

# Prediction Of Fake Job Ad Using Nlp-Based Multilayer Perceptron

Vanapalli Jyothi, K.Sri Devi

PG scholar, Department of MCA, DNR college, Bhimavaram, Andhra Pradesh.

(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

## Abstract

Today, technological and industrial developments have created new and diversified career opportunities for job searchers. Based on suitability, experience, qualification, time, etc., job seekers find out their options with the help of the advertisements for these job offers. The recruitment process is now influenced by the power of social media and the internet recruitment process. To share the job details in electric media Social media and advertisements created new and newer opportunity. To share the job ads, the growth of opportunity has increased the number of fake postings that may cause harassment of job seekers. Due to the security and consistency of their professional, academic, and personal information, people may show interest in new job postings. To attain people's reliability and belief, electronic and social media face an extremely hard challenge—the true motive of valid job postings. To make life easier and more developed, there are technologies around us but for professional life, they create an unsecured environment. For recruiting new employees, employees will be a great advancement if the job ads can be properly filtered, predicting fake job ads. This paper proposes various data mining techniques and classification algorithm such as random forest classifiers, NB classifiers, SVM, DT, K-nearest neighbors, and multi-layer perceptron, for the prediction of whether job advertisements are real or fraudulent. On the

Employment Scam Aegean Dataset (EMSCAD) we have examined 18000 samples. For this classification task as a classifier, the deep neural network performs great.

Keywords: deep neural networks, neural networks, machine learning, fake job prediction,

## I. INTRODUCTION

Now days, for the job seekers there is large number of opportunity for diverse and new jobs are opened in the field of industry and technology. Depending on the suitability, experience, qualification, time etc job seekers find out the job offers with the help of the advertisements. The effects of social media and online hiring procedures have a growing impact on the hiring process. Social media has a huge impact on recruitment processes since their advertisements determine how well they are completed. [1]

To find the perfect job for the job seeker, the post of a fake job creates inconsistency that may cause wastage of time. To face difficulties, an automated system in the field of human resource management that predicts false job postings opens a new window. [2]

The proliferation of online job platforms has led to an exponential rise in job postings across various digital mediums. Detecting and mitigating these deceptive job advertisements has become imperative to safeguard job seekers and maintain the integrity of online job marketplaces. This paper focuses on leveraging natural language processing

(NLP) techniques in conjunction with a multilayer perceptron (MLP) model to predict and identify fake job advertisements. By harnessing the power of NLP-driven text analysis and the predictive capacity of MLP, this study aims to develop a robust framework capable of discerning fraudulent job postings from authentic ones. The identification of fake job advertisements presents significant challenges due to the diverse nature of fraudulent postings.

From the job seekers, fraudulent people try to take illegal money from them. In the UK, a survey by ActionFraud shows that more than 67% of people are at great risk through online ads, but they are unaware of the job scams or job postings. Due to the job scam, almost 700,000 job seekers in the UK have complained of losing over \$500,000. Most of the fresh graduate students are targeted by fraud because they try to get jobs immediately, and for that, they are ready to pay extra money for them. Cybercrime prevention or avoidance strategies are ineffective in reducing this offence since job scams regularly alter their methods. [3]

If the fraudsters want to take the information of other people, such as their date of birth, income tax details, bank details, and insurance details, they create fake job advertisements. When the fraudster asks to pay money in advance, they show reasons like management costs, information security checking costs, admin charges, management costs, etc. In such cases, fraudsters act like themselves as employers and ask people about their driving license, bank statements, passport details, etc. as a pre-employment check. Fraudulently convincing someone to pay the amount in their personal account and saying that we will transfer it back is nothing but a legal money-mulling scam. [4]

In the online recruitment fraud (ORF) in recent times, employment scams scam is a major issue. Now days, most companies post the vacancies for jobs online, so these can be accessed timely and easily by job seekers. Since it is one of the ways of insinuating people to fraud, job-seekers offer employment taking money from them. On the fake job posting, to address the problem of identifying classification techniques, supervised learning algorithms are considered.

## II. LITERATURE SURVEY

**Habiba et al.** [1] proposed to predict if a job posting is real or fraudulent using different classification algorithms and data mining techniques like DNN, multilayer perceptron, RF classifiers, NB classifiers, SVM, DT, and KNN. On the Employment Scam Aegean Dataset (EMSCAD), we have examined 18000 samples. DNN performs great classifier in classification step. For this deep neural network classifier we have used three dense layers. The trained classifier shows 98% classification accuracy to predict a fraudulent job post. [1]

**Amaar et al.** [2] proposed to analyze whether these job ads are real or fraudulent using six machine learning models. Fake job advertisements are a famous scam. For these fake job openings, people apply, waste their money, send their data to scammers, and pay application fees to scammers. From online recruitment portals to detecting fraudulent job ads, we proposed a methodology that NLP and supervised ML techniques. To analyze the overall performance of the classifiers, we compared all models with both TF-IDF and BoW features. In this study, one of the challenges is the dataset we used. The model over fitted the majority of class

data due to an unbalanced ratio of fake and real job posting samples. By using TF-IDF as feature extraction and ADASYN as oversampling, ETC achieved 99.9% accuracy through experimental analysis.

**Mehboob et al. [3]** tackle the scam or fraud detection problem in the recruitment. In this paper, various important features of organisations, types of compensation, and job descriptions are proposed. An effective recruitment fraud detection model using the extreme gradient boosting method is constructed. It creates an algorithm that extracts the necessary components from job advertisements and tests it with three samples. For a two-step feature selection strategy, the features are further considered. Selected characteristics combined into a hybrid composition showed 97.94% accuracy and outperformed three state-of-the-art baselines. The most effective indicators the study finds are “organization\_type,” “company\_profile,” “salary\_range,” has multiple jobs,” and “required education.”

**Ranparia et al. [4]** used machine learning to minimize the number of such frauds and predict fake jobs so that students can stay alert and, if required, take informed decisions. In the job posting to analyse the pattern and sentiments, the model will use NLP. As a SNN the model will be trained and the Glove algorithm used. To predict jobs posted on LinkedIn, we will use a trained model in the real world to understand their accuracy. To improve the model, we worked on different methods to make it realistic and robust.

**Sudhakar et al. [5]** for classifying actual news and phoney information propose a novel algorithm. Based on machine learning algorithms these studies deal with SVM, logistic regression, novel

ensemble approach. It is separated into 620 sample size values for each group. With binary classes and a dataset of 10,000 records, consider the experiment. For the proposed novel approach, the results show a loss value of 05% and a better accuracy value of 95% compared with other algorithms. The results demonstrate that the suggested method is an ensemble strategy that, by changing certain parameters, combines AdaBoost with decision tree techniques to produce a much higher accuracy score.

**S. Vidros et. Al. [6]** Hiring is an essential procedure that was just recently moved to the cloud. In particular, the automated systems in charge of completing the online hiring process for new hires seek to improve the hiring process's speed, accuracy, and economy. However, the internet disclosure of these customary business practices has created new vulnerabilities that could jeopardize applicants' privacy and damage an organization's reputation. So far, the most typical incidence of online recruitment fraud (ORF) is employment scam. In contrast to pertinent internet fraud issues, ORF has not yet gotten the appropriate attention and has remained mainly unexplored up until this point. The current paper addresses this need by defining and outlining the features of this important and topical new area of cyber security research.

**B. Alghamdi et. Al. [7]** This work aims to prevent financial loss and invasion of privacy for both individuals and organizations by developing a reliable framework that can identify potential fraud in online recruitment settings. This study makes a significant contribution to the detection of online recruitment fraud (ORF) by presenting an ensemble strategy based on a RF classifier as a dependable

detection model. The present nature of online recruitment fraud detection sets it apart from previous forms of electronic fraud detection, as does the paucity of research on the subject. In order to accomplish the goals of this investigation, the researcher suggested a detection model. Support vector machines are utilized for feature selection, and Random Forest ensemble classifiers are used for classification and detection. The acquired accuracy was 97.41%, according to the results.

### III. PROPOSED METHOD

#### 2.1 EMSCAD dataset

The Employment Scam Aegean Dataset (EMSCAD), a freely available set of 17,880 real job advertisements, aims to give the research community a clear picture of the employment scam problem and can serve as a useful test bed for scientists working in the subject. The EMSCAD dataset was used in this project to train the system; the dataset values, such as pay and job description, are displayed in the subsequent rows, while the dataset column names are displayed in the row with the highest value.

#### 2.2 Data Pre-processing

For machine learning model preparing the raw data and make it suitable this process I known as data

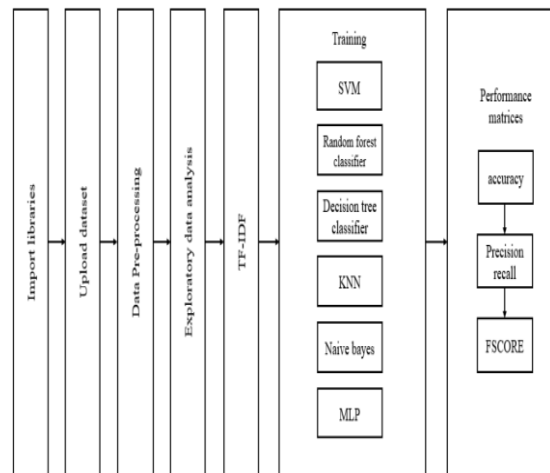
pre-processing. While creating a machine learning model it is a first and crucial step. Not all of the time we find clean, prepared data while starting a machine learning project. It is mandatory to clean it while doing any operation with data and put it into the formatted way. For the training and testing purpose we divide our dataset.

#### 2.3 TF-IDF Feature extraction

Term Frequency—Inverse Document Frequency is referred to as TF-IDF. For information retrieval it is one of the most important techniques to represent how important a specific phrase or word. TF-IDF does not immediately transform data into features. First, it transforms the dataset into raw strings, or vectors, with a unique vector for every word. The characteristic will then be retrieved using a specific method, such as cosine similarity, which is applicable to vectors, etc.

#### 2.4 Multilayer perceptron

One of the most popular neural network topologies in MDSS is MLP, which is a member of the supervised neural network class. The network of nodes (processing elements) that make up the multilayer perceptron is organized in layers.



**Fig. 1: Block diagram of proposed system**

### III. RESULT

### IV. CONCLUSION

At present, all over the world, fraud job detection is a great concern. In this paper, we examined the effects of fake job postings, which can be a very profitable area of research, creating many problems in detecting fake job postings. We worked with the EMSCAD dataset, which comprises real-life fraudulent job postings. In this paper, we examined machine learning algorithms such as a neural network, Random Forest, Naive Bayes, KNN, and SVM idea known as MLP. This research demonstrated the evaluation of MLP-based classifiers and machine learning.

### REFERENCES

[1]S. U. Habiba, M. K. Islam and F. Tasnim, "A Comparative Study on Fake Job Post Prediction Using Different Data Mining Techniques," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques

(ICREST), 2021, pp. 543-546, doi: 10.1109/ICREST51555.2021.9331230.

[2]Amaar, A., Aljedaani, W., Rustam, F. et al. Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches. Neural Process Lett 54, 2219–2247 (2022). <https://doi.org/10.1007/s11063-021-10727-z>

[3]Mehboob, A., Malik, M.S.I. Smart Fraud Detection Framework for Job Recruitments. Arab J Sci Eng 46, 3067–3078 (2021). <https://doi.org/10.1007/s13369-020-04998-2>

[4]D. Ranparia, S. Kumari and A. Sahani, "Fake Job Prediction using Sequential Network," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 339-343, doi: 10.1109/ICIIS51140.2020.9342738.

[5]Sudhakar, M., Kaliyamurthie, K.P. (2023). Efficient Prediction of Fake News Using Novel Ensemble Technique Based on Machine Learning Algorithm. In: Kaiser, M.S., Xie, J., Rathore, V.S. (eds) Information and Communication Technology for Competitive Strategies (ICTCS

2021). LectureNotes in Networks and Systems, vol 401. Springer, Singapore.  
[https://doi.org/10.1007/978-981-19-0098-3\\_1](https://doi.org/10.1007/978-981-19-0098-3_1)

[6] S. Vidros, C. Kolias, 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.

[7]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/jis.2019.103009>.

[8]Tin Van Huynh<sup>1</sup>, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen<sup>1</sup>, and Anh Gia-Tuan Nguyen,

“Job Prediction: From Deep Neural Network Models to Applications”, *RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020.

[9]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.

[10]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