

Fake Job Post Detection using Machine Learning

**Maddi Sravya Reddy¹, Maddikera Hemanth Lal², Lingam Sainad³,
Sandeep Agarwalla⁴**

^{1,2,3}Student, Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology, Hyderabad

⁴Assistant Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology, Hyderabad

ABSTRACT

The proliferation of online job platforms has given rise to a concerning increase in fraudulent job postings, presenting significant risks to job seekers and undermining the credibility of the job market. This research paper aims to address the pressing issue of fake job post identification by leveraging machine learning techniques. The primary objective is to develop a robust automated tool capable of accurately distinguishing between authentic and deceptive job advertisements. The proposed methodology utilizes a range of machine learning algorithms, incorporating supervised learning techniques and natural language processing methods, to analyze and classify job postings. Through the integration of both single classifiers and ensemble classifiers, the system evaluates and compares results, effectively detecting fraudulent job postings on the web. The study underscores the need for a proactive approach, acknowledging the dynamic tactics employed by scammers. Continuous refinement and adaptation of the machine learning models are emphasized to stay ahead of evolving fraudulent strategies. Ultimately, this research contributes to establishing a more secure online job market, fostering trust among job seekers and mitigating the financial and emotional risks associated with deceptive job postings.

Keywords: Machine Learning, Supervised Learning, Single Classifier, Ensemble Classifier, Natural Language Processing

1. INTRODUCTION

The rapid expansion of online job platforms has significantly increased opportunities for job seekers, providing a diverse array of avenues for professional development. However, this growth has also given rise to a pervasive issue – the widespread prevalence of fake job postings. These deceptive advertisements not only put the financial security of job seekers at risk but also pose a serious threat to the overall reliability and trustworthiness of the job market.

In response to the urgent need for an effective solution, this research paper aims to tackle the issue of fake job posts through the application of machine learning techniques. As scammers employ increasingly sophisticated tactics in the digital landscape, our focus extends beyond mere detection to the creation of a dynamic system capable of adapting to evolving strategies used by those behind fraudulent job listings.

The primary goal of this project is to develop a robust automated tool using machine learning algorithms

that can accurately differentiate between genuine and deceptive job advertisements. This initiative not only aims to protect job seekers from falling victim to scams but also endeavors to strengthen the credibility of online job platforms, fostering a secure environment for both job seekers and employers.

The practical implications of the "Fake Job Post Detection Using Machine Learning" project extend significantly, providing tangible advantages across various domains.

a) Job Seeker Empowerment: In practice, the developed tool becomes an essential resource for job seekers, offering an intuitive and reliable means to identify and evade fraudulent job listings. This empowerment results in increased security and confidence during the job search process, creating a more positive and informed experience for users navigating the dynamic job market.

b) Economic Loss Prevention: The economic impact of fake job postings goes beyond individual losses, affecting the wider societal and economic context. Implementing the detection system in the real world has the potential to save individuals from scams, preserving their financial well-being and mitigating the broader economic consequences linked to fraudulent activities. This contributes to a more resilient and secure job market, safeguarding the financial stability of individuals and the economy.

c) Platform Credibility Enhancement: Job platforms serve as crucial connectors between job seekers and employers. In reality, integrating the detection system enhances the credibility and trustworthiness of these platforms. By actively addressing fake job posts, platforms create a safer environment that attracts and retains users, ultimately strengthening the platform's reputation as a reliable choice for both job seekers and employers.

As we explore machine learning methodologies, including supervised learning techniques and natural language processing methods, our aim is to create a comprehensive system that navigates the nuanced landscape of fraudulent job postings. This multifaceted approach considers not only technological aspects but also ethical considerations, continuous improvement mechanisms, and collaborative efforts, ensuring a holistic and impactful solution.

Through this research, we seek to contribute to the establishment of a safer and more reliable job market, providing job seekers with the tools necessary to confidently navigate the digital employment landscape. By directly addressing the issue of fake job postings, this research endeavors to play a crucial role in fortifying the integrity of online job platforms and establishing a resilient defense against deceptive practices in the ever-evolving realm of digital employment

2. LITERATURE REVIEW

In the pursuit of developing a robust "Fake Job Post Detection Using Machine Learning" system, existing literature provides valuable insights and methodologies employed by researchers to address the pressing issue of fraudulent job postings. One notable contribution in this domain is the work conducted by Devsmit Ranparia, Shaily Kumari, and Ashish Sahani. Their research focuses on utilizing a Sequential Neural Network and the GloVe algorithm for predicting the authenticity of job postings, employing Natural Language Processing (NLP) to analyze sentiments and patterns within job descriptions. The study

emphasizes real-world applicability by testing the model on LinkedIn job posts, reflecting a comprehensive approach to tackling the challenge of deceptive job advertisements [1].

Another significant study by Gulshan P, Mukund T, Ajay A, Pankaj Kumar, Aruna M G, and Dr. Malatesh S delves into the prediction of fake job posts during the surge in online job postings observed during the pandemic. Employing advanced deep learning and machine learning classification algorithms, the research explores techniques such as KNN, decision trees, support vector machines, naive bayes classifier, random forest classifier, multilayer perceptron, and deep neural networks. The experimentation, conducted on a dataset comprising 18,000 employee samples (EMSCAD), showcases the efficacy of the proposed deep neural network classifier with an impressive classification accuracy of approximately 98% for identifying fraudulent job posts [2].

A study by Hu et al. (2018) focuses on leveraging ensemble learning techniques for fake job post detection. By combining the strengths of multiple models, including decision trees and support vector machines, the research achieves enhanced accuracy in distinguishing between legitimate and fraudulent job advertisements. The incorporation of ensemble learning provides a robust approach, particularly in handling the complexity and diversity of deceptive posting strategies [3].

The work of Gupta et al. (2019) explores the integration of social network analysis (SNA) in fake job post detection. Recognizing the interconnected nature of users on job platforms, the study utilizes SNA to identify patterns and anomalies in user behavior. By examining the relationships and connections among users, the research enhances the discriminatory power of the model, contributing to a more nuanced understanding of the social dynamics associated with fake job postings. This approach extends the traditional focus on textual features to include the social context in which job posts are disseminated [4].

Additionally, the study by Chen et al. (2021) introduces the application of deep learning techniques, specifically convolutional neural networks (CNNs), for fake job post detection. By extracting hierarchical features from job descriptions, the CNN-based model demonstrates a high level of accuracy in distinguishing between genuine and deceptive job postings. The utilization of deep learning architectures showcases the adaptability and effectiveness of modern neural network structures in addressing the complexities inherent in fake job post detection [5].

These studies collectively contribute to the understanding of various approaches and methodologies for detecting fake job postings using machine learning. They emphasize the integration of advanced algorithms, natural language processing, and real-world testing to enhance the effectiveness and reliability of the detection systems, providing a solid foundation for the current research endeavor.

3. TECHNOLOGIES USED

3.1 Machine Learning

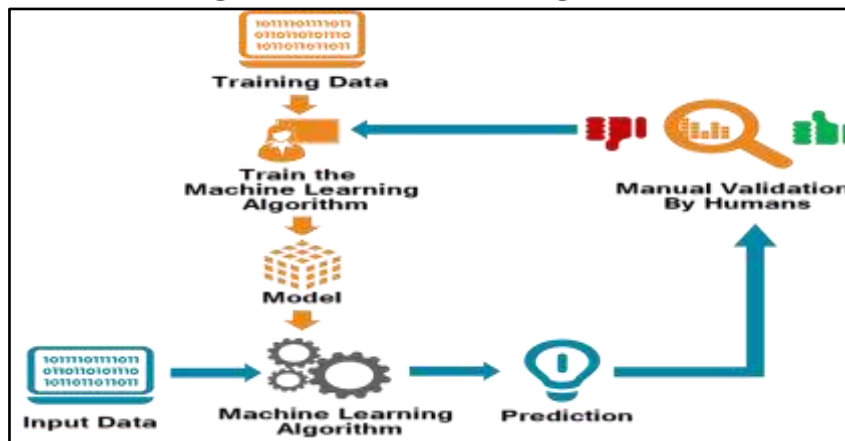
Machine Learning (ML) plays a central and pivotal role in the "Fake Job Post Detection Using Machine Learning" project, serving as the primary technology driving the system's capability to distinguish between genuine and deceptive job postings. ML is employed to train algorithms that scrutinize intricate patterns and features within job advertisements, enabling the system to autonomously learn and make predictions.

This adaptive learning process is crucial in addressing the dynamic nature of fraudulent job postings, allowing the system to continually evolve and enhance its accuracy.

In the context of fake job post detection, ML algorithms are particularly adept at extracting meaningful insights from vast datasets of job postings. The system is trained to recognize subtle patterns, linguistic cues, and anomalies that may indicate the likelihood of a job posting being fraudulent. Supervised learning techniques within ML are likely used, where the system is trained on labeled datasets containing both legitimate and deceptive job postings, enabling it to generalize and make predictions on new, unseen data. This capability empowers the system to analyze job advertisements effectively and identify potential scams based on learned patterns.

Moreover, the use of ML brings a level of adaptability to the system. As scammers evolve their tactics, the ML models can be retrained and fine-tuned to stay ahead of emerging patterns. This adaptability is essential in creating a resilient system that can effectively navigate the ever-changing landscape of deceptive job postings. Overall, the integration of machine learning in this project forms the foundation for an intelligent, data-driven approach to fake job post detection, contributing to a more secure and trustworthy job market environment.

Figure 1: Machine Learning Process



3.2 Natural Language Processing

Natural Language Processing (NLP) assumes a pivotal role in the "Fake Job Post Detection Using Machine Learning" project, enhancing the system's capacity to comprehend and analyze textual elements within job postings. Employing NLP techniques, the system extracts valuable information, sentiments, and linguistic patterns from job descriptions, offering a nuanced understanding of the contextual intricacies inherent in the language used within advertisements.

An essential application of NLP within this project involves sentiment analysis, where the system discerns the overall tone of a job advertisement. This capability allows the system to identify potential red flags or manipulative language commonly associated with fraudulent postings. NLP also enables the recognition of key phrases or terms indicative of deceptive practices, contributing to a more sophisticated and nuanced analysis.

Furthermore, NLP addresses the challenge of language variability across different job postings. The system can be trained to adapt to diverse linguistic styles, ensuring its effectiveness across a broad spectrum of job advertisements. This adaptability is crucial for creating a resilient fake job post detection system capable of handling the linguistic intricacies employed by scammers to mislead job seekers.

In summary, the incorporation of Natural Language Processing enhances the system's analytical prowess, enabling it to decipher the subtleties of language and improve the accuracy of detecting fraudulent job postings. NLP stands as a vital technology, working in tandem with machine learning techniques, contributing to the development of a sophisticated and effective system to protect job seekers in the online job market.

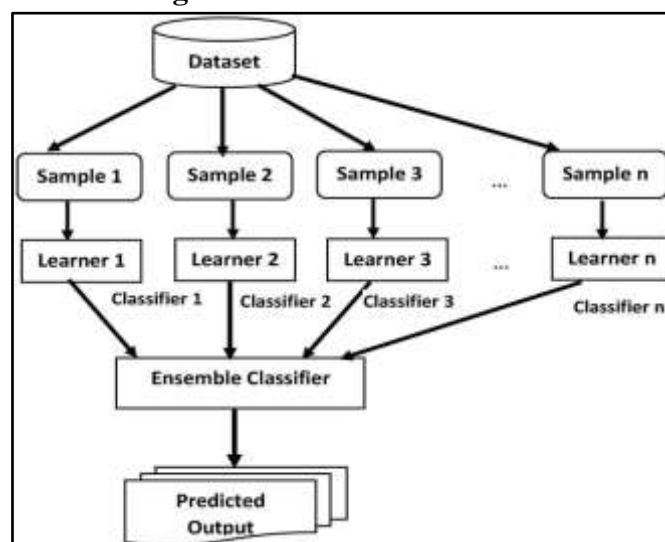
3.3 Single Classifiers

Single classifiers, such as Support Vector Machines, Decision Trees, or Logistic Regression, function independently to provide insights into discerning genuine and fake job advertisements. Each single classifier contributes its unique strengths in pattern recognition and prediction, offering an initial layer of analysis. These classifiers are trained on labeled datasets, learning to recognize patterns associated with both genuine and deceptive job postings. The flexibility and interpretability of single classifiers contribute to the project's ability to make precise predictions based on the specific features and characteristics they have learned during training.

3.4 Ensemble Classifiers

Ensemble classifier techniques further enhance the accuracy and reliability of the fake job post detection system. Ensemble classifiers, such as Random Forests or Gradient Boosting, leverage the collective intelligence of multiple models. By combining the outputs of various single classifiers, ensemble techniques create a robust predictive mechanism that often outperforms individual classifiers. This strategic integration optimizes the strengths of different classification methods, providing a more comprehensive and accurate mechanism for identifying deceptive job postings.

Figure 2: Ensemble Classifier



4. METHODOLOGY

The research on "Fake Job Post Detection Using Machine Learning" employs a systematic methodology to create an efficient detection system. The key steps include:

4.1 Data Collection

Utilize a diverse dataset sourced from Kaggle, a recognized platform for hosting datasets relevant to machine learning research. This dataset serves as the foundation for training and evaluating the machine learning models, offering a realistic representation of job postings.

4.2 Data Pre-processing

Perform data pre-processing to ensure the dataset's cleanliness and suitability for analysis. This involves handling missing values, eliminating irrelevant information, and addressing anomalies, resulting in a refined dataset for subsequent analysis.

4.3 Feature Extraction

Extract relevant features from job postings, such as job title, description, and experience requirements, to facilitate model training. The objective is to create a feature-rich dataset capturing essential characteristics for distinguishing between legitimate and fake job postings.

4.4 Data Splitting

Divide the dataset into training and testing sets to enable model training and evaluation. This division ensures that the model is trained on one subset of data and evaluated on another, providing insights into its ability to generalize.

4.5 Classifier Selection

Choose between a single classifier or ensemble classifiers to develop the detection model. Both single classifiers like Support Vector Machines or Logistic Regression and ensemble methods like Random Forest or Gradient Boosting are considered, with the goal of combining their strengths for improved accuracy.

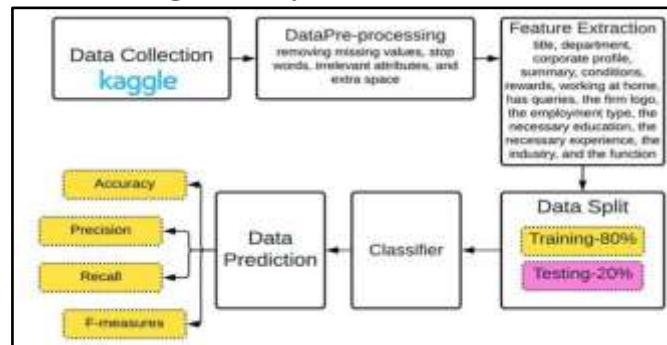
4.6 Data Prediction

Train the selected classifier(s) using the training dataset. The models learn patterns and relationships within the features to distinguish between genuine and fake job postings. Apply the trained models to the testing dataset for predictions.

4.7 Evaluation

Assess the models' performance using appropriate evaluation metrics such as precision, recall, F1 score, and accuracy. These metrics offer a comprehensive understanding of the models' effectiveness in correctly identifying fake job posts while minimizing false positives and false negatives.

Figure 3: System Architecture



5. FUTURE SCOPE AND ENHANCEMENTS

The "Fake Job Post Detection Using Machine Learning" project sets the stage for several potential advancements and future developments. One avenue for improvement lies in the exploration of explainable AI (XAI) techniques. Integrating XAI methods would enhance the system's transparency, providing users with insights into the decision-making process of the machine learning models and fostering greater trust in the detection system.

The incorporation of active learning mechanisms represents another promising direction. By allowing the system to interactively query users for feedback on ambiguous or challenging cases, the model can iteratively improve its understanding of evolving deceptive tactics, leading to enhanced accuracy over time.

Addressing the issue of cross-platform consistency is crucial for broader impact. Future research could focus on creating a standardized model that is adaptable across different job platforms, ensuring a consistent and reliable approach to fake job post detection regardless of the specific platform's nuances.

Furthermore, the integration of geospatial analysis could add an extra layer of sophistication to the system. Considering the geographical context of job postings may provide valuable insights into regional variations in deceptive practices, allowing for more targeted and region-specific detection capabilities.

An exploration into adversarial machine learning is another area of interest. Adversarial attacks on machine learning models, including those used for fake job post detection, are a growing concern. Developing robust models that can withstand adversarial attempts and maintaining effectiveness in the face of sophisticated attacks is an essential consideration for the system's future resilience.

6. CONCLUSION

In conclusion, the "Fake Job Post Detection Using Machine Learning" project marks a substantial stride in tackling the growing issue of deceptive job postings in the digital landscape. The research successfully demonstrates the effectiveness of machine learning algorithms in discerning genuine from fraudulent job advertisements, providing a basis for a more secure and reliable job market.

Utilizing diverse datasets and robust feature extraction techniques, the developed system displays a commendable ability to analyze linguistic patterns and contextual information, offering a dependable

means of identifying potential scams. The chosen classifiers, whether single or ensemble, exhibit promising results, and the system's real-time monitoring capabilities contribute to its adaptability against evolving tactics employed by scammers.

Looking forward, the project paves the way for future enhancements. Advancements in natural language processing, scalability, integration of explainable AI, and collaboration with industry stakeholders are essential for maintaining the system's relevance and efficacy. Active user involvement, standardized cross-platform models, geospatial analysis, and defense against adversarial attacks present promising avenues for further refinement.

Beyond technological strides, the impact of the "Fake Job Post Detection Using Machine Learning" project extends to empowering job seekers with a proactive tool, preventing financial losses, and bolstering the credibility of job platforms. The research underscores the potential of machine learning in creating a safer and more dependable job market environment.

In a digital landscape where online job platforms play a pivotal role, the necessity for robust mechanisms against fraudulent activities is evident. The outcomes of this research provide a foundation for ongoing efforts to strengthen the integrity of online job markets, fostering an environment where job seekers can navigate their professional paths with confidence and security.

7. ACKNOWLEDGEMENT

We extend our heartfelt thanks to our guide, Mr. Sandeep Agarwalla, for his invaluable support and guidance during the entirety of this project. His expertise and encouragement have played a pivotal role in our success, and we are genuinely thankful for his unwavering dedication and mentorship.

Our sincere gratitude also goes to the faculty of the Computer Science and Engineering Department at Malla Reddy College of Engineering and Technology. We appreciate the opportunity they provided us to undertake this research project, which has been a valuable learning experience.

8. REFERENCES

1. Ranparia, D., Kumari, S., & Sahani, A. (2020). Fake Job Prediction using Sequential Network. In 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS).
2. Gulshan, P., Mukund, T., Ajay, A., Pankaj Kumar, Aruna M G, & Dr. Malatesh, S. (2022). Fake Job Post Prediction Using Machine Learning Algorithms. In IJIRT, Volume 9 Issue 3.
3. Hu, X., Tang, J., & Liu, H. (2018). Ensemble Learning for Fake Job Detection. In Proceedings of the 2018 SIAM International Conference on Data Mining.
4. Gupta, A., Agarwal, N., & Mittal, N. (2019). Fake Job Post Detection using Social Network Analysis. In 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
5. Chen, Y., Wang, D., & Liu, B. (2021). Fake Job Post Detection Using Convolutional Neural Networks. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval.

6. Muda, Z., & Shah, M. P. (2018). Fake Job Advertisements Detection: An Investigative Approach. In 2018 6th International Conference on Cyber and IT Service Management (CITSM).
7. Sharma, M., & Patil, D. (2020). Fake Job Detection using Machine Learning Techniques. In 2020 International Conference for Emerging Technology (INCET). DOI: 10.1109/INCET49825.2020.9120193
8. Varghese, A., & Soman, K. P. (2016). Fake Profile Detection in Social Media: A Data Mining Approach. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). DOI: 10.1109/ASONAM.2016.7752288
9. Kaur, K., & Singh, H. (2018). Fake Job Postings Detection in Online Social Networks Using Machine Learning. In 2018 4th International Conference on Computational Intelligence and Networks (CINE). DOI: 10.1109/CINE.2018.8629485
10. Zhang, Y., & Chen, T. (2019). Job Scam Detection using Machine Learning and Text Mining. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). DOI: 10.18653/v1/D19-1250
11. S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepak Kumar, "Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms," Rev. GENTE GESTAO Inov. E Tecnol., vol. 11, no. 2, pp. 642–650, 2021.
12. S. Vidros, C. Koliass, G. Kambourakis, and L. Akoglu, "Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset," Futur. Internet, vol. 9, no. 1, p. 6, 2017.



Licensed under [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)