DOI:

Unmasking Fraudulent Job Ads: A Critical Review of Machine Learning Techniques for Detecting Fake Jobs

Mr. Dipjyoti Deka: Assistant Professor – Dept. of Computer Science and Engineering – Assam down town University, Guwahati, India

Rituparna Seal: Dept. of Computer Science and Engineering, Assam down town University, Guwahati, India

Shubham Banik: Dept. of Computer Science and Engineering, Assam down town University, Guwahati, India

Abstract

In today's technology-driven world, where everything is just a few clicks away, online job postings have also increased clearly, allowing job seekers to apply for jobs via various online job portals. While this has made job hunting easier, the rise of fraudulent job advertisements has also augmented tremendously. Fraudulent job advertisements are created to deceive job seekers by extracting their personal information for several malicious purposes or monetary gain. It has become the need of the hour to protect job seekers from potential financial and identity theft by detecting these fraudulent job advertisements. This paper focuses on reviewing some recent research on the detection of fraudulent job advertisements using machine learning approaches. In this paper, seven research papers were analyzed, focusing on the datasets, feature engineering techniques, machine learning algorithms, and evaluation metrics used to detect fraudulent job advertisements. The paper concludes by highlighting the current challenges and future directions for research in this area.

Keywords

Machine learning, Fraudulent job advertisements, Fake job advertisements, Natural Language Processing, Logic Regression, Random Forest, Naive Bayes, Decision Tree, Gradient Boosting, Support Vector Classifier, Feature Engineering, Feature Extraction

I. INTRODUCTION

The rise in online job platforms is evident these days. It has become extremely convenient for job seekers to hunt for jobs and apply for job opportunities at their own pace of time, anywhere globally. Nonetheless, the availability of innumerable job postings has also led to an increase in fraudulent job advertisements. Fraudulent job advertisements are generally created to deceive job seekers and extract their personal information or money for personal gain. This not only is a huge crisis for job seekers but also the legitimate organizations that offer genuine opportunities to candidates, as the fraudsters try to disguise themselves as a legitimate organization and seek unethical particulars on their name, which challenges the reputation of the genuine organizations.

A job hunt is a significant part of every individual and their career, and job recruitment is a crucial process for any organization or business. Therefore, it is essential to ensure that the job advertisement that is posted online is genuine and not fake. Fake job advertisements can cause some serious major issues for both job seekers and employers. While job postings online are increasing at a tremendous pace, it is becoming more and more complicated to differentiate between genuine and fake job advertisements. Job scams are a mounting problem in today's technology-driven world. With the augment of remote work and online job portals, almost all individuals are turning to the Internet to find employment opportunities. Unfortunately, this has also led to an enhancement in job scams, where fraudsters craft fake job postings to lure unsuspecting victims into giving away personal information or paying a huge sum for non-existent jobs.

The dilemma with such scams is that they can be tricky to detect since they pretend to be from a valid source. Although various job portals previously warn job seekers regarding some common pointers of fake job postings and the masterminds behind them, the

fraudsters are becoming gradually more refined in their tactics, making it convoluted for job seekers to make a distinction between genuine job postings and fake ones. This can lead to individuals wasting both time and money on deceitful job applications, as well as putting their personal information in jeopardy. There are also evident cases of Cybercrimes related to the same scenario where job seekers are asked to fill in particularly sensitive information. As soon as the fraudsters acquire the extremely sensitive information they first request the job seekers for a certain amount which if denied, they tend to threaten the individuals claiming that if the respective amount is not paid they would misuse their information in unethical and illegal deeds.

In recent years, some researchers have been carried out using machine learning approaches to detect fraudulent job advertisements. By using a machine learning approach fake jobs can be detected, by automating the process. Various techniques and algorithms in machine learning can give high accuracy and the detection technique can be improved. Various machine learning algorithms have been used to detect fraudulent job postings, with various researchers proposing different approaches to tackle the problem.

II. RELATED WORKS

In this review paper, seven research papers are being analyzed that have proposed machine learning algorithms to detect fake job advertisements. These papers include "Detection of Fake Job Advertisements Using Machine Learning Algorithms" by Baraneetharan a machine learning-based approach was proposed to detect fake job advertisements using text and image features. [1]

"A Smart Secured Framework for Detecting and Averting Online Recruitment Fraud Using Ensemble Machine Learning Techniques" by Ullah where a smart secured framework for detecting and averting online recruitment fraud using ensemble machine learning techniques was proposed. [2]

"A machine learning approach to detecting fraudulent job types" by Naudé, Adebayo, and Nanda proposed a machine learning approach to detecting fraudulent job types. [3]

"Dealing with the class imbalance problem in the detection of fake job descriptions" by Vo and Sharma addressed the class imbalance problem in the detection of fake job advertisements. [4]

"Detection of fake job postings by utilizing machine learning and natural language processing approaches" by Amaar et al. proposed a machine learning and natural language processing-based approach to detecting fake job postings. [5]

"Classifying Fake and real job advertisements using Machine Learning" by Kumar proposed a machine learning-based approach to classifying fake and real job advertisements. [6]

"Fake Job Recruitment Detection Using Machine Learning Approach" by Dutta and Bandyopadhyay proposed a machine learning approach for detecting fake job recruitment by using some algorithms. [7]

III. DATASETS

The research papers that we reviewed made use of different datasets for detecting fraudulent job advertisements.

Baraneetharan (2022) used a dataset of 45,465 job advertisements, which consisted of 19,867 fake and 25,598 real job advertisements. [1]

Ullah (2023) used a dataset of 10,000 job advertisements, which consisted of 5,000 real and 5,000 fake job advertisements. [2]

Naude et al. (2022) used a dataset of 1,027 job advertisements, which consisted of 315 fraudulent job advertisements and 712 legitimate job advertisements. [3]

Vo and Sharma (2021) used a dataset of 16,660 job advertisements, which consisted of 4,902 fake and 11,758 real job advertisements. [4]

Amaar et al. (2022) used two datasets for their study on detecting fake job postings. The first dataset contains 17,880 job postings, out of which 866 are labeled as fake job postings. The second dataset contains 11,000 job postings, collected from various online job portals and social media platforms. Out of these 11,000 postings, 7,828 are genuine job postings, and 3,172 are labeled as fake job postings. [5]

Kumar (2020) used a dataset of 17,880 real-life job postings of which 17,014 are real and 866 are fake. The dataset was further processed and uploaded on Kaggle and is available publicly. [6]

Dutta and Bandyopadhyay (2020) used a dataset consisting of 17,880 job postings, out of which 866 job postings were identified as fake. [7]

The authors collected the dataset from various job portals and classified the job postings into two categories: fake and genuine. They used this dataset

to train and evaluate their proposed Machine learning approach for fake job recruitment detection.

IV. ANALYSIS OF TECHNIQUES

All the seven research papers that we selected for this review paper proposed some machine learning algorithms to detect fake job advertisements. All the papers were unique in their way and used different approaches to detect fraudulent job advertisements. Some of the papers made an effort on detecting fake job advertisements using feature extraction and feature engineering technique which included the features such as job titles, job descriptions, company names, and various other features.

Some papers also used Natural Language Processing techniques to analyze the job posting text and detect if it is a genuine or fake job posting. Furthermore, some of the papers used supervised learning algorithms and machine learning models such as decision trees, random forests, or logistic regression. While others used unsupervised learning algorithms like k-means clustering.

Baraneetharan (2022) who used a dataset of 45,465 job advertisements, in the paper proposed a machine learning-based approach to detect fake job advertisements using the Feature engineering technique. In the research, the features that were used to achieve the goal were job titles, company names, and job descriptions. The techniques that were used were decision trees, random forest, and logistic regression algorithms to train and test the model. While all the techniques that were used gave a satisfactory result it was found that the best performance of all three was obtained with the decision tree algorithm. The decision tree algorithm gave an accuracy of 96.2% which even if is not the highest among the other researchers, is quite satisfactory and gave pretty good results. [1]

Ullah (2023) proposed a smart and secure framework for detecting and averting online recruitment fraud in the paper. In the research paper ensemble Machine learning techniques were used. Feature extraction techniques were also used in this paper, and the author used various job features such as job titles, job descriptions, and company names to train the model. To improve the performance of the model, the author also made use of techniques like bagging and boosting. The proposed model achieved an accuracy of 92.5%. Considering the less amount of datasets that were taken which is only 10,000, the results were agreeable. [2]

Naude, Adebayo, and Nanda (2022), in their research, proposed a Machine learning approach to detect fraudulent job types. In their research work, the authors tried to use and implement unsupervised learning techniques such as k-means clustering to cluster job advertisements and detect fraudulent jobs. It is the method by which the unlabeled dataset is grouped into different clusters based on some rules. The model that was proposed by the authors had an accuracy of 89%, which is the lowest compared to the other researchers. However, the technique that has been proposed is highly efficient and if used on a higher number of datasets can potentially achieve a higher accuracy rate. [3]

Vo and Sharma (2021) proposed a method to deal with the class imbalance problem in the detection of fake job descriptions particularly. They proposed a technique called "SMOTE" (Synthetic Minority Over-sampling Technique) to overcome this problem. In this technique, the authors used a combination of oversampling and undersampling to balance the dataset and improve the performance of the models. Feature extraction techniques were also used in the proposed model by the authors to train the model. The proposed approach achieved an overall accuracy of 95% in detecting fake job descriptions, using the Random Forest algorithm. The precision, recall, and F1-score were also high for both the minority class (fake job descriptions) and the majority class (genuine job descriptions). [4]

Amaar et al. (2022) proposed a machine learning and natural language processing (NLP) approach to detect fake job postings. The authors performed feature engineering, data cleaning, and data preprocessing before applying machine learning algorithms. The authors used two datasets and both datasets were preprocessed and transformed into numerical features before being used to train and evaluate the machine learning models. The authors used several machine learning algorithms, including Naive Bayes, Decision Tree, Random Forest, and Gradient Boosting, to evaluate the performance of their proposed approaches. On the first dataset, which contains 17,880 job postings, out of which 866 were labeled as fake, the proposed approach achieved an accuracy of 96.1% using the Random Forest classifier. On the second dataset, which contains 11,000 job postings with 7,828 labeled as genuine and 3,172 labeled as fake, the proposed approach achieved an accuracy of 97.8% using the Gradient Boosting classifier. [5]

In the study by Kumar (2020), the author used machine learning algorithms to classify fake and real job advertisements. The author applied different machine learning algorithms, such as Vector

Classifier, decision tree, and random forest, and evaluated the performance of each algorithm. The Support Vector Classifier gave an accuracy score of 97.78% for 5354 test observations. It is also noteworthy that it was able to correctly predict the class labels for 5245 job postings. [6]

The proposed approach in the paper by Dutta and Bandyopadhyay (2020), used several machine learning algorithms, including Naive Bayes, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). The paper reported the accuracy of each algorithm on the testing set, among which experimental results indicate that the Random Forest classifier outperforms its peer classification tool. The proposed approach achieved an accuracy of 98.27% which is much higher than the existing methods. [7]

V. CONCLUSION

Last but not the least, the studies reviewed in this paper demonstrates the potential of Machine learning algorithms in detecting fake job advertisements. However, it is difficult to address which proposed method is best since all the methods that were proposed by the researchers used different datasets, methods, and technologies. In terms of the best-performing model across the reviewed papers, several algorithms were commonly used, such as logistic regression, decision trees, random forests, and support vector machines.

Nevertheless, some studies achieved better accuracy by using ensemble learning techniques, such as bagging and boosting, or by incorporating NLP techniques. Fake job postings are a major problem that affects job seekers and also the market for job advertisements. Our review paper discussed various Machine learning approaches that can be used to detect fake job postings. It has been observed that the use of a combination of techniques gives better accuracy and results than using a single approach. Conversely, it is important to note that the classifiers are heavily dependent on the quality and the size of the dataset, and also the choice of machine learning algorithms.

VI. FUTURE SCOPE

The research in this field is still evolving, and more research is needed to develop robust models that can detect fake job postings with high accuracy and can handle noisy and incomplete data. It is also the need of the hour to explore other data sources, such as social media, which has become a hub for fraudulent job advertisements these days. This will improve the accuracy of fraudulent job detection models. Social media platforms are also highly significant in providing additional information about the job poster, such as employment history, or other credentials, which can be utilized to validate the authenticity of job postings. Fraudsters are becoming more cunning with the passing days and have developed strategic ways to impersonate a genuine job posting or lure victims in phishing scams. There is still room for a lot of development, as new types of fraud are emerging.

In near future, we aim to perform research and develop a model using machine learning that addresses all the mentioned problems in this industry. We will be evaluating some machine learning algorithms and try to develop a model with the algorithm that gives the best performance. We might also use a combination of different algorithms to achieve our goal.

References

[1] Baraneetharan, None E. "[PDF] Detection of Fake Job Advertisements Using Machine Learning Algorithms by None E. Baraneetharan · 10.36548/Jaicn.2022.3.006 · OA.Mg." OA.Mg · Open Access for Everyone · Download and Read over 240 Million Research Papers, 2022, https://oa.mg/work/10.36548/jaicn.2022.3.006.

[2] Ullah1, Zahid. "A Smart Secured Framework for Detecting and Averting Online Recruitment Fraud Using Ensemble Machine Learning Techniques [PeerJ]." PeerJ Computer Science, 8 Feb. 2023, https://peerj.com/articles/cs-1234/.

[3] Naudé, M., Adebayo, K.J. and Nanda, R. (2022) A machine learning approach to detecting fraudulent job types - ai & society, SpringerLink. Springer London. Available at: https://link.springer.com/article/10.1007/s00146-02201469-0

[4] Vo, A. and Sharma, R. (2021) Dealing with the class imbalance problem in the detection of fake job ... Available at: https://www.researchgate.net/publication/350399453_Dealing_with_the_Class_Imbalance_Problem_in_the_Detection_of_Fake_Job_Descriptions

[5] Amaar, A. et al. (2022) Detection of fake job postings by utilizing machine learning and ... Available at: https://www.researchgate.net/publication/357576631_Detection_of_Fake_Job_Postings_by_Utilizing_Machine_Learning_and_Natural_Language_Processing_Approaches

[6] Kumar, D.V. (2020) Classifying fake and real job advertisements using machine learning, Analytics India Magazine. Available at: https://analyticsindiamag.com/classifying-fake-and-real-job-advertisements-using-machine-learning/

[7] Shawni Dutta, Prof.Samir Kumar Bandyopadhyay "Fake Job Recruitment Detection Using Machine Learning Approach" International Journal of Engineering Trends and Technology 68.4(2020):48-53