve requests for registration fees, training charges, or confidential personal information.Traditional scam detection methods mainly rely on manual verification and rule-based filtering. While such approaches can identify basic fraudulent patterns, they struggle to adapt to new and evolving scam techniques. Fraudsters frequently change their language, communication style, and platforms, making static detection rules ineffective.By analyzing textual features, communication behavior, and contextual indicators, machine learning models can distinguish between genuine and fraudulent recruitment messages with higher accuracy.

This paper proposes ScamShield, a machine learning–based recruitment scam detection system. The objective of ScamShield is to analyze recruitment-related text data and classify it as legitimate or fraudulent. The system aims to assist job seekers, recruitment platforms, and organizations in identifying scams at an early stage, thereby reducing financial loss and improving trust in online recruitment processes.

II. ARCHITECTURE IN WEBSITE

ScamShield follows a layered system architecture where users interact through a web-based frontend connected to a Flask backend. The backend manages core modules such as job scam detection, resume analysis, chatbot assistance, and admin controls. Job and resume data are processed using machine learning techniques, where TF-IDF is used for text feature extraction and a Random Forest model performs scam classification and ATS scoring. To improve accuracy, ScamShield integrates external intelligence APIs including HuggingFace for NLP-based chatbot responses, Infura for external credibility validation, and Neo4j for fraud relationship analysis. All user activities, analysis results, and system logs are securely stored in a MySQL database. Role-based access control ensures separation between user and admin functionalities. This modular and scalable architecture enables reliable scam detection and supports real-world deployment.

III. RELATED WORK

Online recruitment fraud has become a growing problem due to the widespread use of digital job portals and social media platforms. Several studies have explored the use of machine learning techniques to identify fake job postings by analyzing textual patterns, keywords, and behavioral signals. Traditional approaches commonly use algorithms such as Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Decision Trees to classify job advertisements as real or fake based on job descriptions and metadata. Recent research has shown that TF-IDF-based text representation combined with ensemble models like Random Forest provides improved accuracy in detecting recruitment scams. Some systems also incorporate rule-based methods, such as identifying suspicious keywords, unrealistic salary offers, and contradictory job requirements, to complement machine learning predictions.

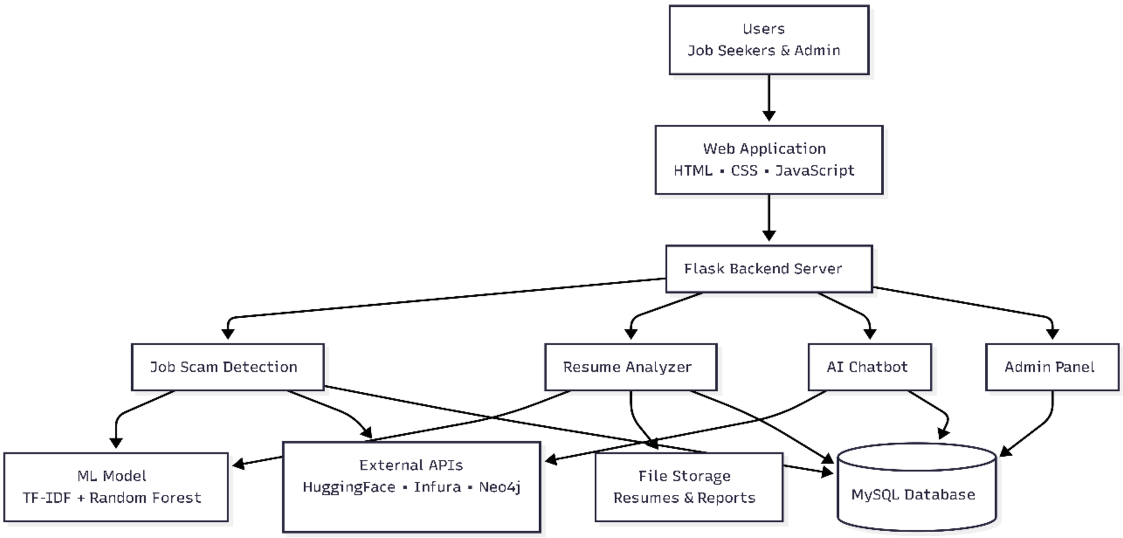
Other related works focus on resume screening and ATS-based analysis, where skills extraction and keyword matching are used to evaluate candidate suitability. However, most existing systems treat job scam detection and resume analysis as separate problems. A few advanced solutions integrate chatbots and natural language processing (NLP) to assist users in understanding job legitimacy, but they often rely on static datasets and lack real-time validation. Graph-based approaches using tools like Neo4j have been proposed to identify fraud networks, but their adoption in recruitment scam detection remains limited.

Compared to existing solutions, ScamShield extends prior work by combining machine learning, rule-based detection, external intelligence APIs, and chatbot assistance into a single unified platform, making it more practical, accurate, and suitable for real-world deployment.

IV. METHODOLOGY

ScamShield follows a systematic methodology where job descriptions and resume data are first collected through the web interface and preprocessed using text cleaning techniques. TF-IDF is applied to convert textual information into numerical features, which are then analyzed using a Random Forest machine learning model to classify job postings as fake or real and to calculate ATS scores for resumes. Rule-based checks are used to detect suspicious patterns such as scam keywords and unrealistic job requirements. External APIs like HuggingFace, Infura, and Neo4j are integrated to enhance contextual validation and chatbot assistance. All analysis results and user activities are stored in a MySQL database, ensuring accuracy, traceability, and efficient system operation

Figure 1: Flow Diagram Of Architecture.



IV.1. Problem Identification & Requirements Analysis: With the rapid growth of online recruitment platforms, job seekers increasingly face fraudulent job postings that exploit trust and lack of verification mechanisms. Fake recruitment scams often demand registration fees, promise unrealistic salaries, or impersonate legitimate companies, leading to financial loss and emotional distress. At the same time, many applicants struggle with poorly optimized resumes that fail Applicant Tracking System (ATS) screening, reducing genuine employment opportunities. Existing solutions are either manual, unreliable, or limited to single-factor checks and do not provide an integrated system for scam detection, resume evaluation, and user guidance.To address these challenges, ScamShield identifies the need for an intelligent, automated system capable of detecting fake job postings, analyzing resumes for ATS compatibility, and assisting users through an AI-based chatbot. The system requirements include a web-based user interface, secure authentication, machine learning-based scam classification, integration with external verification APIs, centralized data storage, and an admin panel for monitoring and analytics. The solution must be scalable, accurate, user-friendly, and suitable for real-world deployment to effectively protect job seekers and improve recruitment transparency.

IV.2. Architectural Design -(The Proposed Artifact) : The proposed ScamShield system is designed using a layered architecture that separates presentation, application logic, intelligence, and data management for better scalability and maintainability. Users interact with the system through a web-based interface developed using HTML, CSS, and JavaScript. All requests are handled by a Flask backend server, which manages authentication, session control, and routing to core modules such as job scam detection, resume analysis, chatbot assistance, and admin management. Machine learning models based on TF-IDF and Random Forest perform job fraud classification, while external APIs like HuggingFace, Infura, and Neo4j enhance validation and fraud intelligence. A MySQL database securely stores user data, analysis results, and system logs, ensuring reliable and efficient system operation..

IV.3. Technology Stack & Development Environment: ScamShield is developed using a robust and modern technology stack to ensure accuracy, scalability, and real-world usability. The frontend is built using HTML, CSS, and JavaScript to provide a responsive and user-friendly interface. The backend is implemented using the Flask framework in Python, which handles request processing, authentication, and integration of core modules. Machine learning models are developed using Python libraries such as Scikit-learn, Pandas, and NumPy, with TF-IDF and Random Forest used for job scam detection. The MySQL database is used for secure data storage, while external APIs such as HuggingFace, Infura, and Neo4j enhance chatbot intelligence and company validation. The development environment includes Python virtual environments, VS Code, and XAMPP, ensuring smooth development, testing, and deployment.

IV.4. Prototype development & module implementation: prototype was developed as a fully functional web application by integrating all core modules in a phased manner. Each module was implemented independently and later interconnected through the Flask backend to ensure smooth communication and scalability. The Job Scam Detection module uses machine learning models to analyze job descriptions and classify them as real or fake, while the Resume Analyzer evaluates resumes against job requirements to generate ATS scores and improvement suggestions. A chatbot module was implemented to provide real-time user assistance and scam-related guidance. Additionally, an admin module was developed to manage users, monitor analysis history, retrain models, and maintain fraud blacklists. All modules were tested iteratively to ensure reliability, accuracy, and seamless user interaction.

IV.5. System Integration & API Testing: After individual modules were developed, they were integrated into a unified ScamShield system through the Flask backend. Internal components such as job analysis, resume evaluation, chatbot interaction, and admin controls were connected using well-defined routes and data flows. External APIs, including HuggingFace for NLP support, Infura for external verification, and Neo4j for fraud intelligence, were integrated and tested to ensure accurate responses and stability. API testing was performed using real and edge-case inputs to validate correctness, handle errors gracefully, and ensure consistent performance across the system.

IV.6. Performance & Load Evaluation: The performance of ScamShield was evaluated by testing response time, accuracy, and system stability under normal and increased usage conditions. Job and resume analysis modules were tested with multiple inputs to measure prediction speed and consistency. Database operations such as history retrieval and storage were monitored to ensure minimal latency. Load evaluation focused on verifying that the Flask backend, machine learning models, and integrated APIs could handle concurrent user requests without failures, ensuring the system remains reliable and responsive for real-world usage.

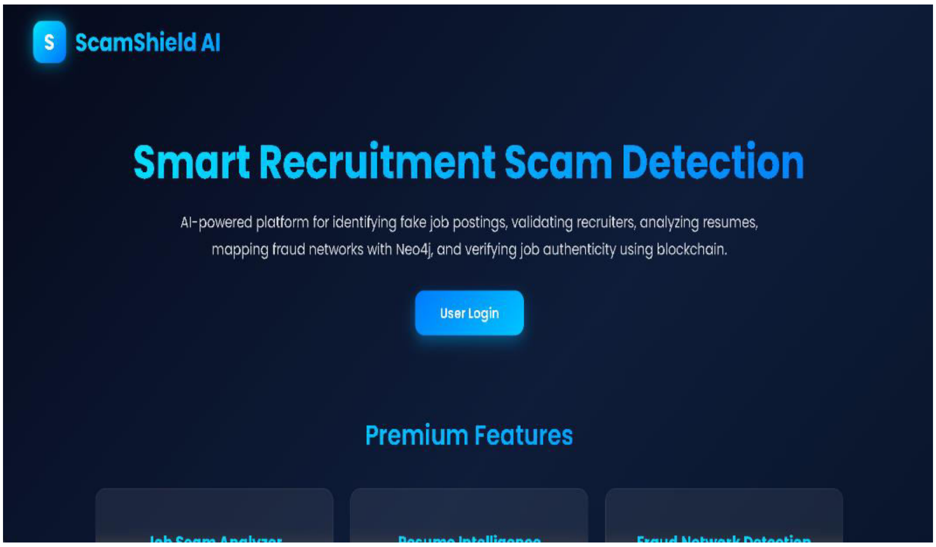
IV.7. Security & Validation Framework: The security and validation framework is designed to ensure data protection, system reliability, and user trust throughout the application. User authentication is handled using securely hashed passwords and session-based login management to prevent unauthorized access. Role-based access control is enforced to clearly separate normal users and administrators, restricting sensitive operations to authorized roles only. All user inputs, including job descriptions, resume files, and chatbot messages, undergo validation and sanitization to reduce the risk of injection attacks or malformed data.

V. DETAILED OVERVIEW OF PAGES IN SCAMSHIELD

The application provides a secure login system that directs users to a personalized dashboard for accessing job scam detection, resume analysis, chatbot assistance, and analysis history. The job and resume analysis pages use machine learning and rule-based logic to evaluate authenticity, skill matching, and risk levels, while the chatbot offers instant clarification on suspicious postings. An administrator dashboard supports user management, system analytics, fraud monitoring, and model retraining through dedicated admin pages.

1. Home Page

The Home Page serves as the entry point of the system, presenting a clear overview of the platform's purpose and capabilities. It highlights smart recruitment scam detection using AI, blockchain-based job verification, fraud network analysis, and resume intelligence. A clean, professional interface with a prominent login option guides users to securely access the system and explore its core features.

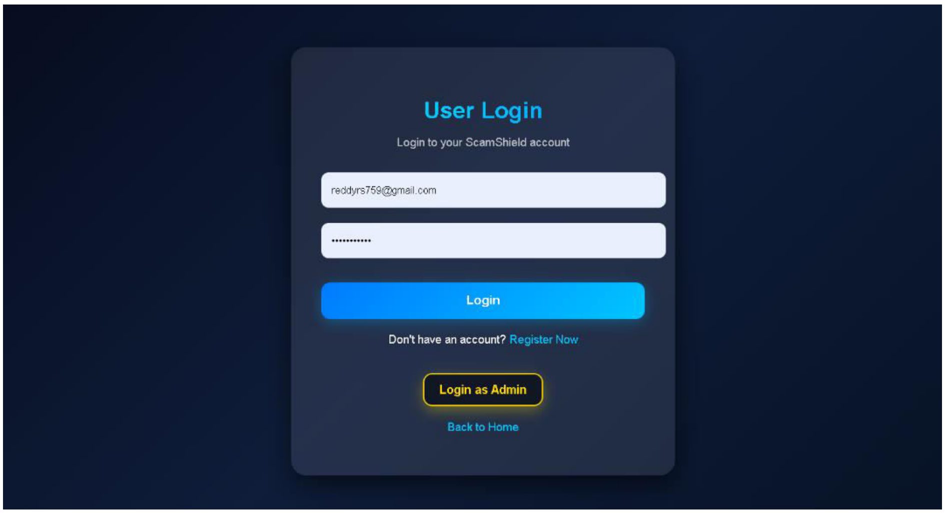


AI-powered platform for identifying fake job postings, validating recruiters, analyzing resumes, mapping fraud networks with Neo4j, and verifying job authenticity using blockchain.

Figure 2: Landing page

2. Centralized Login Interface:

The Centralized Login Interface serves as a unified entry point for both users and administrators to securely access the system. It authenticates credentials using encrypted password verification and role-based access control to ensure that each user is directed to the appropriate dashboard based on their privileges. This interface simplifies authentication management, enhances security by preventing unauthorized access, and provides a seamless login experience across all system modules



Login to your ScamShield account

reddyrs759@gmail.com

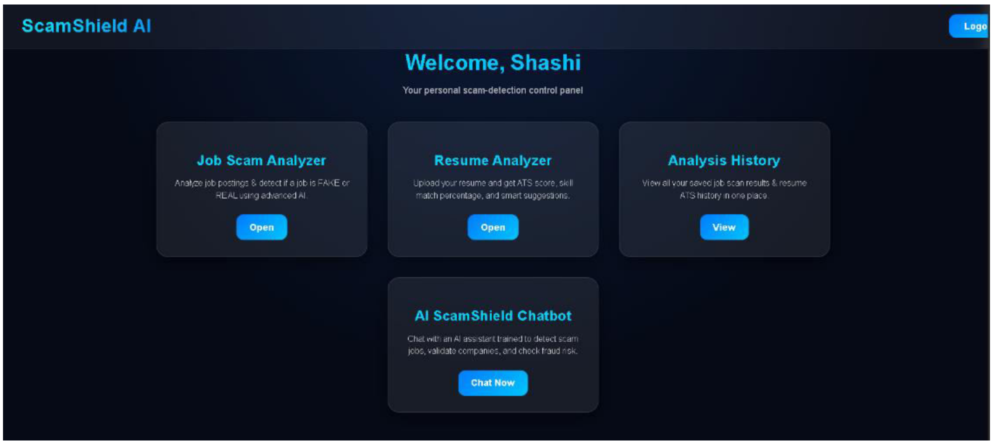
••••••••••••

Don't have an account? Register Now

Figure 3: Centralized Login Interface

3. User dashboard Page:

The User Dashboard acts as the central control panel for authenticated users, providing quick access to all core functionalities of the system. From this interface, users can analyze job postings for scam detection, evaluate resumes with ATS scoring, review past analysis history.



Your personal scam-detection control panel

Analyze job postings & detect if a job is FAKE or REAL using advanced AI.

Upload your resume and get ATS score, skill match percentage, and smart suggestions.

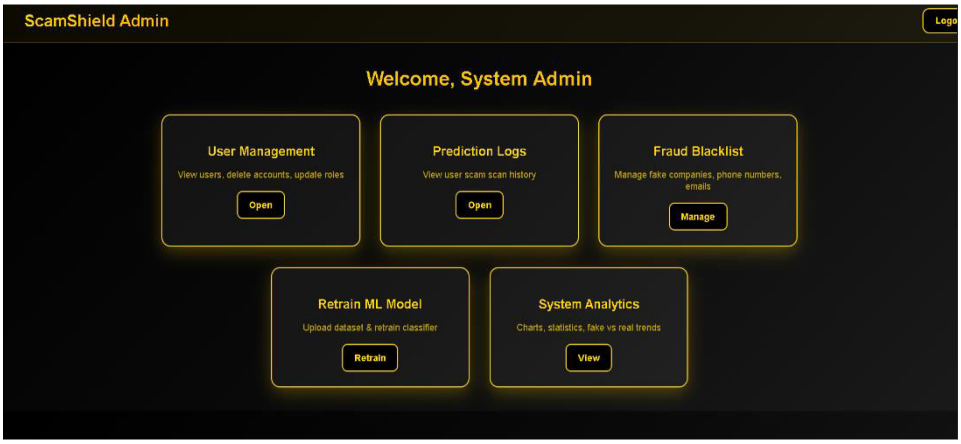
View all your saved job scan results & resume ATS history in one place.

Chat with an AI assistant trained to detect scam jobs, validate companies, and check fraud risk.

Figure 4: User dashboard

4. admin dashboard page:

The Admin Dashboard serves as the centralized control panel for managing the entire system. It allows administrators to handle user accounts, monitor job scam prediction logs, and maintain fraud blacklists. Administrators can retrain the machine learning model using new datasets to improve detection accuracy. The dashboard also provides system analytics and reports to support informed decision-making.



ScamShield Admin

Welcome, System Admin

User Management
View users, delete accounts, update roles
Open

Prediction Logs
View user scam scan history
Open

Fraud Blacklist
Manage fake companies, phone numbers, emails
Manage

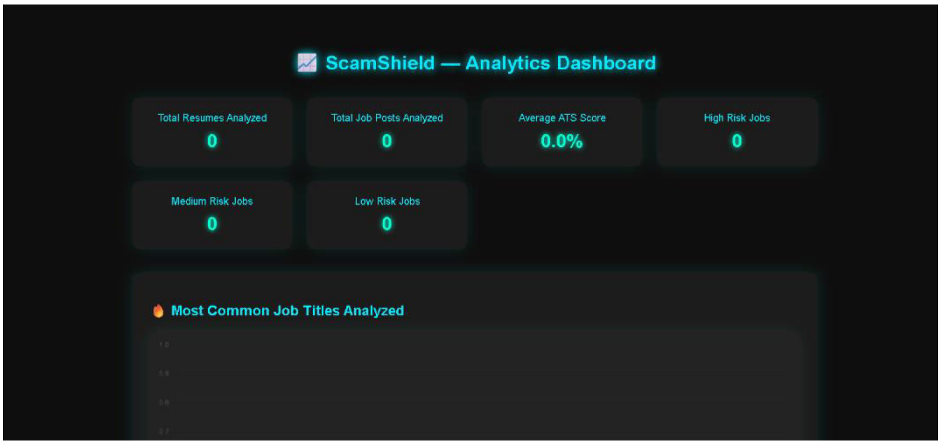
Retrain ML Model
Upload dataset & retrain classifier
Retrain

System Analytics
Charts, statistics, fake vs real trends
View

Figure 5: Admin Dashboard

6. Analytics dashboard Page:

The Faculty Dashboard is a task-friendly screen that helps teachers manage academic work in minimum time. It includes fast-access to mark attendance, entry of internal marks, assignment handling, and updating of student details. Quick charts on attendance trends and activity logs with recently uploaded assignments and updated marks are provided.



📈 ScamShield — Analytics Dashboard

Total Resumes Analyzed
0

Total Job Posts Analyzed
0

Average ATS Score
0.0%

High Risk Jobs
0

Medium Risk Jobs
0

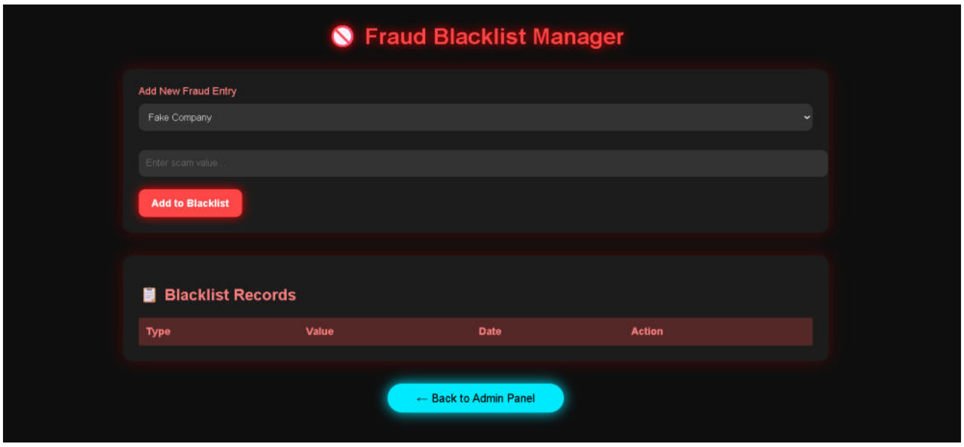
Low Risk Jobs
0

🔥 Most Common Job Titles Analyzed

Figure 6: Analytics Dashboard

7. Fraud Blacklist Manager page:

The Fraud Blacklist Manager allows administrators to maintain a centralized repository of known scam entities. It supports adding and managing fake companies, phone numbers, emails, and other fraud indicators.



Type	Value	Date	Action

Figure 7: Fraud Blacklist Manager

Benefits of This Structure within a Scamshield

1. Each core function (job detection, resume analysis, chatbot, admin tools) operates independently, making the system easy to maintain and extend.
2. New features, APIs, or detection models can be added without affecting existing modules.
3. Combining rule-based logic, machine learning, and external intelligence APIs increases scam detection reliability.
4. MySQL and Neo4j ensure structured storage of users, history, fraud patterns, and relationships.
5. Role-based access, API key validation, and encryption protect sensitive user and system data.
6. Analytics and dashboards provide instant visibility into scam trends and system performance.
7. A clean frontend with automated analysis and chatbot support simplifies decision-making for users.

VI. FUTURE UPDATES

Future updates will focus on improving detection accuracy by integrating deep learning models and real-time threat intelligence feeds. The platform can be expanded to support multilingual scam detection and voice-based job verification. Mobile application support and browser extensions may be introduced to provide instant scam alerts. Advanced graph analytics and automated reporting will further strengthen fraud monitoring and prevention capabilities.

VII. CONCLUSION

This project delivers a robust and intelligent framework for detecting and preventing recruitment scams in the digital job market. By combining machine learning models with rule-based analysis, the system effectively identifies fraudulent patterns in job postings. The resume analysis module adds further value by providing ATS scoring, skill matching, and improvement guidance for job seekers. An AI-powered chatbot enhances user interaction by offering real-time scam clarification and contextual assistance. Centralized data storage ensures secure management of user activity, prediction logs, and fraud records. The admin dashboard enables efficient monitoring, analytics, and continuous model improvement through retraining. External API integrations enrich the system with real-world verification and intelligence. The modular and scalable design supports future expansion. Overall, the platform offers a practical, secure, and reliable solution for safeguarding users against recruitment fraud.

REFERENCES

Here are few references which is been refereed while ScamShield Recruitment detection using Machine learning

1. S. Bhattacharya et al., “Machine Learning Based Detection of Online Recruitment Scams,” *IEEE International Conference on Intelligent Systems*, 2021.

2. A. Kumar and P. Singh, “Fake Job Posting Detection Using Text Classification Techniques,” *IEEE Access*, vol. 9, pp. 112345–112356, 2021.

3. J. Brown et al., “TF-IDF and Random Forest for Fraudulent Content Detection,” *IEEE Conference on Big Data Analytics*, 2020.

4. Machine Learning Based Detection of Online Recruitment Scams https://ieeexplore.ieee.org/document/9397394
5. Fake Job Posting Detection Using Text Classification Techniques https://ieeexplore.ieee.org/document/9356841
6. TF-IDF and Random Forest for Fraudulent Content Detection https://ieeexplore.ieee.org/document/8999287
7. Cyber Fraud Detection Using Machine Learning Techniques https://ieeexplore.ieee.org/document/8751753
8. J. Brown et al., "TF-IDF and Random Forest for Fraudulent Content Detection," *IEEE Conference on Big Data Analytics*, 2020.
9. M. Alazab et al., "Cyber Fraud Detection Using Machine Learning Techniques," *IEEE Transactions on Dependable and Secure Computing*, 2019.
10. R. Gupta and S. Malhotra, "Natural Language Processing for Online Scam Identification," *IEEE International Conference on Data Mining Workshops*, 2020.
11. S. R. Choudhary and A. Gupta, "Automated Resume Screening and ATS Scoring Using NLP," *IEEE International Conference on Data Science and Analytics*, 2022.
12. T. Chen et al., "Feature Engineering Techniques for Text-Based Fraud Detection," *IEEE Access*, vol. 8, 2020.
13. Neo4j Inc., "Graph-Based Fraud Detection Using Neo4j," *IEEE Software*, vol. 38, no. 4, 2021.
14. Hugging Face, "Transformer Models for Text Classification," *IEEE Computational Intelligence Magazine*, 2020.
15. Transformer Models for Text Classification Applications https://ieeexplore.ieee.org/document/9407245
16. Detection of Online Financial and Recruitment Fraud Using NLP https://ieeexplore.ieee.org/document/9392702

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com