International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Fake Job Posting Detection

Omkar Kulkari, Pranav Shinde, Ravichandra Mulage, Saurabh Martande, Sneha Kanawade, Arti Singh,

Dr D.Y. Patil Institute of Engineering, Management and Research, Akurdi, Pune

opkulkarni0104@gmail.com, pranavshinde4019@gmail.com, ravichandramulage00@gmail.com, smartande8055@gmail.com, sneha.kanwade@dypiemr.ac.in , arti.singh@dypiemr.ac.in,

Abstract:

The Fake Job Posting Detection Project is a new response to the limitations of traditional work patterns in an evolving and changing society. He acknowledges that the lack of flexibility in existing names and descriptions can stifle creativity and limit many interests. Additionally, a biased relationship with the job can limit understanding of the value of the job.

To solve these problems, the Fake Jobs program offers fake job titles and descriptions that go beyond traditional standards. These role perspectives combine diversity and expertise, allowing individuals to explore unconventional careers. The program's main goal is to engage in a discussion about innovation and personal fulfillment in the professional world, encouraging people to rethink how their interests and skills are incorporated into career choices. Essentially, it encourages breaking away from traditional work patterns and creating an environment where people can reach their potential in their work.

Keywords: Web Scrapping, Online recruitment fraud, Fraud detection, Employment scam, Contextual features.

1. Introduction:

It is very difficult to analyze and predict online recruitment information from various online sites. Determining whether a job posting is genuine requires in-depth research. If such fraudulent ads are detected and removed, job seekers will be able to focus on the real jobs advertised. Especially due to the economic crisis and the impact of the coronavirus, there was a decrease in the number of jobs and an increase in the number of unemployed people. This has led to some computer fraud and digital crimes where many applicants lose. Recently, many companies prefer to announce job openings online so that job seekers can easily reach them and access existing connections. However, scammers make mistakes in this regard by using company and business information to trick candidates into obtaining personal and financial information. Scammers often use job seekers' money to secure jobs and/or infiltrate bank accounts. Also increase their reputation and credibility by publishing fake ads for well-known companies. This article is for spotting inaccurate online recruiting information. This search for fake jobs has had a huge impact astechs identify fake jobs and use them to tell people not to apply for them. Our Reveal app uses input (URL) from the user to check if the job posting is real or fake and then gets listed on specific websites like True, Dab, quikr, and more. Once the login URL link is verified, the login page will start, the URL will be converted to Html format and we will save the main content. This information will be verified with the raw data provided on our advertising website.

1.1. Contribution of the proposed model

Primary Contributions of the Fake Job Detection Model:-

1. **Increased Effectiveness:**

- The fake job search model improves the overall results of the online recruitment process by providing a reliable and fast process to distinguish between real and fake jobs. This helps provide a safer and more reliable environment for job seekers, allowing them to focus their efforts on legitimate opportunities without fear of falling victim to fraud.

2. **Increased Accuracy:**

- This model improves the accuracy of identifying fake hiring information using advanced algorithms and data analysis. Using machine learning and pattern recognition, the model is able to detect many negative aspects and indicators about job postings, thereby increasing the accuracy of distinguishing between real and fake pseudonyms. This increase in accuracy is necessary to build trust in the job search community.

3. **Scalability:**

- The fake job detection model demonstrates the ability to adapt to various online recruitment platforms and websites. Whether it's a popular job site, company website, or social media platform, this model can scale its operations to cover a wide range of online recruiting information.

4. **Discovering Hidden Insights:**

- The fake job detection model demonstrates the ability to adapt to various online recruitment platforms and websites. Whether it's a popular job site, company website, or social media platform, this model can scale its operations to cover a wide range of online recruiting information. This scalability adapts to the ever-changing online recruiting landscape, ensuring the solution remains effective and relevant across multiple digital platforms**.**

5. **Giving Decision-Makers More Power:**

- This model provides decision makers in the hiring process with a tool that supports their decision-making skills. Recruiters, recruiters, and job managers have more control over the quality and accuracy of job openings on their platforms. This management not only improves the overall user experience but also helps maintain reputation and trust in online recruiting.

2. Literature Review:

Index no.	Paper Title	Main Focus	Relevance to Diamond Price Prediction
[1]	Fake Job Recruitment Detection Using Machine Learning Approach	Fake Recruitment detection using machine learning methods	To avoid fraudulent posts for jobs on the internet, an automated tool using machine learning-based classification techniques is proposed in the paper. Different classifiers are used to check fraudulent posts on the web and the results of those classifiers are compared to identifying the best employment scam detection model. It helps in detecting fake job posts from an enormous number of posts.

[2]	Detection of Fake Job Recruitment Using Machine Learning	Using machine learning to predict fake jobs	An application using machine learning-based categorization algorithms are presented in the project to prevent fraudulent job postings online. The outputs of various classifiers are evaluated to determine the best employment scam detection model.
[3]	Fake Job Detection Using Machine Learning Approach	Analyzing data of different fake jobs using various data analysis techniques and detecting using machine learning algorithms.	Advertising new job openings has recently become a very prevalent problem in the modern world as a result of advancements in social communication and modern technologies. Therefore, everyone will have a lot of reason to be concerned about bogus job postings. Fake job-posing prediction presents a variety of difficulties, much like many other categorization problems.

3. Related work:

Fake job postings have become a significant threat to job seekers, wasting time, effort, and potentially exposing personal information. Automatic detection of these fraudulent postings is crucial for maintaining a safe and secure online job search experience.

Several research efforts have explored techniques for fake job posting detection. Here's a breakdown of some key approaches:

Machine Learning: This is a prominent approach, employing supervised learning algorithms like Random Forest, Support Vector Machines (SVM), and Naive Bayes. These algorithms are trained on datasets containing real and fake job postings, allowing them to identify patterns and features indicative of fraudulent activity.

Knowledge-Based Detection: This strategy leverages existing knowledge bases and fact-checking websites to verify information within job postings. It can check the legitimacy of company websites, addresses, and other details mentioned in the post.

Natural Language Processing (NLP) technology: NLP is widely used to analyze the content of job postings for signs of fraud. Yang et al. (2020) proposed a new method based on semantic analysis and pattern analysis to detect suspicious patterns in job descriptions. Additionally, Zhang et al. (2021) studied how to use emotional intelligence and cues to distinguish legitimate job postings from fake ones.

Image-based methods: Some researchers have adopted image models to model relationships between organizations such as job postings, employers, and people. For example, Chen et al. (2019) developed a graphical representation of job posting data and used a graph embedding technique to identify negative patterns indicating fraud.

Crowdsourcing and human intelligence: Many studies have explored the integration of crowdsourcing and human intelligence to improve the search for information sources. not true. Liu et al. (2017) proposed a hybrid approach that combines machine learning algorithms with human feedback to improve classification accuracy.

Data mining and pattern recognition: Researchers use data mining and pattern recognition techniques to identify recurring patterns in fake ads. For example, Kim et al.

(2018) investigated the popularity of some scam-related keywords and phrases and developed rules for flagging suspicious names.

Cross-domain analysis: Some studies have attempted cross-domain analysis using data from different sources such as information sources, company websites, and user reviews to verify whether the legitimacy of the study is sufficient. This collaboration holds promise for reducing the limitations of individual devices.

In summary, the literature on detecting misinformation is diverse, including various methods such as machine learning and NLP techniques for image patterns, and human-computer interaction techniques. Despite significant progress, challenges such as insufficient data, class conflicts, and ever-changing fraud strategies remain, and more research, more research, and innovation are needed in this area.

Models	Algorithms	Description	Features	Research Gap
J48 Decision Tree	Decision Tree	The J48 decision tree algorithm consistently showed improvement in performance measures (accuracy, precision, recall) when contextual features were included.	Content-based and contextual features	Previous studies have overlooked the impact of localized factors in fake job detection, but this study addresses this gap by considering contextual features in the analysis.
Naive Bayes	Naive Bayes	Despite a slight decrease in recall, the Naive Bayes algorithm still demonstrated overall improvement in accuracy and precision when contextual features were incorporated.	Content-based and contextual features	Prior research lacked emphasis on localized factors, but this study fills this gap by examining the impact of contextual features on detecting fake job postings.
Random Forest	Random Forest	Among the tested algorithms, Random Forest exhibited the best performance, with significant enhancements in accuracy, precision, and recall upon inclusion of contextual features.	Content-based and contextual features	Previous studies did not sufficiently explore the influence of localized factors on fake job detection. This study contributes by examining how contextual features improve detection accuracy.

Table 3.1 Systematic Analysis of the ML models

4. System Methodology:

4.1. The system model contains the following components:

1. SaaS WebApp
2. Web Scraping
3. Data preprocessing
4. Machine learning libraries

4.2. System Architecture:

4.2.1. The flow of the architecture is as follows:

- Login as URL: It is related to how system information is written. In this case, the system retrieves the job posting information from the following URL.
- Web scraping: This step involves extracting relevant information from a particular URL. This may include a description in the job posting, salary range, company name, and other details.
- Preliminary data: After the data is captured, the machine must first be prepared for learning. This may include cleaning up files, deleting irrelevant files, and formatting regular files.
- Feature extraction: In this stage, important features are extracted from the previous data. A machine learning model uses these features to identify patterns that distinguish real job postings from fake ones. Learn and train data using ML algorithms: Learn machine learning algorithms here. Preliminary data is divided into two groups: training and testing. The training process is used to train machine learning models to identify features that differentiate real posts from fake posts. Once the model is trained, it is evaluated using a benchmark to measure its accuracy in classifying new tasks.
- Data Extraction - Detection of Fake or Real Jobs: Finally, once the model is trained, new job postings can be fed into the system. Features are extracted from these novels and the training model is used to classify the novel as true or false.

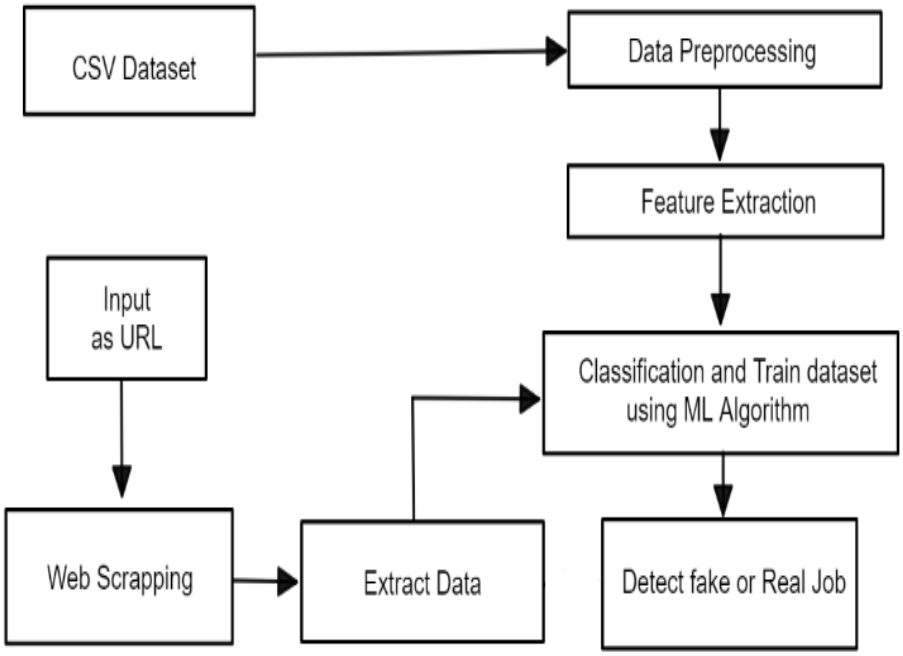

Fig. 1.1: System architecture model

4.2.2. E2E (End to End) Data flow architecture:

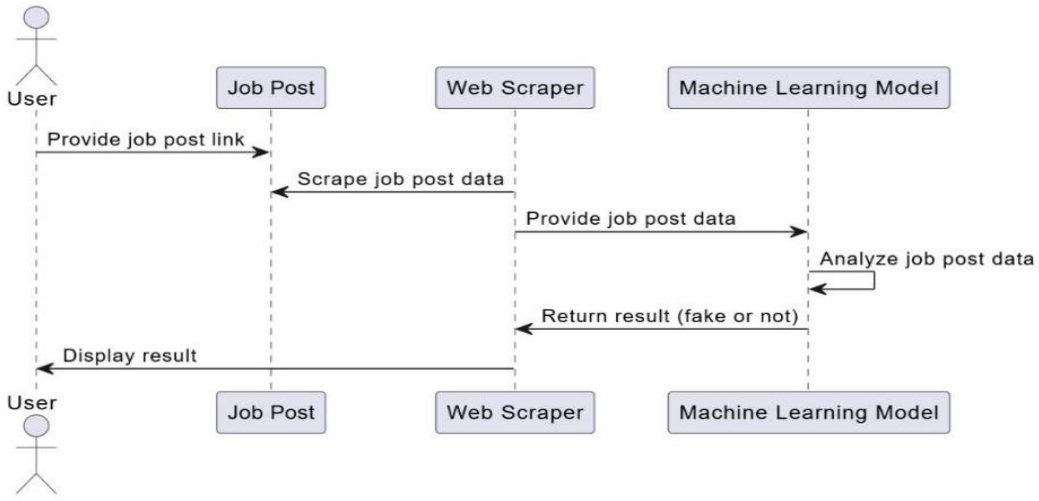

Fig. 1.2. Data Flow Architecture

4.2.3. Evaluation Parameters

Sr. No.	Parameters	Description
1.	F1 Score	The harmonic mean of precision and recall. Provides a balance between precision and recall, suitable for imbalanced datasets.
2.	Confusion Matrix	A table that shows the number of true positives, false positives, true negatives, and false negatives. Offers detailed insights into the model's performance across different classes.
3.	RMSE	Root Mean Square Error. Measures the average magnitude of the errors between predicted and actual values. Lower values indicate better model performance.
4.	MAE	Mean Absolute Error. Measures the average magnitude of the errors between predicted and actual values. Similar to RMSE but does not penalize large errors as heavily

5. Results and Discussion

The research paper investigates the detection of fraudulent job postings online, which pose a significant challenge to job seekers amidst the proliferation of digital recruitment platforms. By introducing the web app, users can input job posting URLs to verify their authenticity, thus mitigating the risk of falling victim to scams. The study emphasizes the importance of technological solutions and collaborative efforts to combat fraudulent activities effectively. Through continuous monitoring and adaptation of detection techniques, the research contributes to creating a safer online job market for job seekers. Future research may focus on refining detection algorithms and fostering greater collaboration across industry.

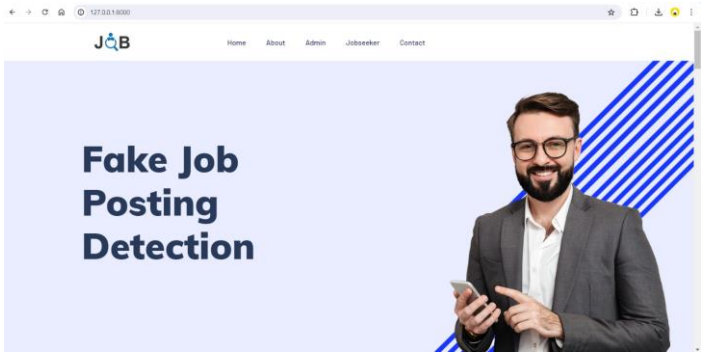

Fig. 1.3. Homepage of the application

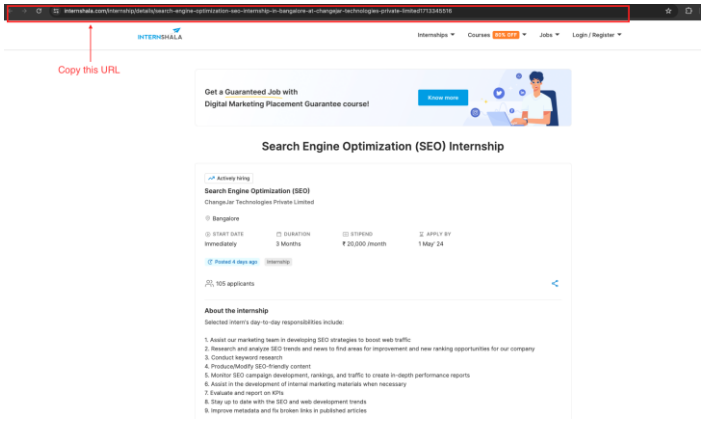

Fig. 1.4. Copy URL from Internshala

Fig. 1.5. URL pasting page

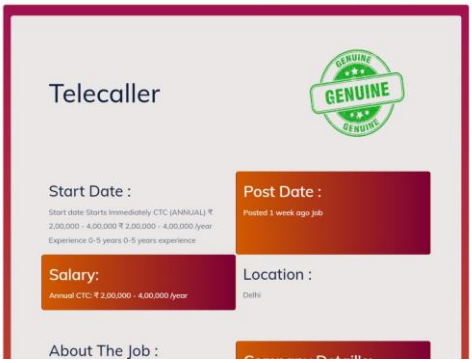

Fig. 1.6. Genuine Result after detection

Fig. 1.7. Fake Result after detection

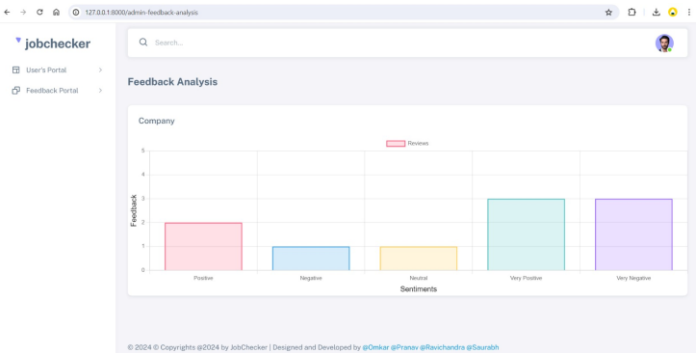

Fig. 1.8. Graph report of feedback

5.1. Advantages

- Rich data integration: Integration of different cases across various online recruitment platforms help increase the depth and accuracy of fraud detection and provide more information.

- Psychological Research: Considering Psychological Aspects in Future Research

Reflects a positive approach to understanding cheating behavior, which can lead to many negative and impactful investigations

5.2. Limitations

- Adaptation Race Challenge: The continuous race between fraudsters' adaptation and detection strategy evolution highlights the need for an agile and responsive system to stay ahead in a dynamic environment.

- Incomplete Medium Coverage: The acknowledgment of potential gaps in the coverage of fraudulent behaviors in mediums such as emails or smartphone text messages suggests that the current framework may not be fully comprehensive, requiring expansion into these areas.

6. Conclusion & Future Scope

6.1 Conclusion:

Platforms such as online job portals or social media for job advertisements are an exciting way of attracting potential candidates on which many enterprise companies are dependent on the hiring process. Fake job scam detection at an early stage can save job seekers and make them only apply to legitimate companies. For this purpose, various machine-learning techniques were utilized in this paper. Specifically, supervised learning algorithms classifiers were used for scam detection. This paper experimented with different algorithms such as naïve Bayes, SVM, decision tree, random forest, and K-Nearest Neighbor.

6.2 Future scope:

Integration with Job Search Platforms: Collaboration with job search platforms like LinkedIn, Indeed, or other job boards to integrate fake job posting detection tools directly into their systems can help in real-time monitoring and removal of suspicious listings.

User Feedback and Crowdsourcing: Incorporating user feedback and crowdsourced data to improve the accuracy of the detection system. Users can report suspicious job postings, and this feedback can be used to refine the algorithms.

7. References:

[1]Alghamdi, B., & Alharby, F. (2019). An intelligent model for online recruitment fraud detection. Journal of Information Security, 10(03), 155. https://www.scirp.org/journal/paperinformation.aspx?paperid=93637

[2]Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. Information Sciences, 497, 38-55. https://app.dimensions.ai/details/publication/pub.1114201506

[3]Islam, M. R., Liu, S., Wang, X., & Xu, G. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. Social Network Analysis and Mining, 10(1), 1-20. https://doi.org/10.1007/s13278-020-00696-x

[4]Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. arXiv preprint arXiv:1804.08559.

[5]Mitra, T., & Gilbert, E. (2015). Credbank: A large-scale social media corpus with associated credibility annotations. In Ninth international AAAI conference on web and social media. https://ojs.aaai.org/index.php/ICWSM/article/view/14625

[6]Mujtaba, G., & Ryu, E. S. (2020). Client-driven personalized trailer framework using thumbnail containers. IEEE Access, 8, 60417-60427. https://ieeexplore.ieee.org/document/9046852

[7]Mujtaba, G., & Ryu, E. S. (2021). Human Character-oriented Animated GIF Generation Framework. In 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC) (pp. 1-6). IEEEE

[8] Thangiah, Murugan; Basri, Shuib; Sulaiman, Suziah, “A framework to detect cybercrime in the virtual environment”, International Conference on Computer & Information Science (ICCIS), 2012, pp. 553–557.

[9] Datta, Priyanka; Panda, Surya Narayan; Tanwar, Sarvesh; Kaushal, Rajesh Kumar, “A Technical Review Report on Cyber Crimes in India”, International Conference on Emerging Smart Computing and Informatics (ESCI), 2020, pp. 269–275.

[10] Sajal, Sayeed Z.; Jahan, Israt; Nygard, Kendall E, “A Survey on Cyber Security Threats and Challenges in Modern Society”, IEEE International Conference on Electro Information Technology (EIT), 2019, pp. 525–528.

[11] Gunjan, Vinit Kumar; Kumar, Amit; Rao, Allam Appa, “Present & Future Paradigms of Cyber Crime & Security Majors - Growth Rising Trends”, 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, 2014, pp. 89–94.

[12] Govil, Jivesh; Govil, Jivika, “Ramifications of cybercrime and suggestive preventive measures”, IEEE International Conference on Electro/Information Technology, 2007, pp. 610–615.

[13] Rok Bojanc and Borka Jerman-Blažič, “Standard approach for quantification of the ICT security investment for cybercrime prevention”, Second International Conference on the Digital Society, 2008, pp. 7-14

[14] Sinchul Back, Jennifer LaPrade, “The Future of Cybercrime Prevention Strategies: Human Factors and A Holistic Approach to Cyber Intelligence”, International Journal of Cybersecurity Intelligence and Cybercrime, 2019, pp.1-4.

[15] M. Sathiyanarayanan, C. Turkay, and O. Fadahunsi. "Design of Small Multiples Matrix-based Visualisation to Understand E-mail Socioorganisational Relationships." In 2018 10th International Conference on Communication Systems & Networks (COMSNETS). 2017.