



# Prediction of Fake Job Ad using NLP-based Multilayer Perceptron

**R. SANDHYA**- Student, Department of COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE), Geethanjali  
College of Engineering and Technology,

**N. VINAY KUMAR**- Student, Department of COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE), Geethanjali  
College of Engineering and Technology,

**B. PRAVALLIKA**- Student, Department of COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE), Geethanjali  
College of Engineering and Technology

**MR. G. VIJAY KUMAR**- Senior Assistant Professor, Department of COMPUTER SCIENCE AND ENGINEERING (DATA  
SCIENCE), Geethanjali College of Engineering and Technology

## Abstract

In modern times, the development in the field of industry and technology has opened a huge opportunity for new and diverse jobs for job seekers. With the help of advertisements of these job offers, job seekers find out their options depending on their time, qualification, experience, suitability, etc. The recruitment process is now influenced by the power of the 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 opportunities to share job details. However, the rapid growth of opportunities to share job ads has increased the percentage of fraudulent job postings, which causes harassment to job seekers. People lack interest in new job postings due to concerns about the 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 reliability. Technologies are around us to make our lives easy and developed but not to create an unsecured environment for professional life. If job ads can be filtered properly to predict false job ads, this would be a great advancement for recruiting new employees. Therefore, this project proposes to use different data mining techniques and classification algorithms like K-nearest neighbor, decision

tree, support vector machine, naive Bayes classifier, random forest classifier, and multi-layer perceptron to predict if a job advertisement is real or fraudulent. We have experimented on the 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 ad.

## 1. INTRODUCTION

In modern times, the development in the field of industry and technology has opened a huge opportunity for new and diverse jobs for 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. The 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 [1]. Social media and advertisements in electronic media have created newer and newer opportunities to share job details. However, the rapid growth of opportunities to share job posts has increased the percentage of fraud job postings which causes harassment to job seekers. So, people lack in

showing interest to new job postings due to preserving 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 reliability. Technologies are around us to make our life easy and developed but not to create an unsecured environment for professional life. If job 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].

### 1.2 Scope of the Project:

The scope of this project involves developing an automated system to predict fraudulent job postings, addressing the rising concern of job scams in online recruitment. It encompasses data collection and preprocessing, feature extraction, model development using various machine learning algorithms, evaluation, and validation. The project includes implementing the predictive model into a user-friendly system for real-time analysis of job postings, testing for robustness and accuracy, and deployment in a production environment. Documentation and maintenance planning ensure the system's usability and sustainability, ultimately providing job seekers and recruiters with a reliable tool to identify and mitigate fraudulent job postings, thus enhancing the integrity of the recruitment process.

## 2. LITERATURE SURVEY

Habiba et al. [6] 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. [6]

Amaar et al. [7] 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. By using TF-IDF as feature extraction and ADASYN as oversampling, ETC achieved 99.9% accuracy through experimental analysis.

Mehboob et al. [8] tackle the scam or fraud detection problem in the recruitment. In this paper, various important features of organizations, types of compensation, and job descriptions are proposed. An effective recruitment fraud detection model using the extreme gradient boosting method is constructed. The most effective indicators the study finds are "organization\_type," "company\_profile," "salary\_range," has multiple jobs," and "required education."

Ranparia et al. [9] 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 analyze 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.

Sudhakar et al. [10] 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.

## 3. OVERVIEW OF THE SYSTEM

### 3.1 Existing System

The existing system for job advertisement typically relies on manual review and human judgment to detect fraudulent job postings. This process can be time-consuming and prone to errors, leading to inefficiencies and potential risks for job seekers who may fall victim to scams. Moreover, the lack of automated tools means that fraudulent job postings may go undetected, posing a continuous threat to the integrity of the recruitment process.

#### 3.1.1 Disadvantages of Existing System

*Reliance on manual review:* The existing system relies heavily on manual review and human judgment to detect fraudulent job postings. This process can be time-

consuming and prone to errors, leading to inefficiencies in the recruitment process.

*Limited scalability:* With the increasing volume of job postings on online platforms, manual review becomes increasingly challenging and may not scale effectively to handle the growing workload.

*Higher risk of fraud:* The lack of automated tools and algorithms in the existing system increases the risk of fraudulent job postings going undetected, potentially exposing job seekers to scams and financial losses.

*Inconsistent detection:* Due to the subjective nature of manual review, the detection of fraudulent job postings may vary depending on the individual reviewer, leading to inconsistencies and inaccuracies in identifying scams.

### 3.2 Proposed System

To address the limitations of the existing system, we propose the development of an automated job scam detection system using machine learning and data mining techniques. The proposed system will leverage algorithms such as K-nearest neighbor, decision tree, support vector machine, naive Bayes classifier, random forest classifier, and multi-layer perceptron to predict if a job advertisement is real or fraudulent. By analyzing various features extracted from job postings, such as job descriptions, company profiles, salary ranges, and required education, the system aims to accurately identify fraudulent postings and mitigate the risks associated with online job scams. Additionally, the proposed system will offer a user-friendly interface for real-time analysis of job postings, enhancing the efficiency and reliability of the recruitment process for both job seekers and recruiters.

#### 3.2.1 Advantages of Proposed System

*Automation:* The proposed system offers automation through the use of machine learning and data mining techniques, reducing the reliance on manual review and improving the efficiency of fraud detection in job postings.

*Enhanced accuracy:* By leveraging advanced algorithms such as K-nearest neighbor, decision tree, support vector machine, naive Bayes classifier, random forest classifier, and multi-layer perceptron, the proposed system aims to achieve higher accuracy in identifying fraudulent job postings compared to manual methods.

*Scalability:* The automated nature of the proposed system allows for scalability to handle large volumes of job postings, ensuring timely and thorough analysis without sacrificing accuracy.

*Real-time analysis:* The system provides real-time analysis of job postings, allowing for prompt detection and mitigation of fraudulent activities, thereby reducing the

risks for job seekers and maintaining the integrity of the recruitment process.

### 3.3 Proposed System Design

In this project work, there are three modules and each module has specific functions, they are:

1. EMSCAD dataset collection
2. Data Pre-processing
3. TF-IDF Feature extraction
4. Multilayer perceptron

#### 3.3.1 EMSCAD dataset collection

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.

#### 3.3.2 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.

#### 3.3.3 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.

#### 3.3.4 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.

### 4. RESULT SCREEN SHOTS

### 3.4 Architecture

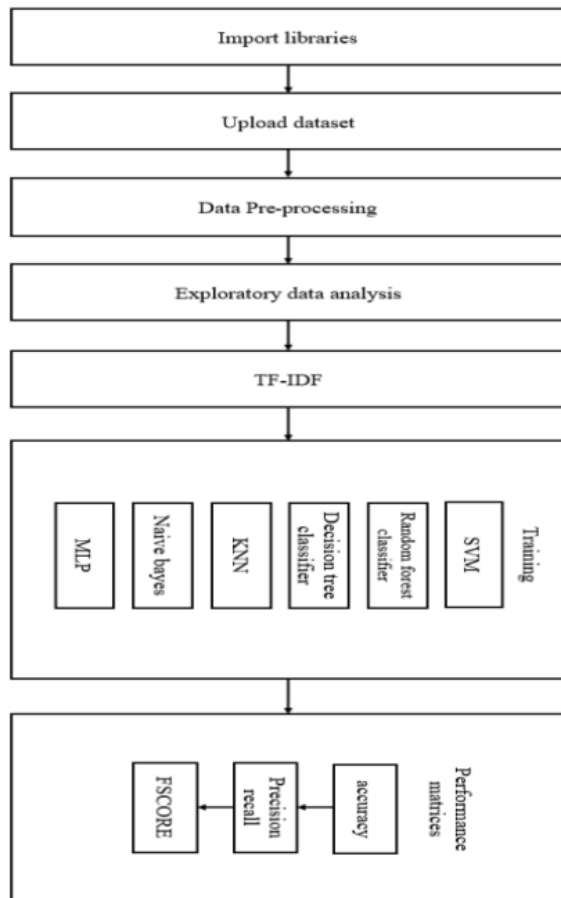
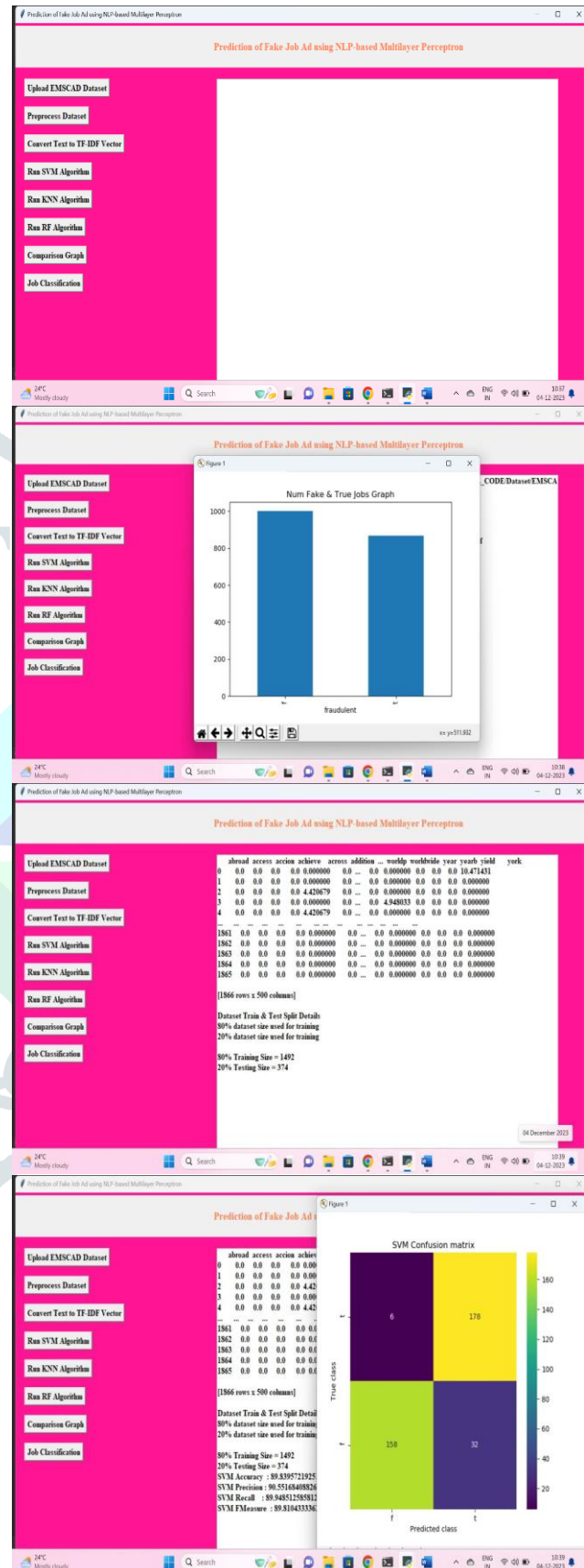


Fig 1: System Architecture



### 5. CONCLUSION

Job scam detection has emerged as a critical concern globally, prompting extensive research efforts to combat fraudulent job postings. In this project, we analyzed the impact of job scams using the EMSCAD dataset, which contains real-life fake job posts. Through experimentation with machine learning algorithms such as SVM, KNN, Naive Bayes, Random Forest, and the MLP neural network concept, we evaluated the effectiveness of various classifiers. Looking ahead, future endeavors could involve adapting data processing and model training methods to address the evolving nature of online scams, incorporating advanced NLP techniques and model architectures to capture intricate language patterns in fraudulent job postings, and enhancing transparency and user trust through Explainable AI techniques. By focusing on these aspects, the project aims to make significant contributions to combating fake job ads and improving the overall job search experience for individuals.

### 6. REFERENCES

- [1] 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.
- [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/jis.2019.103009>.
- [3] 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.
- [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] 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.
- [6] 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.
- [7] 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>.
- [8] 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>.



[9] 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.

[10] 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). Lecture Notes 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).

