

Fake Job Posting Detection

Ram Prasath S¹ and Dr M N Nachappa²

PG Student, Department of MSc CS-IT¹

Professor, School of CS & IT²

Jain (Deemed-to-be University), Bangalore, India

ram.sivaraj2002@gmail.com¹ and mn.nachappa@jainuniversity.ac.in²

Abstract: *In the realm of online job platforms, the rise of fraudulent job postings poses a significant challenge, undermining the credibility and reliability of these platforms. To address this issue, we propose a machine learning- based solution that leverages the power of Random Forest, Logistic Regression, and Decision Tree classifiers. Through the compilation of a comprehensive dataset containing labeled job postings, we embark on a journey of preprocessing and feature engineering to extract pertinent information from the postings, including textual attributes, geographic details, salary indications, and company profiles. Splitting the dataset into training and testing subsets enables us to meticulously train and evaluate the performance of each classifier, utilizing established metrics such as accuracy, precision, recall, and F1-score to quantify their efficacy in discerning between authentic and fake job listings.*

Our study goes beyond mere model training and evaluation, delving into the intricacies of imbalanced data handling and the practicalities of model deployment and maintenance. By examining the comparative strengths and weaknesses of Random Forest, Logistic Regression, and Decision Tree classifiers, we provide actionable insights for enhancing the integrity of online job platforms through advanced machine learning techniques. With our approach, we aim to not only detect and mitigate the prevalence of fake job postings but also to fortify the trust and credibility of online job- seeking platforms, thereby fostering a more secure and reliable environment for both job seekers and employers alike.

Keywords: NLP (Natural Language Processing), Text classification, Sentiment analysis, Topic modelling , Text pre processing , Word embeddings, Supervised learning, Unsupervised learning, Semi-supervised learning, Deep learning, Convolutional Neural Networks (CNN), Fake news datasets, Real news datasets

REFERENCES

- [1] B. Alghamdi and F. Alharby, —An Intelligent Model for Online Recruitment Fraud Detection,” J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019, doi : 10.4236/jis.2019.103009.
- [2] I. Rish, — An Empirical Study of the Naïve Bayes Classifier An empirical study of the naïve Bayes classifier, || no. January 2001, pp. 41 - 46, 2014.
- [3] D. E. Walters, —Bayes’ s Theorem and the Analysis of Binomial Random Variables, || Biometrical J., vol. 30, no. 7, pp. 817 - 825, 1988, doi: 10.1002/bimj.4710300710.
- [4] F. Murtagh, —Multilayer perceptrons for classification and regression, || Neurocomputing, vol. 2, no. 5 - 6, pp. 183 - 197, 1991, doi: 10.1016/0925-2312(91)90023-5.
- [5] P. Cunningham and S. J. Delany, —K -Nearest Neighbour Classifiers, || Multi . Classif. Syst., no. May, pp. 1 - 17, 2007, doi: 10.1016/S0031-3203(00)00099-6.
- [6] H. Sharma and S. Kumar, —A Survey on Decision Tree Algorithms of Classification in Data Mining, || Int. J. Sci. Res., vol. 5, no. 4, pp. 2094 - 2097, 2016, doi: 10.21275/v5i4.nov162954.
- [7] E. G. Dada, J. S. Bassi, H. Chiroma , S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, “Machine learning for email spam filtering: review, approaches and open research problems, || Heliyon, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.
- [8] L. Breiman, —ST4_Method_Random_Forest, || Mach. Learn., vol. 45, no. 1, pp. 5 - 32, 2001, doi: 10.1017/CBO9781107415324.004.

