Guarding Against Job Scams : Harnessing Decision Tree Analysis

Mr. G.Sravan Kumar[1],T.Pranitha[2], S.Sai Sridhar Naidu[3], M.Madhava Swamy[4]
Scholar[2,3,4], Associate Professor[4], Assistant Professor1
Department of Computer Science and Engineering,
Nalla Narasimha Reddy Education Society's Group of Institutions,Hyderabad, India

Abstract—The rise of online job portals has made job hunting easier but has also led to an increase in fake job postings. These scams can trick job seekers into losing money or sharing personal. This study aims to create a system that can predict and identify fake job postings using data mining techniques. We analysed a large number of job advertisements using various data mining methods to find patterns that indicate whether a job posting is real or fake. We used text mining to extract important features from the job descriptions, like common words, sentence structures, and writing styles. Natural language processing (NLP) helped us understand the context and meaning behind the text. We then applied machine learning algorithms, such as decision trees, naïve bayes, and neural networks, to classify the job postings as real or fake based on these features. We found that specific words, unusual job titles, and how contact information is presented are key indicators of fake job postings. Our study highlights the effectiveness of data mining in protecting job seekers from online scams. The model we developed can be used by job portals to automatically detect and flag suspicious job postings, making job hunting safer for everyone. Future work will focus on using more data, improving the way we extract features, and testing advanced methods to make our predictions even more accurate. This research helps in the fight against online fraud in the job market, providing a practical solution to protect job seekers from scams.

Keywords—Natural Language Processing(NLP), Decision tree, Fraud detection, Job Portals, Text Mining.

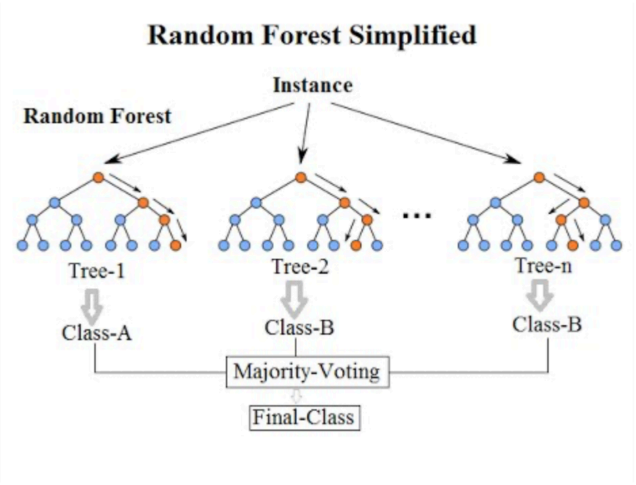
1.INTRODUCTION

In modern time, the development in the field of industry and technology has opened a huge opportunity for new diverse jobs for the job seekers. With the help of the advertisements of these job offers, job seekers find out their options depending on their time, qualification,experience, suitability etc. Recruitment process is now influenced by the power of internet and social media. Since the successful completion of a recruitment process is dependent on its advertisement, the impact of social media over this is tremendous. Social media and advertisements in electronic media have created newer and newer opportunity to share job details. Instead of this, rapid growth of opportunity to share job posts has increased the percentage of fraud job postings which causes harassment to the job seekers. So, people lacks in showing interest to new job postings due to preserve security and consistency of their personal, academic and professional information. Thus the true motive of valid job postings through social and electronic media faces an extremely hard challenge to attain people's belief and reliabilities make more reliability.Technologies are around us to make our life easy and developed but not to create unsecured environment for professional life. If jobs posts can be filtered properly predicting false job posts, this will be a great advancement for recruiting new employees. . Fake job posts create inconsistency for the job seeker to find their preferable jobs causing a huge waste of their time. An automated system to predict false job post opens a new window to face difficulties in the field of Human Resource Management.

2.RELATED RESEARCH

1. Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset

The critical process of hiring has relatively recently been ported to the cloud. Specifically, the automated systems responsible for completing the recruitment of new employees in an online fashion,aim to make the hiring process more immediate, accurate and cost-efficient.However, the online exposure of such traditional business procedures has introduced new points of failure that may lead to privacy loss for applicants and harm the reputation of organisations. So far, the most common case of Online Recruitment Frauds (ORF), is employment scam. Unlike relevant online fraud problems, the tackling of ORF has not yet received the proper attention, remaining largely unexplored until now.



2. Spotting fake reviews via collective positive-unlabeled learning

Online reviews have become an increasingly important resource for decision making and product designing.But reviews systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, ground truth of large scale datasets is still unavailable and most of existing approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. Working with Dianping1 , the largest Chinese review hosting site, we present the first reported work on fake review detection in Chinese with filtered reviews from Dianping's fake review detection system. Dianping's algorithm has a very high precision, but the recall is hard to know. This means that all fake reviews detected by the system are almost certainly fake but the remaining reviews (unknown set) may not be all genuine.

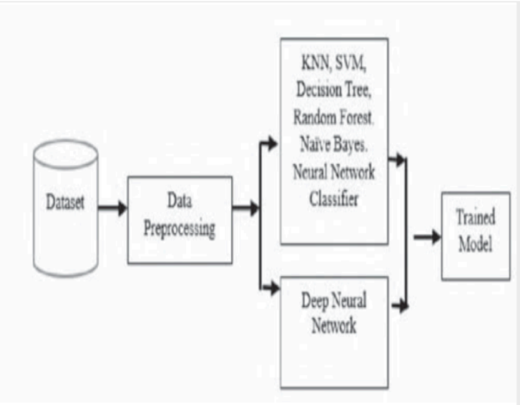

Dataset
Data Preprocessing
KNN, SVM, Decision Tree, Random Forest, Naive Bayes, Neural Network Classifier
Deep Neural Network
Trained Model

Fig: System Architecture

3. An Intelligent Model for Online Recruitment Fraud Detection

This study research attempts to prohibit privacy and loss of money for individuals and organisation by creating a reliable model which can detect the fraud exposure in the online recruitment environments.This research presents a major contribution represented in a reliable detection model using ensemble approach based on Random forest classifier to detect Online Recruitment Fraud (ORF). The detection of Online Recruitment Fraud is characterised by other types of electronic fraud detection by its modern and the scarcity of studies on this concept. The researcher proposed the detection model to achieve the objectives of this study. For feature selection, support vector machine method is used and for classification and detection, ensemble classifier using Random Forest is employed.

As more individuals rely on these platforms to seek employment opportunities, scammers have increasingly exploited this by posting fake jobs, leading to financial losses, identity theft, and privacy breaches for unsuspecting job seekers. Traditional methods, which often rely on manual moderation or rule-based systems, are inefficient and cannot scale to the large number of job postings. This paper proposes an intelligent, automated solution using machine learning techniques to detect fraudulent job postings. By analysing various features such as job descriptions, titles, and other metadata, the model aims to classify postings as either legitimate or fraudulent, offering a robust defense against evolving scam tactics. The goal is to safeguard both job seekers and the credibility of online recruitment platforms.

3. METHODOLOGY

3.1. AIM OF THE PROJECT

The main Aim of this project is to detect fake job prediction. Blacklisted domain detector to label post containing blacklisted URLs.

3.2. PROPOSED SYSTEM

In the proposed system, the system proposes a semi-supervised framework for fake posts detection. The proposed framework mainly consists of two main modules: 1) four lightweight detectors in the fake posts detection module for detecting fake posts in real time and 2)updating module to periodically update the detection models based on the confidently labeled posts from the previous time window. The detectors are designed based on our observations made from a collection of 14 million posts, and the detectors are computationally effective, suitable for real-time detection.More importantly, our detectors utilise classification techniques at two levels, posts level and cluster level. Here, a cluster is a group of posts with similar characteristics. With this flexible design, any features that may be effective in fake detection can be easily incorporated into the detection framework. The framework starts with a small set of labeled samples and updates the detection models in a semi-supervised manner by utilising the confidently labeled posts from the previous time window

3.3. ADVANTAGES OF PROPOSED SYSTEM

Confidently Labeled Posts : Posts that are labeled by the first three detectors (i.e., blacklisted domain, near duplicate and reliable ham posts) are considered as confidently labeled posts.

4.RESULTS

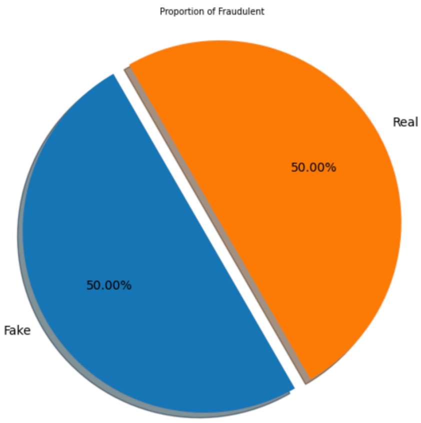


Fig : Pie Chart

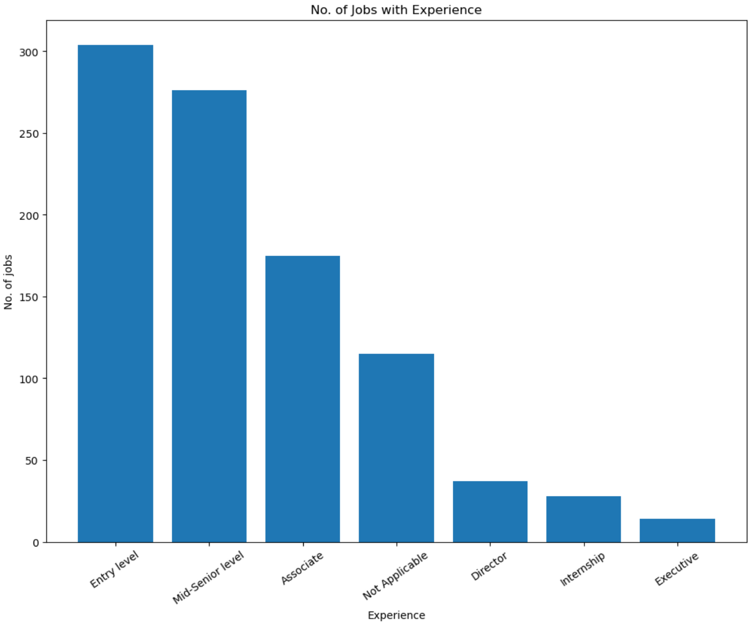


Fig : Frequency Graph

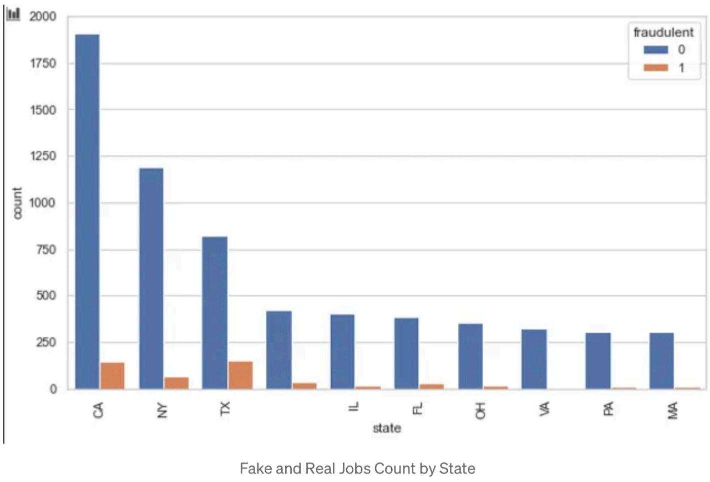


Fig : Bar Graph

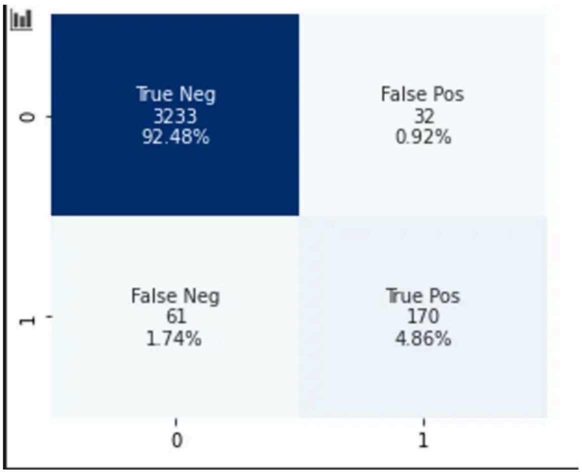


Fig : Visualisation

5.CONCLUSION

Job scam detection has become a great concern all over the world at present,We have analysed the impacts of job scam which can be a very prosperous area in research filed creating a lot of challenges to detect fraudulent job posts. I will provide a characterisation of the users of this labeled collection, bringing to the light several attributes useful to differentiate fake job and non-fake job. We leverage our characterisation study towards a spammer detection mechanism. Using a classification technique, we were able to correctly identify a significant fraction of the fake job while incurring in a negligible fraction of misclassification of legitimate users. I also investigate different tradeoffs for our classification approach and the impact of different attribute sets. Our results show that even with different subsets of attributes, our approach is able to detect fake job with high accuracy.

6. REFERENCES

[1] S. Vidros, C. Kolilias , G. Kambourakis ,and L. Akoglu, “Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset”, Future Internet 2017, 9, 6; doi:10.3390/fi9010006.

[2] B. Alghamdi, F. Alharby, “An Intelligent Model for Online Recruitment Fraud Detection”, Journal of Information Security, 2019, Vol 10, pp. 155 176, https://doi.org/10.4236/iis.2019.103009 .

[3] Tin Van Huynh1, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen1, and Anh Gia- Tuan Nguyen, “Job Prediction: From Deep Neural Network Models to Applications”, RIVF International Conference on Computing and Communication Technologies (RIVF), 2020.

[4] Jiawei Zhang, Bowen Dong, Philip S.Yu, “FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network”, IEEE 36th International Conference on Data Engineering (ICDE),2020.

[5] P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao, “Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification,”Neurocomputing, vol. 174, pp. 806 814,2016.

[6] C. Li, G. Zhan, and Z. Li, “News Text Classification Based on Improved BiLSTM-CNN,” in 2018 9th International Conference on Information Technology in Medicine and Education (ITME), 2018, pp. 890-893.

[7] Yasin, A. and Abuhasan, A. (2016) An Intelligent Classification Model for Phishing Email Detection. International Journal of Network Security& Its Applications, 8, 55-72. https://doi.org/10.5121/imsa.2016.8405

[8] Vong Anh Ho, Duong Huynh-Cong Nguyen, Danh Hoang Nguyen, Linh Thi- Van Pham, Duc-Vu Nguyen, Kiet Van Nguyen, and Ngan LuuThuy Nguyen."Emotion Recognition for Vietnamese Social Media Text", arXivPrepr. arXiv:1911.09339, 2019.

[9] Li, H.; Chen, Z.; Liu, B.; Wei, X.; Shao, J. Spotting fake reviews via collective positive-unlabeled learning. In Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM), Shenzhen, China, 14-17

[10] Ott, M.; Cardie, C.; Hancock, J. Estimating the prevalence of deception in online review communities. InProceedings of the 21st international conference on World Wide Web, Lyon, France, 16-20 April 2012; ACM: New York, NY, USA,2012; pp. 201-210.

[11] Thin Van Dang, Vu Duc Nguyen, Kiet Van Nguyen and Ngan LuuThuy Nguyen,“Deep learning for aspect detection on vietnamese reviews” in In Proceeding of the 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), 2018, pp. 104-109.