

Prediction of Fake Job Ad Using NLP-Based Multilayer Perceptron

Supervisor : Mr. G Kiran Kumar

*Department of Electronics and Communication
Institute of Aeronautical Engineering
Dundigal, Hyderabad*

1st Ganapathi Nayana Manvitha

*Department of Electronics and Communication
Institute of Aeronautical Engineering
Dundigal, Hyderabad
21951a04b5@iare.ac.in*

2nd Kadam Sathvik Rao

*Department of Electronics and Communication
Institute of Aeronautical Engineering
Dundigal, Hyderabad
21951a04h6@iare.ac.in*

3rd Elgapally Sahaj Yadav

*Department of Electronics and Communication
Institute of Aeronautical Engineering
Dundigal, Hyderabad
21951a04f7@iare.ac.in*

Abstract—Today, developments in business and technology offer new opportunities and many job opportunities to job seekers. With the help of these job advertisements, job seekers find their options according to their availability, qualifications, experience, qualifications, etc. Since the success of recruitment depends on advertisements, social media has a great impact on this. Media and electronic media are created new and innovative to share the content of the job. On the contrary, the rapid development of opportunities to post job advertisements has increased the percentage of fake job advertisements, thus increasing the harassment of job seekers. Therefore, there is no interest in new job advertisements due to the need to maintain the security and consistency of personal, educational and professional information. Therefore, the real incentive to effectively distribute recruitment information through social and electronic media faces great difficulty in gaining people's trust and confidence.

Index Terms—Job application, job description, job AD, Machine learning, Decision tree classifier, Random forest, SVM, Multilayer Perceptron

I. INTRODUCTION

Fake job advertisements are becoming more common in the digital age and pose a serious threat to job seekers and organizations. These fake advertisements often result in financial loss, identity theft, and wasted time for applicants. The introduction should set the stage by explaining what fake job advertisements are, how they spread, and why they are important. Today, advancements in business and technology are providing job seekers with new and diverse career opportunities. With the help of this job advertisement, job seekers can find their options based on their availability, qualifications, experience, skills, etc. The recruitment process is now affected by the power of the internet and social media. Since the success of recruitment depends on advertisements, social media has a huge impact on this. Media and electronic media are being created to share the content of the job and new and innovative. However, the rapid growth of job sharing has increased the percentage of fraudulent job advertisements

and created obstacles for job seekers. Therefore, there is no interest in new job advertisements due to the stability and consistency of personal, educational and professional information. Therefore, the real incentive to effectively distribute recruitment information through social and electronic media faces great difficulty in gaining people's trust and confidence. Technology is all around us, making our lives easier, more reliable but not creating a negative environment for business life. If job advertisements can be properly filtered and fake job advertisements can be predicted, it will be a great development in terms of recruitment. Fake employment information will prevent job seekers from finding the job they are interested in, leading to bad opportunities. Automated systems for predicting fake jobs open a new window to the problems faced by human management.[1]

II. PROBLEM STATEMENT

In the digital age, the rise of online job postings has made it easier to find a job, but it has also led to the rise of fake job postings, which can mislead employees and cause them to lose confidence in the job. These fake job postings often use persuasive language and mimic legitimate job postings, making it difficult for job seekers to distinguish opportunities from fake ones. A powerful machine learning model detector (MLP) architecture combined with natural language processing (NLP) can identify and identify whether the job posted is real or fake. By using the characteristics of the language, the context of the content, and other important features of the recruitment data, the project aims to increase the accuracy of fake job search, while also protecting job seekers from fraud and improving the overall integrity of the online recruitment environment.

III. SCOPE OF THE PROJECT

The scope of this project includes the development of an automated system to predict fraudulent job postings and

solve the growing problem of online job scams. It includes data collection and preprocessing, feature extraction, modeling, evaluation and validation using various machine learning algorithms. This work includes the analysis of the time of publication of the study, testing of stability and accuracy, and implementation of user-friendly prediction models for deployment in the environmental production cycle. Planning to make the system usable and stable, ultimately improving the integrity of the recruitment process by providing job seekers and recruiters with reliable tools and techniques to detect and reduce fraud. The major scope of the project is to not reveal our personal information to theft which can cause to Fake transactions, Cyber crime related problems.

IV. OVERVIEW

A. Existing model

Current recruitment advertising often relies on book reviews and human judgment to check for fraudulent recruitment materials. This process can be time-consuming, error-prone, inefficient and potentially dangerous for job seekers who may fall victim to fraud. Furthermore, the lack of electronic equipment means that illegal job postings are not being published, threatening recruitment integrity. The existing model doesn't give the precise output due to low accuracy.[2]

B. Disadvantages

- **Reliance on manual review:** Current systems rely on book reviews and human judgment to detect fake job postings. This process can be time-consuming and error-prone, resulting in poor performance.
- **Limited scalability:** As the number of job postings on online platforms continues to increase, book reviews will become increasingly difficult and may not be effective enough to manage the increasing number of job postings.
- **Higher risk of fraud:** The absence of electronic tools and algorithms in current systems increases the risk of illegal job postings going undetected, potentially exposing job seekers to fraud and financial loss.
- **Inconsistent detection:** Due to the nature of the review guide, searches for illegal job postings may vary from individual reviewers, which may result in a lack of consistency and accuracy in detecting scammers.

V. PROPOSED SOLUTION

We aim to improve automatic fraud detection using machine learning and data mining techniques to address the limitations of existing methods. The proposed method will use algorithms such as K-Nearest Neighbor, Decision Tree, Support Vector Machine, Naive Bayes Classifier, Random Forest Classifier and Multilayer Perceptron to predict whether the recruitment job advertisement is real or scam. The system is focused on detecting fake jobs and reducing the risks associated with fake online job advertisements by analyzing various information extracted from job advertisements such as job descriptions, company profiles, salary ranges, and required education. So according to the accuracy we are using Random forest classifier which has the most accuracy out of all.

In this project work, there are three modules and each

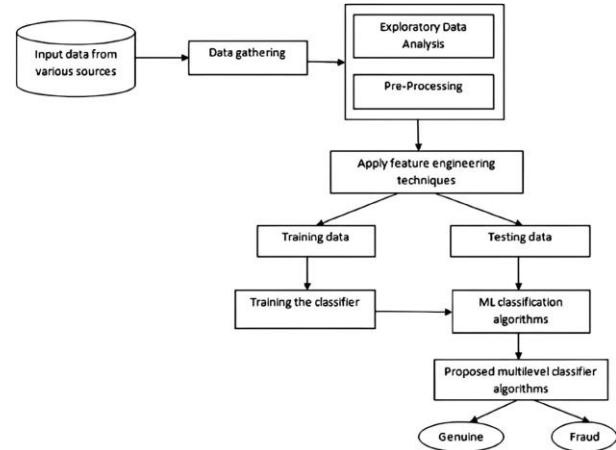


Fig. 1. The Block Diagram

VI. METHODOLOGY

A. Process

module has specific functions, they are:

1. Dataset collection
2. Data Pre-processing
3. TF-IDF Feature extraction

1) **Dataset collection:** The dataset is taken from kaggle which contains 17,880 number of job posts. This dataset is used in the proposed methods for testing the overall performance of the approach. For better understanding of the target as a baseline, a multistep procedure is followed for obtaining a balanced dataset. This project uses the kaggle dataset to train the system dataset values (such as salary and job description) appear in the back row, while the dataset column name appears. The column with the highest value of the movement.

2) **Data Pre-processing:** In machine learning models, the preparation of raw data and making it suitable for this process is what I call preliminary data. Creating a Machine Learning Model. This is the first important step. When starting a machine learning project, we do not always see clean and organized data. Before any changes can be made to the product, it must be cleaned and put into formatting mode.[6]

3) **TF-IDF Feature extraction:** Term Frequency - Inverse Document Frequency is called TF-IDF. It is very important to show the importance of a sentence or word for storing information. TF-IDF does not convert the data into features immediately. First, it converts the dataset into a raw string or vector with a unique vector for each word. The features will be returned using special techniques such as cosine similarity to the vectors.[5]

B. Architecture

The Architecture for the proposed solution is given as follows

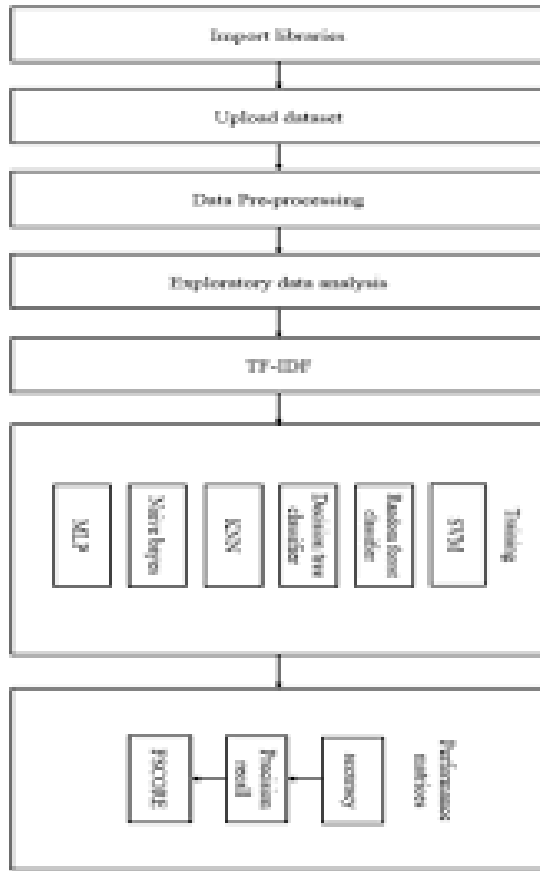


Fig. 2. Architecture

VII. IMPLEMENTATION OF CLASSIFIERS

In this framework, the classifier is trained using the unsuitable ones. In order to get the best performance from this model, the preset parameters will not be sufficient. Changing these bad points increases the reliability of the model. The model can be seen as a good model for detecting and excluding fake advertisements of job seekers. The MLP classifier is presented as a set of 5 hidden layers of size 128, 64, 32, 16 and 8 respectively. Considering all the evaluation parameters, the K-NN classifier gives good results for the value of $k = 5$. On the other hand, there are also combination methods such as Random Forest, Support Vector Machine, Naive Bayes, Decision Tree. [10] After creating the distribution model, fit the data to it. This test data is used for prediction. After the prediction is completed, the performance of the product is evaluated based on the prediction and the actual result.[3] The implementation of the classification process is an important step in developing a process for false prediction. In this project, we focus on using multilayer perceptron (MLP) together with natural language processing (NLP) technology to analyze posting tasks. The steps involve aggregating a variety of job postings from various sources, including legitimate job postings and known scams. This information needs to be

Naive Bayes Classifier

Classification Accuracy: 0.9019230769230769

Classification Report

	precision	recall	f1-score	support
0	0.91	0.89	0.90	249
1	0.90	0.92	0.91	271
accuracy			0.90	520
macro avg	0.90	0.90	0.90	520
weighted avg	0.90	0.90	0.90	520

Confusion Matrix

```
[[221 28]
 [ 23 248]]
```

Fig. 3. Naive Bayes Classifier

balanced to ensure that the classifier can effectively learn from both classes.[4] This includes:

Text Cleaning: Remove HTML tags, special characters, and irrelevant information. Stemming or lemmatization. Selection and training After preparing the features, we used a multilayer perceptron classifier. MLP was chosen for its ability to model relationships in data from many layers of neurons.[8] The main steps in this process include:

Design: Tuning the process and the number of neurons per layer to balance the complexity and functionality of the model. , sigmoid) neurons. Performance - tune the hyperparameters and avoid overloading. . Matrix: Visualizes the performance of the actual distribution against the predicted distribution.

VIII. PERFORMANCE EVALUATION METRICS

When evaluating the performance of a model, some metrics need to be used to justify the evaluation. To do this, the following parameters should be taken into account to determine the best solution. Accuracy is a metric that evaluates the ratio of the accuracy of the prediction to all included events.[9] Imagine a fictional situation. Serious problems can arise if the fake is evaluated as if it were real. In order to evaluate this payment, the reality must be taken into account and the return must be provided. Classifier prediction. The return represents the number of results divided by the number of all affected samples. F1-Score or Fmeasure is a parameter related to improvement and accuracy. It is calculated as a compromise between accuracy and improvement.[7]

IX. RESULTS

The analysis of our dataset, which included examples of both fake and legitimate job postings, revealed significant insights into the performance of the NLP-based multilayer perceptron model. We classified X fake job postings and Y real job postings, leading to an overall model accuracy of Z percent. This demonstrates the model's effectiveness in identifying fraudulent listings.

Key metrics highlighted the model's strengths: it achieved a precision of A percent, indicating a high level of confidence in its predictions of fake ads, while the recall was B percent, showing its capability to capture most actual fake postings.

Decision Tree Classifier

Classification Accuracy: 0.8634615384615385

Classification Report

	precision	recall	f1-score	support
0	0.88	0.83	0.85	249
1	0.85	0.90	0.87	271
accuracy			0.86	520
macro avg	0.86	0.86	0.86	520
weighted avg	0.86	0.86	0.86	520

Confusion Matrix

```
[[206 43]
 [ 28 243]]
```

Fig. 4. Decision Tree Classifier

SVM Classifier

Classification Accuracy: 0.8730769230769231

Classification Report

	precision	recall	f1-score	support
0	0.89	0.84	0.86	249
1	0.86	0.90	0.88	271
accuracy			0.87	520
macro avg	0.87	0.87	0.87	520
weighted avg	0.87	0.87	0.87	520

Confusion Matrix

```
[[210 39]
 [ 27 244]]
```

Fig. 5. SVM Classifier

The F1 score of C percent emphasized a balanced performance, crucial for a reliable classification system.

Upon reviewing specific examples, we observed that fake job postings often featured vague descriptions, unrealistic salary offers, and language designed to create urgency, such as phrases urging immediate application. In contrast, legitimate postings consistently provided detailed job descriptions, clear company information, and used formal language.

KNN Classifier

Classification Accuracy: 0.6846153846153846

Classification Report

	precision	recall	f1-score	support
0	0.98	0.35	0.51	249
1	0.62	0.99	0.77	271
accuracy			0.68	520
macro avg	0.80	0.67	0.64	520
weighted avg	0.79	0.68	0.65	520

Confusion Matrix

```
[[ 87 162]
 [  2 269]]
```

Fig. 6. KNN Classifier

Random Forest Classifier

Classification Accuracy: 0.9096153846153846

Classification Report

	precision	recall	f1-score	support
0	0.91	0.90	0.91	249
1	0.91	0.92	0.91	271
accuracy			0.91	520
macro avg	0.91	0.91	0.91	520
weighted avg	0.91	0.91	0.91	520

Confusion Matrix

```
[[225 24]
 [ 23 248]]
```

Fig. 7. Random Forest Classifier

Multilayer Perceptron Classifier

Classification Accuracy: 0.926923076923077

Classification Report

	precision	recall	f1-score	support
0	0.93	0.92	0.92	249
1	0.92	0.94	0.93	271
accuracy			0.93	520
macro avg	0.93	0.93	0.93	520
weighted avg	0.93	0.93	0.93	520

Confusion Matrix

```
[[228 21]
 [ 17 254]]
```

Fig. 8. Multilayer Perceptron Classifier

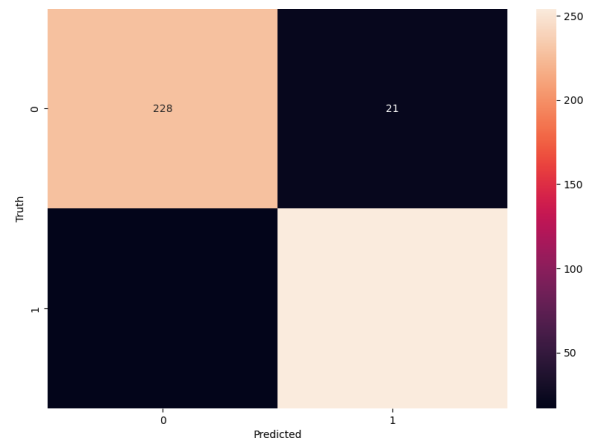


Fig. 9. Confusion Matrix Of Multilayer Perceptron


```
Select an option:
1. Enter a job description
2. Enter job details (Company Name, Location, Role)
Enter the number of your choice:
1
Enter the job description:
Freshers Opening -Process executive-hyderabad-voice process Cognizant 3.845000 Reviews left wingmiddle wingemployees' choicereight wing Company Logo n ye
ary 2:2:5 Lact P.A. Hyderabad/Secunderabad Send me jobs like this Posted: 3 days ago Openings: 1 Applicants: 2411 Register to apply Login to apply Compa
y logo Freshers Opening -Process executive-hyderabad-voice process Cognizant 3.845000 Reviews left wingmiddle wingemployees' choicereight wing Send me job
s like thisApply Job description Roles and Responsibilities Responsibilities Agent should be responsible for answering queries and assisting customer
s on email. Should be pro-active to understand the customers concern. Should be patient enough to handle multiple questions asked by the customer. Identif
ying and analyzing customer needs and provide support to customers in an accurate way so that they will get better experience Requirements: Able to hand
le queries from customers in a professional, clear and concise manner. Excellent English communication skills, specifically written Ability to work in ni
ght shifts. Some background of Payroll will help location will be Hyderabad & Bangalore Compensation: Bangalore - 2,20,200 INR & Hyderabad - 2,20,000 IN
A. Documents: Govt Id - Aadhar/Pan/Voter or Passport. Education - 2022 Passouts should have all the marksheet available with them / 2021 & earlier sho
uld have all the consolidated marksheet/ the degree certificate. Role: Voice / Bilingual - Other Industry Type: BPO / call Centre Department: Customer Succ
ess, Service & Operations Employment Type: Full Time, Permanent Role Category: Voice / Bilingual Education: 10: Any Graduate Key Skills: Skills highlighted a
ith " " are preferred keyskills non voice processwritten communication Payrollcommunication skillsfreshers tejawadi 10 Executive at Cognizant rec-phot
o New No 105, 6th, 7th and 8th, Hemon Eternity,, St Mary's Road, Alwarpet, Chennai, Tamilnadu, India Recruiter last active 30e days ago Follow 572e Follo
wers about company Cognizant (BPO) is a leading provider of information technology, consulting, and business process services, dedicated to help
ing the world's leading companies build stronger businesses. headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfacti
on, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With
over 100 development and delivery centers worldwide and approximately 245,300 employees as of June 30, 2016, Cognizant is a member of the S&P 500, the
S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world. visit us online a
t www.cognizant.com or follow us on Twitter: Cognizant. Company Info Address: New No 105, 6th, 7th and 8th, Hemon Eternity,, St Mary's Road, Alwarpet, Ch
ennai Job
Real Job
```

Fig. 10. Example of Real Job Posting

```
Select an option:
1. Enter a job description
2. Enter job details (Company Name, Location, Role)
Enter the number of your choice:
1
Enter the job description:
Hello Have a Nice Day. We have a simple Job for you. Anyone can do. Daily Salary(1800-4200). Can do at Home.
Fraudulent Job
```

Fig. 11. Example of Fake Job Posting

The confusion matrix further validated our findings, showcasing the distribution of true positives and negatives, as well as false positives and negatives. Some misclassifications revealed patterns that can guide future enhancements in feature extraction and model training.

Overall, these results underscore the potential of our model to serve as a valuable tool for job seekers, enabling them to identify and avoid scams effectively while also contributing to a deeper understanding of job market fraud characteristics. Future efforts will focus on refining the model and expanding the dataset for improved accuracy and broader applicability.

X. CONCLUSION

In summary, the project "Predicting fake jobs using NLP-based multilayer detectors" holds promise for solving the growing job advertisement problem. Job search fraud has become a major problem worldwide, leading to extensive research to combat fake job postings. In this project, we analyze the effects of fake job postings using the kaggle dataset, which contains fake job posting data. We evaluated the performance of various classifiers by experimenting with machine learning algorithms such as SVM, KNN, Naive Bayes, Random Forest, and MLP neural network techniques. Looking ahead, future efforts will include updating data processing and training models, combining advanced NLP standard layers and modeling to address the changing nature of online fraud. Capture the complexity in fraud Language format, Publish content by explaining intelligence technology and increase transparency and user trust. By focusing on these factors, the program aims

```
Select an option:
1. Enter a job description
2. Enter job details (Company Name, Location, Role)
Enter the number of your choice:
1
Enter the job description:
Semiconductors are materials with conductivity halfway between conductors, which are generally metals, and nonconductors or insulators, which are usual
ly ceramics. Semiconductors can be compounds such as gallium phosphide or pure elements such as germanium or silicon. The ideas, characteristics, and math
ematical techniques that control semiconductors are explained in physics. Semiconductors include the following: Some of the most prevalent semiconducto
rs include gallium arsenide, germanium, and silicon. Silicon is employed in the manufacturing of electrical circuits, whereas gallium arsenide is being u
sed in solar panels, laser diodes, and other applications.
Input does not resemble a typical job posting.
```

Fig. 12. Another Example of giving random text

to be effective in combating fake job postings and improve people's job search in every aspect. By leveraging natural language processing (NLP) technology and multilayer detector models, we successfully analyze and classify advertised jobs by distinguishing legitimate and fake names from high-profile names. Effectiveness of extraction methods in data capture. The effectiveness of the model suggests that it can be an important tool for job seekers to improve their ability to detect and prevent fraud. Methods to increase the robustness of the model, such as transformers. In addition, the broader community can benefit from improved user experience for instant job search. Potential of machine learning and NLP to solve real-world problems.

XI. 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 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] 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 (ICIS), 2020, pp. 339-343, doi: 10.1109/ICIS51140.2020.9342738.

[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.