

International Journal of
Engineering Research and Science & Technology



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

A COMPARITIVE STUDY ON FAKE JOB POST PREDICTION USING DIFFERENT DATAMINING TECHNIQUES

¹DR.B. GOHIN, ²GUTTULA VENKATA SATYASREE

¹(Associate Professor), MCA, Swarnandhra College

²MCA, scholar, Swarnandhra College

ABSTRACT

The proliferation of online job boards and other forms of mass communication has made posting openings for new positions a routine occurrence in today's society. Therefore, everyone should be quite worried about the problem of predicting bogus job postings. False job posing prediction has many of the same difficulties as other categorization problems. A variety of data mining and classification algorithms, including KNN, decision trees, support vector machines, naive bayes, random forest, multilayer perceptrons, and deep neural networks, are suggested in this study as ways to determine the authenticity of a job posting. We conducted experiments using the 18000-sample Employment Scam Aegean Dataset

(EMSCAD). This classification challenge is well-suited to deep neural networks as classifiers. This classifier for deep neural networks is built with three thick layers. When it comes to predicting a fake job posting, the trained classifier demonstrates about 98% DNN classification accuracy.

1.INTRODUCTION

The expansion of contemporary industry and technology has provided job-seekers with a plethora of new and varied employment opportunities. employment searchers may narrow their selections based on time, qualifications, experience, appropriateness, etc., by perusing the ads for these employment offerings. The internet and social media have become powerful tools in the recruitment process. The influence of

social media on recruiting processes is substantial since these processes rely on advertisements. There are more and more ways to disseminate information about available jobs thanks to social media and online ads. Rather, the number of fraudulent job ads has grown due to the fast expansion of opportunities to disseminate job advertisements, which in turn creates harassment to those seeking employment. Since individuals value the privacy and integrity of their academic, professional, and personal information, they are hesitant to respond to newly posted job openings. So, getting consumers to trust and believe legitimate job posts on social and electronic media is a huge task. Technology is all around us, and it should be used to improve our lives and make them easier, not to make our professional lives more vulnerable. A huge step forward in the hiring process would be achieved if it were possible to accurately screen job postings to exclude postings that were not legitimate. The irregularity caused by fake job ads makes it very difficult, if not impossible, for job seekers to discover the employment that they desire. A new threat has emerged in the realm of human resource management: an automated system that can forecast fraudulent job postings.

A. Job Scams: False Job Postings

The term "job scam" describes online job ads that are not genuine and are more interested in stealing applicants' personal and professional information than in actually placing them in suitable positions. Scammers may target persons looking for work in an effort to fraudulently collect money from them. More than 67% of individuals who search for employment online without knowing about bogus job postings or job scams are in serious danger, according to a recent poll by Action Fraud in the UK [2]. Roughly 700,000 people in the UK have reported falling prey to work scams, with some losing more than \$500,000. Nearly 300 percent growth in the last two years was recorded in the UK, according to the research [2]. Since they are often looking for a guaranteed job for which they are prepared to pay more, fraudsters target students and recent graduates. Because fraudsters are always coming up with new methods of job scamming, cybercrime prevention and protection strategies are ineffective

2. LITERATURE SURVEY

One. Online Recruitment Fraud Detection: Features, Techniques, and a

Public Database Moving the crucial recruiting process to the cloud is a new development. In particular, the goal of the automated systems that handle the online recruiting of new workers is to streamline, improve, and shorten the hiring process. Yet, new vulnerabilities brought about by the internet's disclosure of such conventional business processes pose a threat to applicants' privacy and companies' credibility. Employment scams have been the most prevalent kind of online recruitment fraud (ORF) so far. In contrast to other relevant online fraud issues, ORF has been mostly unexplored and has not gotten the attention it deserves.

Fraud detection via collective positive-unlabeled learning For both consumers and businesses, online evaluations have grown in importance as a tool for research and development. The problem is that opinion spamming often targets review systems. While supervised learning has been the subject of much study into the detection of fake reviews, the ground truth of large-scale datasets is still lacking, and the majority of current supervised learning algorithms rely on pseudo-reviews instead than actual fraudulent reviews. We provide the first

published effort on Chinese false review identification using filtered reviews from Dianping's fake review detection system, in collaboration with Dianping1, the biggest Chinese review hosting site.

3. EXISTING SYSTEM

False accounts have been the subject of several research on the topic of fake job detection. Determined content polluters by analyzing and using characteristics derived from user demographics, follower/following social network, post content, and the temporal component of user activity. Detected bogus postings on social networks by using user graphs and posts. The ever-changing nature of false news has prompted the use of online learning to combat the problem. For efficient social fake identification, they have used data from both content and networks to refine their fake detection algorithm. In order to take advantage of genuine social media users, Tan suggested an unsupervised fake detection system. Their research demonstrates how quickly false trends may spread over social media. They were able to identify fraudulent patterns by comparing them to the non-fake patterns of real users in the social network and the user link graph. Using language and

URL similarities to cluster postings, Gao et al. discovered big clusters with bursty posting patterns, which they used to identify bogus. Twitter phishing operations have been identified using an incremental clustering-based method.

DISADVANTAGES OF EXISTING SYSTEM:

No semi-supervised learning is available. You can't sort through job ads by category.

3.1 PROPOSED SYSTEM

To identify false postings, the suggested system suggests a semi-supervised framework. The two primary components of the suggested framework are the following:

- 1) a module for real-time detection of false posts using four lightweight detectors; and
- 2) a module for periodic updates to the detection models using the reliably labelled posts from the preceding time window. Our findings from analyzing 14 million postings informed the design of these detectors, which are computationally efficient and well-suited for detection in real time. Much more crucially, our detectors use cluster level and posts level categorization algorithms. In this context, a cluster is a collection of related postings. Thanks to its adaptable nature, the detection

framework may easily integrate any elements that might improve false detection. The system uses the previously-labeled posts to update the detection models semi-supervisedly, starting with a limited batch of labelled samples.

The Benefits of the Suggested Approach

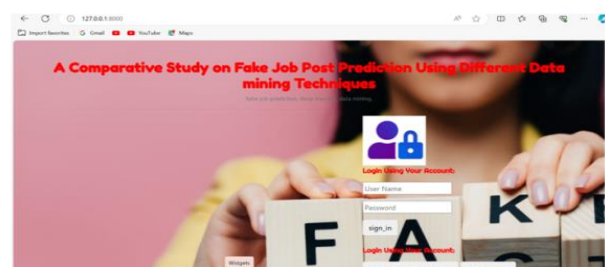
When the first three detectors classify a post as a banned domain, close duplicate, or trustworthy ham, we say the post is confidently tagged.

4. OUTPUT SCREENS

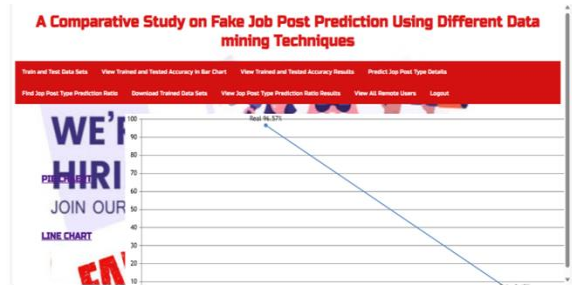
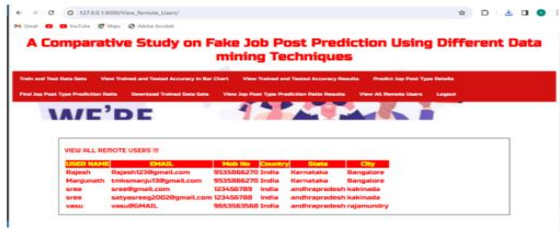
Admin Login:



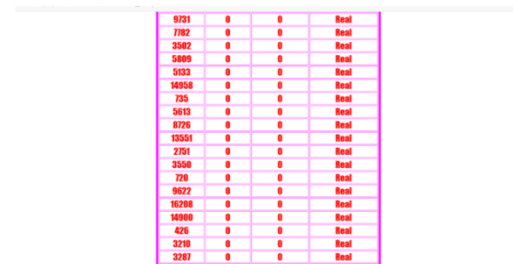
Userlogin:



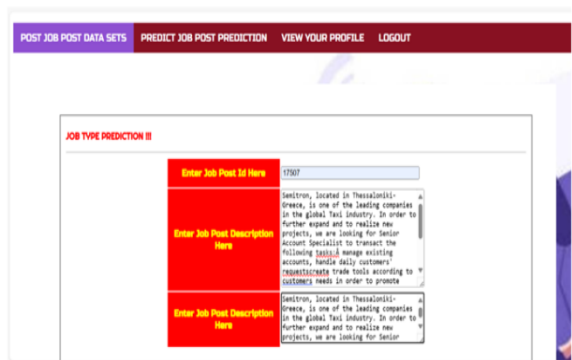
REMOTE USERS



View Jobpost Type Prediction



Job Type Prediction



View Trained And Tested Accuray Results



Prediction Type



5. CONCLUSION

The identification of job scams is becoming a major global problem. We have examined the effects of employment scam in this article. This is a promising topic of study, but it poses significant difficulties in identifying fake job postings. We conducted our experiments using the EMSCAD dataset, which includes actual but doctored job advertisements. In this research, we have tested a deep learning model (Deep Neural Network) in addition to machine learning methods (SVM, KNN, Naïve Bayes, Random Forest, and MLP). A research comparing deep learning-based classifiers with classic machine learning-based ones is presented in this article. Among the more conventional machine learning methods, we discovered that the Random Forest Classifier had the best classification accuracy at 99 percent, while Deep Neural Networks (DNNs) achieved an average of 97.7 percent.

6. REFERENCES

- [1] S. Vidros, C. Koliass, 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. 155176, <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] Scanlon, J.R. and Gerber, M.S., "Automatic Detection of Cyber Recruitment

By Violent Extremitis, Security Informatics, 3, 5, 2014

<https://doi.org/10.1186/s13388-014-0005-5>

[6] Y. Kim, "Convolutional neural networks for sentence classification," arXiv Prepr. arXiv:1408.5882, 2014.

[7] T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A.G.-T. Nguyen, "Hate Speech Detection on Vietnamese Social Media Text using the Bi-GRU-LSTM-CNN Model," arXiv Prepr. arXiv:1911.03644, 2019.

[8] 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. 806814, 2016.

[9] 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.

[10] 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>

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

[12] Thin Van Dang, Vu Duc Nguyen, Kiet Van Nguyen and Ngan Luu-Thuy 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.

[13] Li, H.; Chen, Z.; Liu, B.; Wei, X.; Shao, J. Spotting fake reviews via collective positive-unlabeled learning. In Proceedings of the 2024 IEEE International Conference on Data Mining (ICDM), Shenzhen, China, 14-17 December 2014; pp. 899-904.

[14] Nizamani, S., Memon, N., Glasdam, M. and Nguyen, D.D. (2014) Detection of Fraudulent Emails by Employing Advanced Feature Abundance. *Egyptian Informatics Journal*, Vol. 15, pp. 169-174.