

# A Comparative Study on the Prediction of Fake Job Posts using Various Data Mining Techniques

Kesireddy Samara Simha Reddy<sup>1</sup>, Jangiti Dwarakamai<sup>2</sup>, Motha Rahul<sup>3</sup>,  
CH Yogesh Chowdary<sup>4</sup>, Mr. M. Srinivasa Reddy<sup>5</sup>

UG Students, Department of Information Technology<sup>1,2,3,4</sup>

Assistant Professor, Department of Information Technology<sup>5</sup>

Malla Reddy Engineering College, Hyderabad, India

**Abstract:** *In recent years, due to advancement in modern technology and social communication, advertising new job posts has become very common issue in the present world. So, fake job posting prediction task is going to be a great concern for all. Like many other classification tasks, fake job posing prediction leaves a lot of challenges to face. This paper proposed to use different datamining techniques and classification algorithm like KNN, decision tree, support vector machine, naive bayes classifier, random forest classifier, multi-layer perceptron and deep neural network to predict a job post if it is real or fraudulent. We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. Deep neural network as a classifier, performs great for this classification task. We have used three dense layers for this deep neural network classifier. The trained classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job post. Index Terms--false job prediction, deep learning, data mining.*

**Keywords:** Multi-Layer Perceptron, Data Mining, KNN, decision tree, support vector machine, naive bayes classifier, random forest classifier

## 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 naive 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,|| Mult. 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.
- [9]. B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, —Bagging classifiers for fighting poisoning attacks in adversarial classification tasks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6713 LNCS, pp. 350–359, 2011, doi: 10.1007/978-3-642-21557-5\_37.

- [10]. A. Natekin and A. Knoll, —Gradient boosting machines, a tutorial,|| *Front. Neurorobot.*, vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.
- [11]. N.Hussain, H.T.Mirza, G.Rasool, I.Hussain, and M.Kaleem, Spam review detection techniques: A systematic literature review,|| *Appl. Sci.*, vol. 9, no. 5, pp. 1–26, 2019, doi: 10.3390/app9050987.
- [12]. Mandru, D.B., Aruna Safali, M., Raghavendra Sai, N., Sai Chaitanya Kumar, G. (2022). Assessing Deep Neural Network and Shallow for Network Intrusion Detection Systems in Cyber Security. In: Smys, S., Bestak, R., Palanisamy, R., Kotuliak, I. (eds) *Computer Networks and Inventive Communication Technologies . Lecture Notes on Data Engineering and Communications Technologies*, vol 75. Springer, Singapore. [https://doi.org/10.1007/978-981-16-3728-5\\_52](https://doi.org/10.1007/978-981-16-3728-5_52)
- [13]. Mandru, Deena Babu. "An Improved K-Means Algorithm for Web Page Clustering."
- [14]. Krishna, Paruchuri Jeevan, M. Deena Babu, and G. Manoj Someswar. "Design & Development of an improvised PACK System using TRE Technique for Cloud Computing Users." *International Journal of Research* 3.2 (2016): 384-393.
- [15]. Rakesh, Ganya, M. D. Babu, and G. Manoj Someswar. "A Novel Integrated Attestation Graph Analysis Scheme for Enhancing Result Quality and Higher Attacker Pinpointing Accuracy." *International Journal of Research* 3.2 (2016): 214-225.
- [16]. Mandru, Deena Babu, and Y. K. Krishna. "Enhanced Cluster Ensemble Approach Using Multiple Attributes in Unreliable Categorical Data." *International Journal of Psychosocial Rehabilitation* 23.1 (2019).
- [17]. Mandru, Deena Babu, and Y. S. Krishna. "Multi-view Cluster Approach to Explore Multi-Objective Attributes based on Similarity Measure for High Dimensional Data." *International Journal of Applied Engineering Research* 13.15 (2018): 12289-12295.