

# FAKE JOB POST DETECTION USING MACHINE LEARNING

<sup>1</sup> K.V.Jhansi Rani, B.Tech, M.Tech, Assistant Professor, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>2</sup> Priyanka Rompalli, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>3</sup> Leela Nagababu Kathi, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>4</sup> Abhiram Tholam, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

<sup>5</sup> Ramakrishna Gudla, B.Tech, Department of CSE, Eluru College of Engineering And Technology, Duggirala, Andhra Pradesh-534004.

**Abstract:** Now a days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers as employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF) . To aware about the fake or real job posting which can avoid the end user for getting trap into fake job post .The project proposes an application that uses machine learning-based categorization approaches to avoid fake job postings on the internet. It aids in the detection of fake job postings among a large number of postings. Different classifiers are used for checking fraudulent post in the web and the results of those classifiers are compared for identifying the best employment scam detection model. It helps in detecting fake job posts from an enormous number of posts. Here in our project we implement different Machine Learning techniques and classification algorithm like support vector machine, naive bayes classifier, random forest classifier, logistic regression to predict a job post if it is real or fraudulent. The success rate for classification of the machine training is about 99% in detecting a fraudulent job posting.

## 1. INTRODUCTION

In modern time, the development in the field of industry and technology has opened a huge opportunity for new and diverse jobs for the 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. 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. Social media and advertisements in electronic media have created newer and newer opportunity to share job details. Instead of this, rapid growth of opportunity to share job posts has increased the percentage of fraud job postings which causes harassment to the job seekers. So, people lacks in showing interest to new job postings due to preserve 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 unsecured environment for professional life. If jobs 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. LITERATURE SURVEY

2.1 S. Vidros, C. Koliass, G. Kambourakis and L. Akoglu in 2017 were proposed The critical process of hiring has relatively recently been ported to the cloud. Specifically,

the automated systems responsible for completing the recruitment of new employees in an online fashion, aim to make the hiring process more immediate, accurate and cost-efficient. However, the online exposure of such traditional business procedures has introduced new points of failure that may lead to privacy loss for applicants and harm the reputation of organizations. So far, the most common case of Online Recruitment Frauds (ORF), is employment scam. Unlike relevant online fraud problems, the tackling of ORF has not yet received the proper attention, remaining largely unexplored until now. Responding to this need, the work at hand defines and describes the characteristics of this severe and timely novel cyber security research topic.

2.2 Devsmit Ranparia and Shaily Kumari in 2018 were proposed With increased number of data and privacy breaches day-by-day it becomes extremely difficult for one to stay safe online. Number of victims of fake job posting is increasing drastically day by day. The companies and fraudsters lure the job-seekers by various methods, majority coming from digital job-providing web sites. We target to minimize the number of such frauds by using Machine Learning to predict the chances of a job being fake so that the candidate can stay alert and take informed decisions, if required. The model will use NLP to analyze the sentiments and pattern in the job posting. The model will be trained as a Sequential Neural Network and using very popular GloVe algorithm. Ibrahim M. Nasser, Amjad H. Alzaanin and Ashraf Yunis Maghari in 2019 were proposed that Online recruitment provides job-seekers an efficient search and reach for jobs. It also helps recruiters searching for qualified candidates which improves the

recruitment process. However, employment scam has emerged as a critical issue. Some job posts are legitimate, and others are fraud. In this paper, an Artificial Neural Network based model is proposed to detect fraud job posts. The public Employment Scam Aegean Dataset (EMSCAD) is used with proper text preprocessing techniques for training and testing the proposed model. Our model has precision, recall, and f-measure of 91.84%, 96.02%, and 93.88% respectively.

2.3 Tejasva Bhatia and Jasraj Meena in 2020 were proposed that There are a numerous amount of job postings on the internet and sometime these vacancy postings turns out to be fake. Even on the reputed and trusted job advertising platforms, people fall prey to these fake advertisements . After selection in the job, hiring people start demanding for money and details of bank account. Good number of candidates gets duped and lose loads of money and sometimes even their current jobs. So it would be very helpful to identify if the job listed on the website is real or fake. In this paper, we have used machine learning to detect fraud job vacancy postings. In our proposed ML technique, we have applied two data balancing techniques namely "Adaptive Sympathetic" and "Synthetic Minority Oversampling Technique" in combination with "Term Frequency-Inverse Document Frequency" which is a feature extraction method. 2 Tin Van Huynh, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen and Anh Gia-Tuan Nguyen in 2020 were proposed Determining the job is suitable for a student or a person looking for work based on their job descriptions such as knowledge and skills that are difficult, as well as how employers must find ways to choose the candidates that match the job they require. In this paper, we focus on studying the job prediction using different deep neural network models including TextCNN, Bi-GRU-LSTM-CNN, and Bi-GRU-CNN with various pre-trained word embeddings on the IT job dataset. In addition, we proposed a simple and effective ensemble model combining different deep neural network models. Our experimental results illustrated that our proposed ensemble model achieved the highest result with an F1-score of 72.71%. Moreover, we analyze these experimental results to have insights about this problem to find better solutions in the future.

2.4 P. Santhiya ,S. Kavitha ,T. Aravindh, S. Archana ,Ashwanth V Praveen in 2021 were proposed the The emergence of the World Wide Web and the quick uptake of social media platforms (like Facebook, Instagram and

Whatsapp) have made it possible for information to be disseminated in ways that have never previously been seen in the history of humanity. Online hiring has changed the hiring pattern. In particular, putting job adverts on corporate websites and career portals includes looking for a sizable pool of qualified candidates worldwide. Unfortunately, it has become yet another forum for scammers, which threatens applicants' privacy and harms the reputation of companies. The topic of detecting recruiting fraud and scams is addressed in this case study. Three machine learning models are used in the construction of an effective recruitment fraud detection model, which includes a number of significant organizational, job description, and kind of remuneration features.

2.5 Banu Priya Prathaban, Subash Rajendran, G Lakshmi ,D. Menaka in 2021 were proposed In the recent times it is found to that there is a growing interest in Internet of Things (IoT) and its respective sophisticated cloud architectures. However, such development does not ensure the confidentiality, integrity due to its security vulnerabilities in its improvements. Therefore data breaching occurs rapidly in all sectors. Nowadays, several job-seekers are becoming the prime victim for such digital loophole fraudsters, when there is an advertisement for bulk requirement. In order to reduce frauds and scams that are designed to take advantage of people who seek jobs, a model to predict the genuineness of posting digital vacancies for job is real or fake. This paper presents the implementation of Prediction of Employment Scam Model (POESM), intending for the classification of fraudulent and non-fraudulent digital job posting advertisements. In our work, we used eight techniques such as Logistic Regression, Naive Bayes, Multiple Layer Perceptron, K-Nearest Neighbor, Decision Tree, Random Forests, Adaboost, Gradient Boosting classifiers. To analyze the dataset, supervised machine learning techniques is used to capture several essential information.

2.6 B. Alghamdi and F. Alharby in 2022 were proposed the research attempts to prohibit privacy and loss of money for individuals and organization by creating a reliable model which can detect the fraud exposure in the online recruitment environments. This research presents a major contribution represented in a reliable detection model using ensemble approach based on Random forest classifier to detect Online Recruitment Fraud (ORF). The detection of Online Recruitment Fraud is characterized by other types of electronic fraud detection by its modern and the scarcity

of studies on this concept. The researcher proposed the detection model to achieve the objectives of this study. For feature selection, support vector machine method is used and for classification and detection, ensemble classifier using Random Forest is employed. Shawni Dutta and Samir Kumar Bandyopadhyay in 2022 were proposed To avoid fraudulent post for job in the internet, an automated tool using machine learning based classification techniques is proposed in the paper. Different classifiers are used for checking fraudulent post in the web and the results of those classifiers are compared for identifying the best employment scam detection model. It helps in detecting fake job posts from an enormous number of posts. Two major types of classifiers, such as single classifier and ensemble classifiers are considered for fraudulent job posts detection. However, experimental results indicate that ensemble classifiers are the best classification to detect scams over the single classifiers.

### 3. EXISTING SYSTEM

Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF). In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them. Fraudulent job advertisements can be posted against a reputed company for violating their credibility. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs. In existing system they implemented machine learning algorithms like knn, svm still lack the accuracy in terms for prediction.

### DISADVANTAGES

- 1) The system is implemented by Conventional Machine Learning.
- 2) The system doesn't implement for analyzing large data sets.

### 4. PROPOSED SYSTEM

Employment scam detection will guide job-seekers to get only legitimate offers from companies. For tackling employment scam detection, several machine learning algorithms are proposed as countermeasures in this project. Supervised mechanism is used to exemplify the use of several classifiers for employment scam detection. In this project we propose logistic regression and ensemble classifier which outperforms the existing system.

### ADVANTAGES

- 1) The proposed has been implemented EMSCAD technique which is very accurate and fast.
- 2) The system is very effective due to accurate detection of Fake job posts which creates inconsistency for the job seeker to find their preferable jobs causing a huge waste of their Time.

## 5. UML DIAGRAMS

### 1. CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. It is also known as a structural diagram. Class diagram contains • Classes • Interfaces • Dependency, generalization and association.

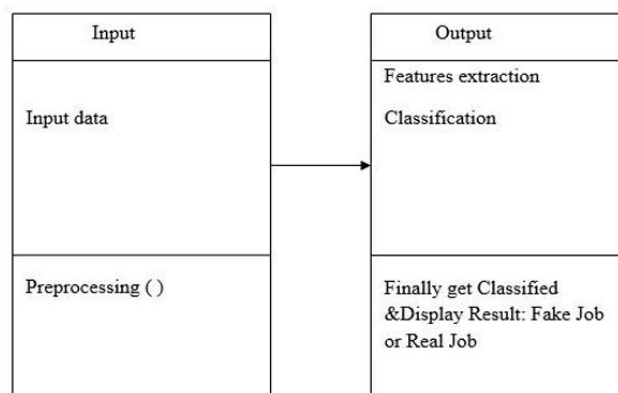


Fig 5.1 shows the class diagram of the project

### 2. USECASE DIAGRAM:

Use Case Diagrams are used to depict the functionality of a system or a part of a system. They are widely used to illustrate the functional requirements of the system and its interaction with external agents (actors). In brief, the purposes of use case diagrams can be said to be as follows

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Identify the external and internal factors influencing the system.

Use case diagrams commonly contains

- Use cases
- Actors
- Dependency, generalization and association relationships.

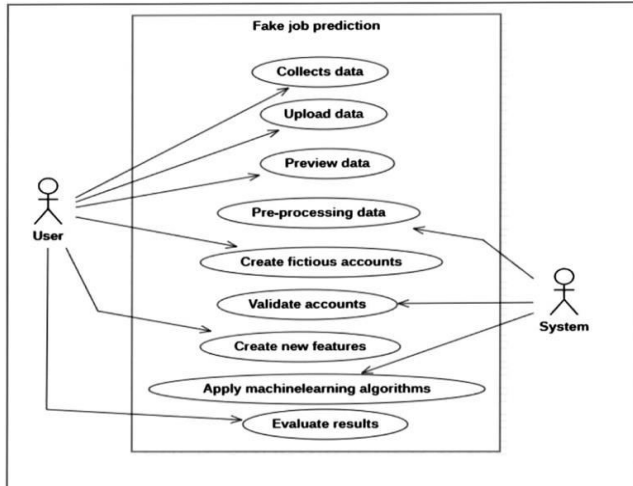


Fig 5.2 Shows the Use case Diagram

**3. SEQUENCE DIAGRAM:**

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. We can also use the terms event diagrams or event scenarios to refer to a sequence diagram. Sequence diagrams describe how and in what order the objects in a system function. Sequence diagrams are used to formalize the behaviour of the system and to visualize the communication among objects. These are useful for identifying additional objects that participate in the use cases. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.

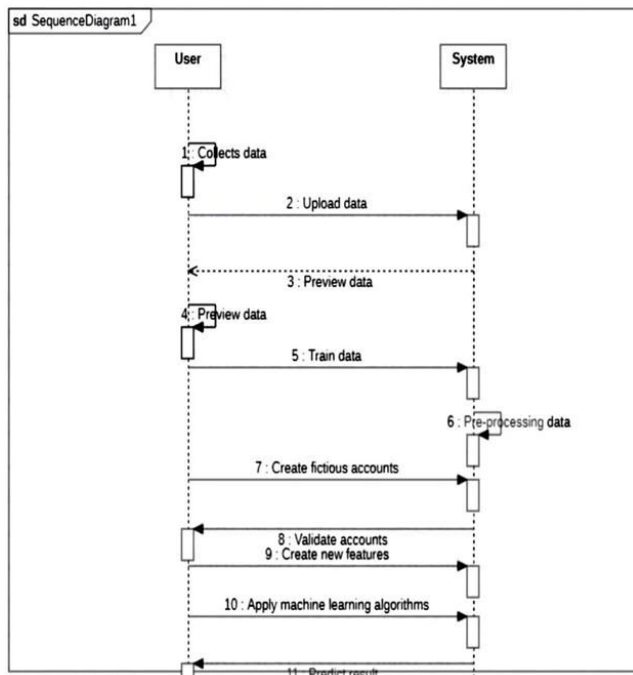


Fig 5.3 Shows the Sequence Diagram

**6. RESULTS**

**6.1 Output Screens**

In below screen click on ‘Upload Tweets Dataset’ button and upload dataset



Fig 6.1 Home Page

The user need to login twice one is Service Provider and another one is Remote User. Username and password is common for any user in Service provider. When the user clicks on the Service Provider button the user need to enter Username and Password.



Fig 6.2 Service Provider Login

After the user logged in first we have to train and test the datasets. Then you can view the Bar chart for different accuracies of different algorithms.

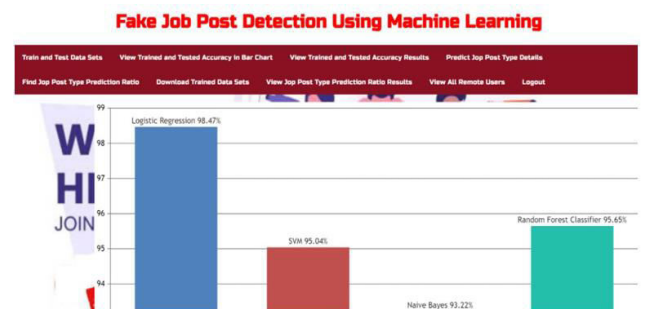


Fig 6.3 Accuracy in bar Chart

In the below screen the accuracy of different Algorithms displayed in the form of Line chart.



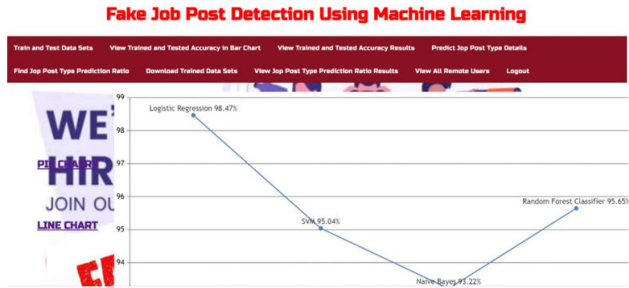


Fig 6.4 Accuracy in Line Chart

In the below screen shows the accuracy in pie chart

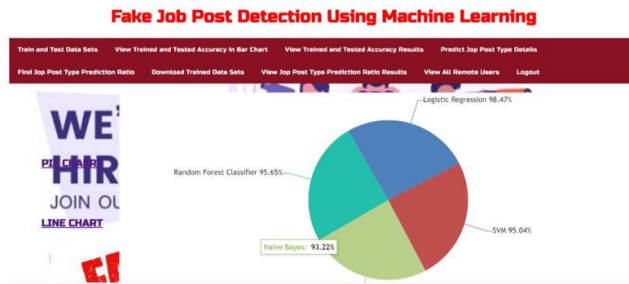


Fig 6.5 Accuracy in Pie Chart

In the above screen shows the prediction result.

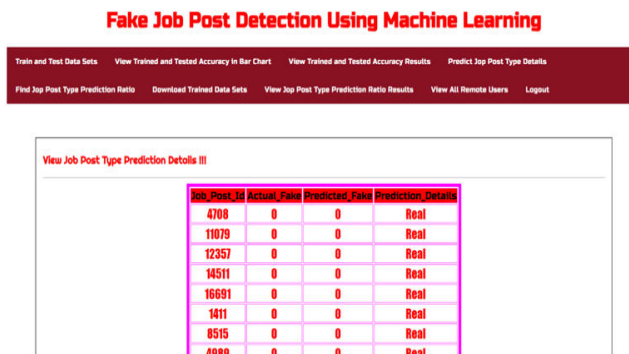


Fig 6.6 Prediction Result

Here is the main part of the project we can predict the job post is real or fake. By choosing different parameters of the job post like employment type, required education, required experience and function.

**7. CONCLUSION**

Job scam detection has become a great concern all over the world at present. In this paper, we have analyzed the impacts of job scam which can be a very prosperous area in research filed creating a lot of challenges to detect fraudulent job posts. We have experimented with EMSCAD dataset which contains real life fake job posts. In the project we have experimented both machine learning

algorithms (SVM, KNN, Naive Bayes, Random Forest .It shows a comparative study on the evaluation of traditional machine learning based classifiers. We have found highest classification accuracy for Random Forest Classifier among traditional machine learning algorithms.

**FUTURE SCOPE**

Detecting fake job posts using machine learning is a pertinent and promising application with significant potential. Here are some future scope areas for your project: 1. Enhanced Model Accuracy : Continuously improve the accuracy of your machine learning models. Explore advanced algorithms such as deep learning, ensemble methods, or hybrid models combining multiple algorithms to achieve better results. 2. Real-time Detection : Develop mechanisms to enable real-time detection of fake job posts. This could involve building efficient streaming data processing pipelines and deploying models to quickly identify and flag suspicious postings as soon as they are posted. 3. Unstructured Data Handling : Improve the model's capability to handle unstructured data such as free-form text in job descriptions. Explore techniques like natural language processing (NLP) for text preprocessing, feature extraction, and sentiment analysis to better understand the content of job postings. 4. Adversarial Robustness : Enhance the robustness of your model against adversarial attacks. Adversarial attacks attempt to deceive machine learning models by making small, carefully crafted perturbations to the input data. Developing defenses against such attacks is crucial for maintaining the reliability of your fake job post detection system. 5. User Feedback Integration : Implement mechanisms for incorporating user feedback into the model. Allow users to flag suspicious job postings and use this feedback to continuously improve the model's performance and adapt to evolving tactics used by scammers. 6. Cross-platform Integration : Extend the application of your model beyond job posting platforms to social media platforms, freelancing websites, and other online platforms where fake job postings may appear. Adapt the model to handle different data formats and sources effectively. 7. Regulatory Compliance : Ensure that your system complies with relevant regulations and privacy standards, particularly concerning the handling of personal data in job postings. Stay informed about legal developments in this area and adapts your system accordingly. Overall, the future scope of Fake job post detection is vast and varied, and further research and development can explore new and innovative applications of this technology.

## 8. REFERENCES

- [1] S. Vidros, C. Kolias, G. Kambourakis and L. Akoglu."Automatic Detection of Online Recruitment Frauds" In 2017 International Conference on Industrial and Information Systems (ICIIS).
- [2] Devsmit Ranparia, Shaily Kumari, Ashish Sahani."Fake Job Prediction using Sequential Network" In 2018 IEEE 15th International Conference on Industrial and Information Systems (ICIIS).
- [3] Ibrahim M. Nasser, Amjad H. Alzaanin and Ashraf Yunis Maghari."Online Recruitment Fraud Detection using ANN" In 2019 Palestinian International Conference on Information and Communication Technology (PICICT).
- [4] Tejasva Bhatia, Jasraj Meena."Detection of Fake Online Recruitment Using Machine Learning Techniques" In 2020 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)
- [5] Tin Van Huynh, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen and Anh Gia-Tuan Nguyen. "Job Prediction: From Deep Neural Network Models to Applications" In RIVF International Conference on Computing and Communication Technologies (RIVF), 2020.
- [6] P. Santhiya,S. Kavitha,T. Aravindh,S. Archana,Ashwanth V Praveen."Fake News Detection Using Machine Learning" in 2021 International Conference on Computer Communication and Informatics (ICCCI).
- [7] Banu Priya Prathaban,Subash Rajendran,G Lakshmi, D.Menaka."Verification of Job Authenticity using Prediction of Online Employment Scam Model (POESM)" in 2021 1st International Conference on Computational Science and Technology (ICCST). 42
- [8] B. Alghamdi and F. Alharby. "An Intelligent Model for Online Recruitment Fraud Detection" in 2022 Journal of Information Security.
- [9] Shawni Dutta and Samir Kumar Bandyopadhyay."Fake Job Recruitment Detection Using Machine Learning Approach" In 2022 International Journal of Engineering Trends and Technology vol. 68, no. 4, pp. 48-53, 2020.
- [10] Sathi Maheswara Reddy,Shaik Mohammed Ali,Kiran Mohan Battula, P.R.N.lakshmana Charan, Manazhy Rashmi."Web App For Predicting Fake Job Posts Using Ensemble Classifiers" In 2023 4th International Conference for Emerging Technology (INCET).