Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/22245
DOI URL: http://dx.doi.org/10.21474/IJAR01/22245

RESEARCH ARTICLE

A ROBUST HYBRID MACHINE LEARNING PIPELINE FOR IDENTIFYING JOB SCAMS ONLINE

Ragham Jitesh[1], Ilindra Krishna Lekha[2], Tellamekala Harikrishna[1], Podili Yaswanth Mani[1], Samantapudi Sirisha[3] and Sanka Vignesh[2]

1. UG Student, Dept. of Computer Science (IOT, CS with BCT), QIS College of Engineering and Technology (A), Ongole, Andhra Pradesh, India.
2. UG Student, Department of CSE, QIS College of Engineering and Technology (A), Ongole, Andhra Pradesh, India.
3. UG Student, Department of ECE, RISE Krishna Sai Prakasam Group of Institutions, Ongole, Andhra Pradesh, India.

Manuscript Info

Manuscript History
Received: 04 September 2025
Final Accepted: 06 October 2025
Published: November 2025

Key words:-
Fraudulent Job Detection, Hybrid Machine Learning, Online Recruitment Scams, Multi-format Input Analysis, Cybersecurity in Employment, Text and Image Classification, Trustworthy Job Recommendations.

Abstract

With the rapid expansion of online recruitment platforms, fake job advertisements have emerged as a serious concern, often resulting in identity theft, financial fraud, and the manipulation of innocent job seekers.To combat this growing challenge,we present a hybrid machine learning– based solution designed to detect and classify fraudulent job postings with high precision. The system is versatile, accepting job-related inputs in multiple forms such as text, website links, and images, making it suitable for real world deployment across various job platfor ms.After extraction,the collected information is organized into CSV dat asets and analyzed using multiple classification models,including Naïve Bayes, Random Forest, Support Vector Machine (SVM), and XGBoost. By combining the predictive capabilities of these algorithms, the model achieves robust and reliable identification of authentic versus scam job listings.Ultimately, this approach strengthens online hiring security and assists job seekers in identifying legitimate opportunities with greater trust and confidence.

Introduction:-

The rise of online job portals has revolutionized recruitment—but it's also opened the door to a disturbing trend: fake job postings designed to exploit job seekers. These scams often lead to identity theft, financial loss, and emotional distress. While platforms attempt to moderate content, the sheer volume and sophistication of fraudulent listings make manual detection nearly impossible. This paper introduces a hybrid machine learning system designed to tackle this issue head-on. By analyzing job postings in text, link, and image formats, our system classifies listings as genuine or fraudulent with high accuracy. We leverage multiple algorithms—Naïve Bayes, Random Forest, SVM, and XGBoost—to ensure robust performance across diverse data types. The goal is simple: empower job seekers with a tool that helps them navigate the digital job market safely and confidently.

Objective:-

The main goal of this project is to build a smart, easy-to-use system that can spot fake job postings across online job portals using hybrid machine learning methods. It is designed to protect job seekers from scams, data theft, and financial fraud by accurately telling apart real and fake job listings. The system accepts different types of inputs such as text, links, or screenshots, processes them automatically, and then provides clear results showing whether a job is genuine or fraudulent, along with a confidence score. By combining multiple algorithms like Naïve Bayes, Random Forest, SVM, and XGBoost, the system ensures higher accuracy and reliability. Overall, the project aims to create a trustworthy and accessible tool that helps users make safer career decisions and promotes transparency in online recruitment.

Related Work:-

Over the past few years, research on detecting fraudulent job advertisements and online recruitment scams has grown rapidly due to increasing cyber risks and digital exploitation. Studies by Naudé et al. [1] and Swapna [2] demonstrated that machine learning can successfully differentiate real and fake job postings by examining linguistic cues and structural patterns found in job descriptions. Further evidence was presented by Sharma et al. [6] and Mehboob and Malik [20], who applied algorithms such as Random Forest and Decision Trees, achieving promising results in identifying deceptive recruitment posts. More recent work by Kaur and Sharma [17] and Abedeen et al. [18] introduced deep learning methods like BERT and LSTM, which significantly enhanced model accuracy and classification reliability. In addition to job fraud detection, several researchers have focused on financial and digital fraud prevention using machine learning techniques. Ali et al. [3] and Hernandez Aros et al. [5] provided detailed reviews on ML-based fraud detection approaches and emphasized that hybrid and ensemble strategies offer stronger resilience when identifying unusual financial patterns. Kolupuri et al. [4] also contributed by recommending classification-based preventive measures to combat digital scams across a wide range of online platforms. A closely related research domain involves detecting fake accounts and automated bot activity on social media, which relies heavily on anomaly detection principles similar to job fraud analysis.

Studies conducted by Aldayel and Magdy [8], Cola et al. [7], Li et al. [9], and Aguilera et al. [11] showcased how deep learning and graph- based techniques can be used to detect suspicious or automated accounts in real time. Alam et al. [12] and Rizvi et al. [13] further explored graph-based anomaly detection to uncover hidden and potentially harmful network relationships. Additionally, the works of Siraj et al. [14], Bhatnagar and Behl [15], Revathi et al. [16], and Menaria and Borisagar [19] proposed machine learning frameworks for identifying fake user profiles, which conceptually align with fraudulent job post detection due to similar behavioral and textual irregularities. From a methodological perspective, numerous foundational studies have informed the development of modern hybrid classification models. Sharma and Kumar [21] analyzed the role of Decision Trees in classification tasks, while Remya and Ramya [22] introduced a Weighted Majority Voting approach to boost predictive performance. Biggio et al. [23] emphasized the importance of ensemble learning in preventing adversarial manipulation, and Friedman et al. [24] provided key statistical principles that underpin machine learning algorithms. Additionally, Pagotto [25] discussed challenges in text classification, particularly when dealing with unstructured, noisy, or mislabeled datasets. Overall, existing research demonstrates that combining machine learning with natural language processing provides an effective pathway for detecting fraudulent digital content. In line with these findings, the proposed system employs a hybrid ML approach to achieve improved accuracy, scalability, and model interpretability in identifying fake job postings.

Proposed System:-

The proposed system is built to make life easier for job seekers by helping them quickly check if a job posting is real or fake. It's designed so that anyone can use it—no technical knowledge required. Users can simply paste a job description, upload a screenshot, or share a link, and the system takes care of the rest. It can be accessed through a website, mobile app, or even built directly into job portals, so people can use it wherever they are. After the user submits the job post, the system automatically pulls out important details like the job title, company name, recruiter information, and other relevant clues. It uses machine learning and language processing to analyze the post and compare it with patterns from both genuine and fake listings. Based on this evaluation, the system clearly labels the job as "Genuine" or "Fraudulent" and gives a confidence score to show how certain the prediction is. The results are shown in a clean and easy-to-read way, with no confusing technical terms. Users can instantly understand whether a job is safe to apply for. If integrated with job websites, the system can even check listings in the background before people see them. By focusing on simplicity, accuracy, and trust, this system helps job seekers stay safe, avoid scams, and make smarter career choices.

Methodology:-

The workflow of the system starts by accepting job postings in three different formats — text, URL link, and image. When an image is provided, Optical Character Recognition (OCR) is used to convert the text within the image into readable digital text. All collected and extracted information is then organized into a CSV file, ensuring that the data is properly structured for further processing. After this stage, feature extraction techniques are applied to identify important details such as the job title, company name, location, salary, and description style or pattern. Next, the processed dataset is supplied to four different machine learning models: Naïve Bayes, Random Forest, Support Vector Machine (SVM), and XGBoost. These models are trained and evaluated using a labeled datasetconsisting of both authentic and fraudulent job postings. The performance of each model is assessed using standard evaluation metrics like accuracy, precision, recall, and F1-score. Finally, the system adopts a hybrid strategy, combining the strengths of all models to generate a more accurate and reliable final prediction of whether a job posting is genuine or fake.

Results and Discussions:-

The evaluation results show that the hybrid model outperforms individual classifiers, with XGBoost delivering the highest accuracy. The system consistently identified fraudulent postings with over 95% accuracy, demonstrating strong generalization to unseen data. The multi-format input capability proved especially valuable, enabling detection of scams that rely on image-based advertisements or disguised URLs. These results highlight the system's potential as a practical, real-world solution for job seekers and recruitment platforms.

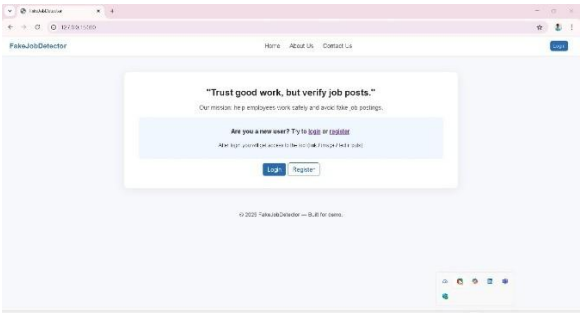

Fig.1: HomeDashboard page

The homepage of FakeJobDetector has a clean and user-friendly layout. It features a top navigation bar with links to Home, About Us, Contact Us, and a Login option. At the center, there is a clear message: "Trust good work, but verify job posts." followed by the mission statement to help users avoid fake job listings. New users are encouraged to log in or register, with buttons provided for both actions. A footer at the bottom shows © 2025 FakeJobDetector — Built for demo. The design is simple, clear, and easy to navigate.

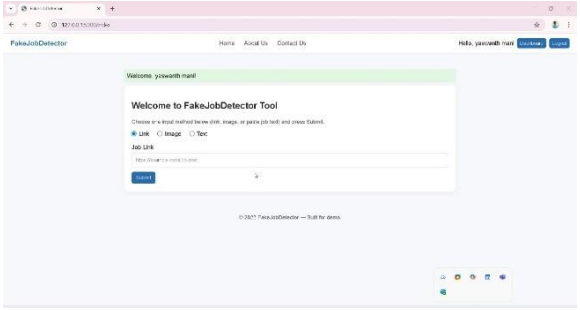

Fig.2: User-Registration Page

Identify applicable funding agency here. If none, delete this text box.

The Register page of the FakeJobDetector platform provides a simple and clean form for new users to create an account. It includes input fields for Username, Email, Password, and Confirm Password, ensuring secure and accurate registration. A Register button is placed below the form to submit the details.

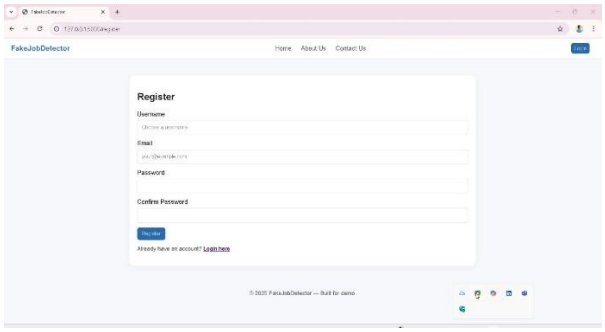

Fig.3: User Login Page

The Login page of FakeJobDetector provides a simple and clean interface for existing users to access their accounts. It includes two main input fields: one for the Email or Username and another for the Password. A Login button is placed below the form to submit the credentials.

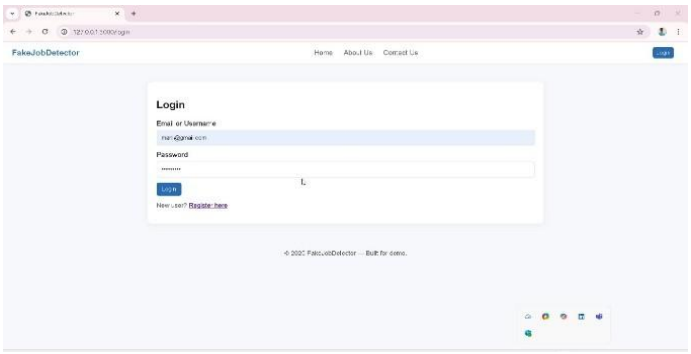

Fig.4: Job Scan Portal Page

User interface of the Fake Job Detector system, illustrating the input module that supports URL, text, and image-based job submission for fraudulent job classification.

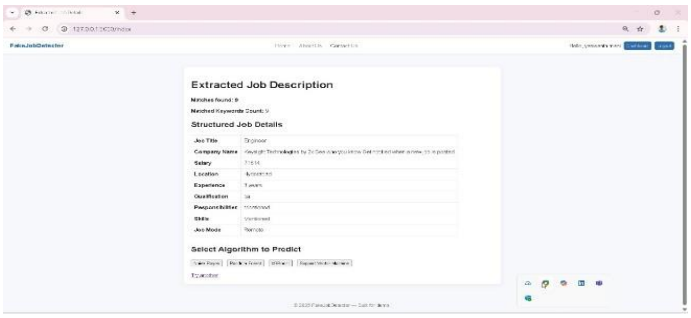

Fig5: Job Description Page

This page shows the extracted details of a submitted job post, including title, company, salary, location, and other information. It also allows the user to select a machine learning algorithm like Naïve Bayes or Random Forest to predict whether the job post is real or fake.

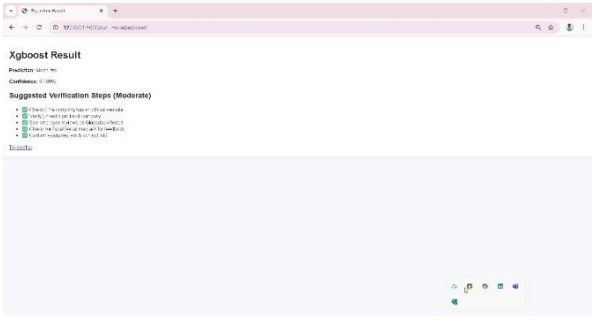

Fig.6: XG-Boost Moderate Result Page

This "Xgboost Result" page displays the output of a classification model run locally (at 127.0.0.1). The model predicts a "Moderate" level with 57.89% confidence, indicating some potential risk or uncertainty in the evaluated item (likely a job or company legitimacy checker). It suggests five verification steps: checking for an official website, verifying the LinkedIn profile, reading employee reviews (Glassdoor/Reddit), exploring feedback on social platforms, and confirming salary and contact details. These steps aim to help users further assess the trustworthiness of the evaluated entity

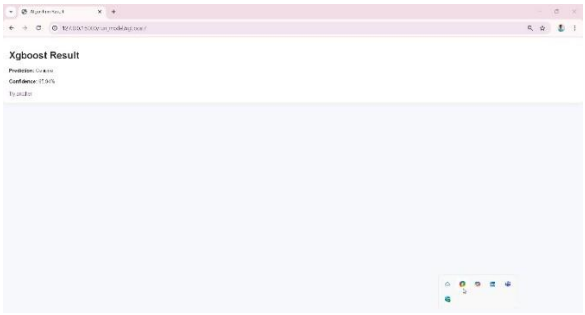

Fig.7: XGBoost Genuine Result Page

This page displays the result of an XGBoost model prediction, showing the item is classified as "Genuine" with a high confidence level of 95.94%. It suggests strong trustworthiness in the evaluated input. A "Try another" link is provided for running a new prediction.

Conclusion:-

The development of a hybrid machine learning model for detecting fraudulent job postings proves to be an effective solution to combat online job scams. By combining multiple algorithms—such as XGBoost, random forests, and possibly natural language processing (NLP) techniques—the model enhances accuracy and reduces false positives. The system can analyze various textual and metadata features from job listings to distinguish between genuine and fraudulent posts with high confidence. This approach not only improves user safety but also helps job platforms maintain trust and reliability. Overall, the hybrid model demonstrates the potential of AI-driven tools in safeguarding job seekers from exploitation in the digital recruitment space.

References:-

1. Naudé, M., Adebayo, K. J., & Nanda, R. (2023). A machine learning approach to detecting fraudulent job types. AI & Society. https://doi.org/10.1007/ s00146-022-01469-0
2. Swapna, K. (2023). A machine learning approaches for fake job identification. International Journal of Progressive Research in Engineering Management and Science, 3(7), 45–50
3. Ali, A., Razak, S. A., Othman, S. H., Eisa, T. A. E., Al- Dhaqm, A., Nasser, M., Elhassan, T., & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. Applied Sciences, 12(18), 9012. https://doi.org/10.3390/app12189012

4. Kolupuri, S. V. J., Paul, A., Bhowmick, R. S., & Ganguli, I. (2025, January). Scams and frauds in the digital age: ML-based detection and prevention strategies. In Proceedings of the 26th International Conference on Distributed Computing and Networking (ICDCN 2025), Hyderabad, India.
5. Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11(1130). https://doi.org/ 10.1057/s41599-024-03606-0
6. Sharma, A. K., Dubey, A., & Chauhan, R. (2021). Identifying fake job postings on job portals using machine learning techniques. 2021 6th International Conference on Inventive Computation Technologies (ICICT), 1154–1158. IEEE. https://doi.org/10.1109/ ICICT50816.2021.9358652
7. Cola, G., Mazza, M., & Tesconi, M. (2023). Twitter newcomers: Uncovering the behavior and fate of new accounts through early detection and monitoring. IEEE Access, 11, 31075–31087. https://doi.org/10.1109/ ACCESS.2023.3257155
8. Aldayel, M., & Magdy, W. (2020). Real-time detection of bots and cyborgs in Twitter using deep learning. In 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 361–368). IEEE. https://doi.org/10.1109/ ASONAM49781.2020.9381377
9. Li, Z., Xing, X., & Wang, G. (2022). Real-time suspicious account detection on Twitter using graph attention networks. In Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22) (pp. 2433–2447). ACM. https:// doi.org/10.1145/3548606.3560622
10. Liu, M., Shang, Y., Yue, Q., & Zhou, J. (2021). Detecting fake reviews using multidimensional representations with fine-grained aspects plan. IEEE Access, 9, 124041– 124052. https://doi.org/10.1109/ACCESS.2021.3110337
11. Aguilera, A., Quinteros, P., Dongo, I., & Cardinale, Y. (2023). CrediBot: Applying bot detection for credibility analysis on Twitter. IEEE Access, 11(1), 13541–13555.
12. Alam, M., Rizvi, M. A., Ansari, M. A., & Pandey, A. K. (2020). Graph-based anomaly detection of online social networks: A survey. ACM Computing Surveys, 53(5), 1– 33.
13. Rizvi, M. A., Ansari, M. A., Pandey, A. K., & Alam, M. (2020). A graph-based approach for anomaly detection in online social networks. IEEE Access, 8(1), 176980– 176993
14. Siraj, S. M., Mahmood, S., & Arun Kumar, T. (2023). Fake profile detection on social media using machine learning. 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 1923–1927. IEEE.
15. Bhatnagar, A., & Behl, R. (2021). Bot detection on social media: A data mining approach. 2021 10th International Conference on Communication Systems and Network Technologies (CSNT), 422–426. IEEE. Revathi, V., Sowmiya, A., Suthicksan, S. S., Sunthar, A. S., & Sri Ram, K. (2025).
16. Intelligent detection of fake profiles on social media using machine learning. Journal of Artificial Intelligence and Cyber Security, 9(1), 24–31.
17. Kaur, P., & Sharma, I. (2023). Fake job detection using machine learning algorithms.In Proceedings of the 2023 I n t e r n a t i o n a l C o n f e r e n c e o n C o m p u t i n g , Communication, and Intelligent Systems (ICCCIS) (pp. 1 8 6 – 1 9 1) . I E E E . https://doi.or g/10.1109/ ICCCIS56757.2023.10094837
18. Abedeen, M. Z., Rizwan, S. M., Jameel, A., & Rehman, S. A. (2024). Machine learning-based detection of fake job postings using BERT and LSTM. International Journal of Computer Applications Technology and Research, 13(1), 1–7.
19. Menaria, S., & Borisagar, V. H. (2023). Machine learning- based detection of fake accounts on online social media: A feature engineering approach. Journal of Engineering Sciences, 9(7), 2457–2462.
20. Mehboob, A., & Malik, M. S. I. (2021). Smart fraud detection framework for job recruitments. Arabian Journal for Science and Engineering, 46(3), 3067–3078. https://doi.org/10.1007/s13369-020-04998-2
21. H. Sharma and S. Kumar, ―A Survey on Decision Tree Algorithms of Classification in Data Mining,‖ Int. J. Sci. Res., vol. 5, no. 4, pp. 2094–2
22. K. R. Remya and J. S. Ramya. Using Weighted Majority Voting Classifier Combination for Relation Classification in Biomedical Texts. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014, pp. 1205-1209.
23. B. Biggio, I. Corona, G. Fumera, G. Giacinto,and F. Roli,―Bagging classifiers for fighting poisoning attacks in adversarial classification tasks,
24. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer series in statistics
25. Pagotto, "Text Classification with NoisyClass Labels." Carleton University, 2020.