



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 [ijrset@gmail.com](mailto:ijrset@gmail.com)

 [www.ijrset.com](http://www.ijrset.com)

# Fake Job Post Prediction Using Machine Learning

Harshitha B <sup>1</sup>, Latha R <sup>2</sup>

P.G. Student, Department of Master of Computer Application, Sir M Visvesvaraya Institute of Technology,  
Karnataka, India<sup>1</sup>

Associate Professor, Department of Master of Computer Application, Sir M Visvesvaraya Institute of Technology,  
Karnataka, India<sup>2</sup>

**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 project proposed to Random Forest Classifier to predict a job post if it is real or fraudulent. We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. The trained classifier shows approximately 98% classification accuracy to predict a fraudulent job post.

**KEYWORDS:** Fake Job Post, Random Forest Classifier, Machine Learning, Legitimate Job, Decision Tree.

## I. INTRODUCTION

Now-a-days, getting a job is difficult. Before going to any interview you have to apply for a job, get registered then further go for an interview. The first and foremost step is to apply for a job according to the requirements of a company and as per the field a user wants to get a job in it. When you explore on internet you may find several job postings, those job postings may be a phony jobs or legitimate jobs. User may not find it easy as it is hard to say, the posted job is a fake or legitimate. So, we require a software to detect which is the fake job and which isn't, helping a number of people not to disclose their personal details to anyone by being aware of the fake job postings.

## II. RELATED WORK

1)Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset

AUTHORS : S Vidros, c Koliass, G Kambourakis, and L Akoglu

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 organization. So for, the most common case of Online Recruitment Frauds(ORF),is employment scam.

2)An Intelligent Model for Online Recruitment Fraud Detection

AUTHORS : B Alghamdi, F Alharby

This study 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. The 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).

## III. METHODOLOGY

### KNN:

The k-nearest neighbors algorithm (k-nn) is a versatile technique used for both classification and regression in pattern recognition. It determines the output based on the k closest training examples in the feature space. In classification, it assigns an object to the class most common among its k nearest neighbors. In regression, it estimates the property value by averaging the values of the k nearest neighbors.

**Random forest classifier:**

A random forest is a powerful ensemble learning method that combines multiple decision tree classifiers. It improves predictive accuracy and reduces overfitting by fitting these decision trees on different subsets of the dataset and using averaging to generate the final predictions. The size of the subsets, known as sub-samples, can be controlled using the max\_samples parameter. By default, if bootstrap=true, random forest uses sub-samples, but if bootstrap=false, it utilizes the entire dataset to build each tree.

**XGboost:**

Xgboost is a highly effective ensemble machine learning algorithm that is based on decision trees and utilizes a gradient boosting framework. While artificial neural networks excel in prediction problems with unstructured data like images and text, decision tree-based algorithms are currently regarded as the top performers for small-to-medium structured or tabular data.

**Evaluation:**

After creating a model, it is crucial to validate it using real-time data values. This process is commonly referred to as validation, where the predicted value of the model is compared with the actual output value. By assessing the agreement between the predicted and actual values, the model's performance and accuracy can be evaluated.

**IV. EXPERIMENTAL RESULTS**

This phase is very important and crucial part in any project as it gives the functionality to the design of the project. Each screenshots represents the part of project which is there in our project.

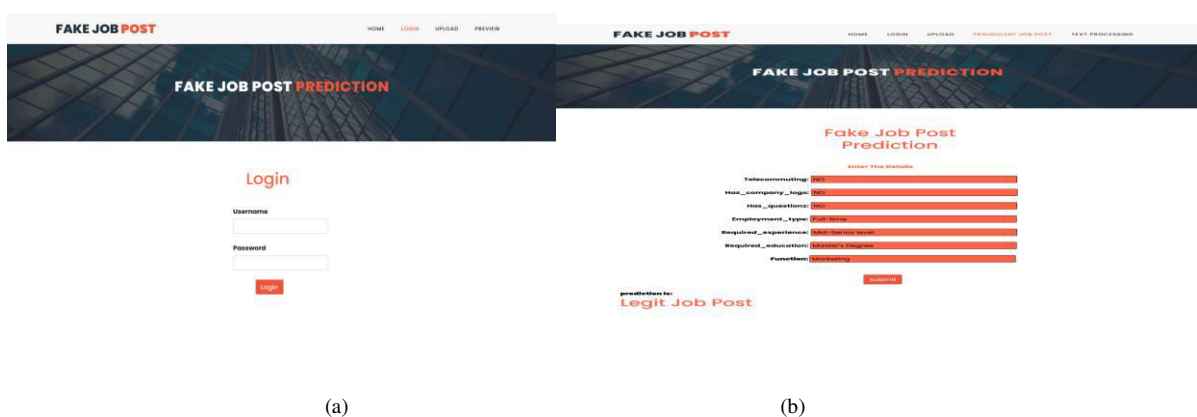


Fig 1 : 1. Username/Email: Users can quickly input either their registered username or email address in this area to access their accounts.

By users who may have forgotten their username but remember their email address and vice versa this flexibility offers a seamless login experience.

2. Password: The platform uses a strong password field to ensure the highest level of security.

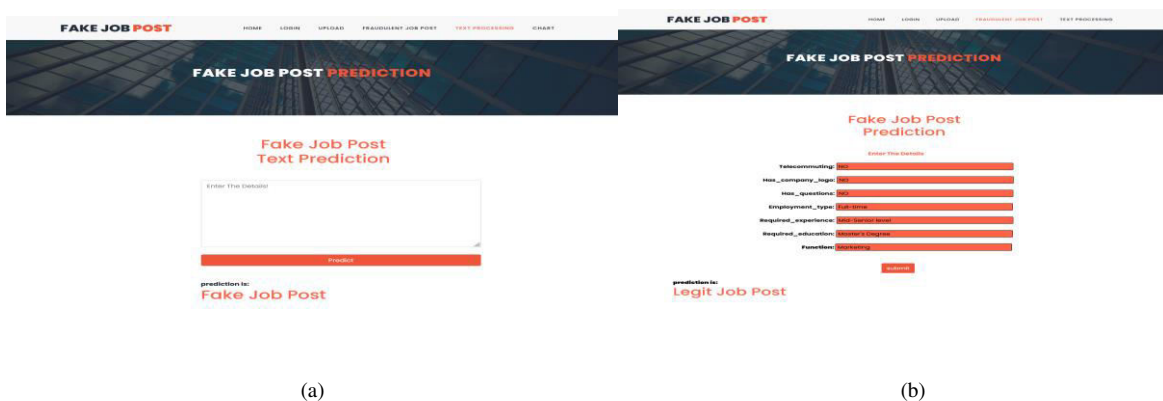


Fig 2 : It has seven attributes using this it displays the job whether the job is real or fake. This page displays the job is legit job post.

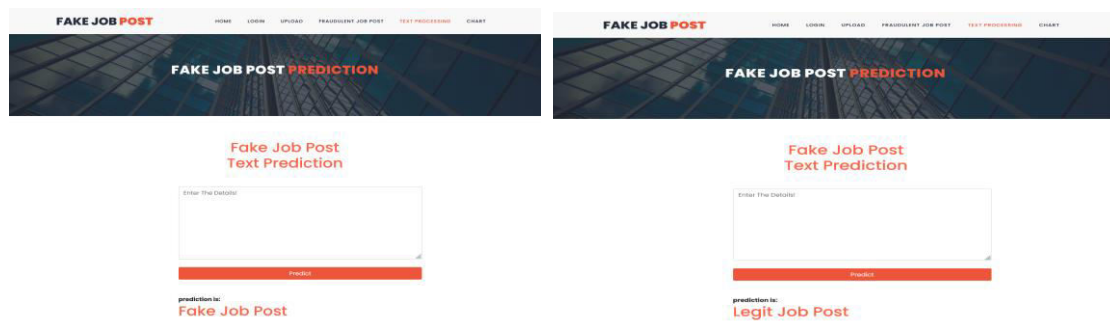
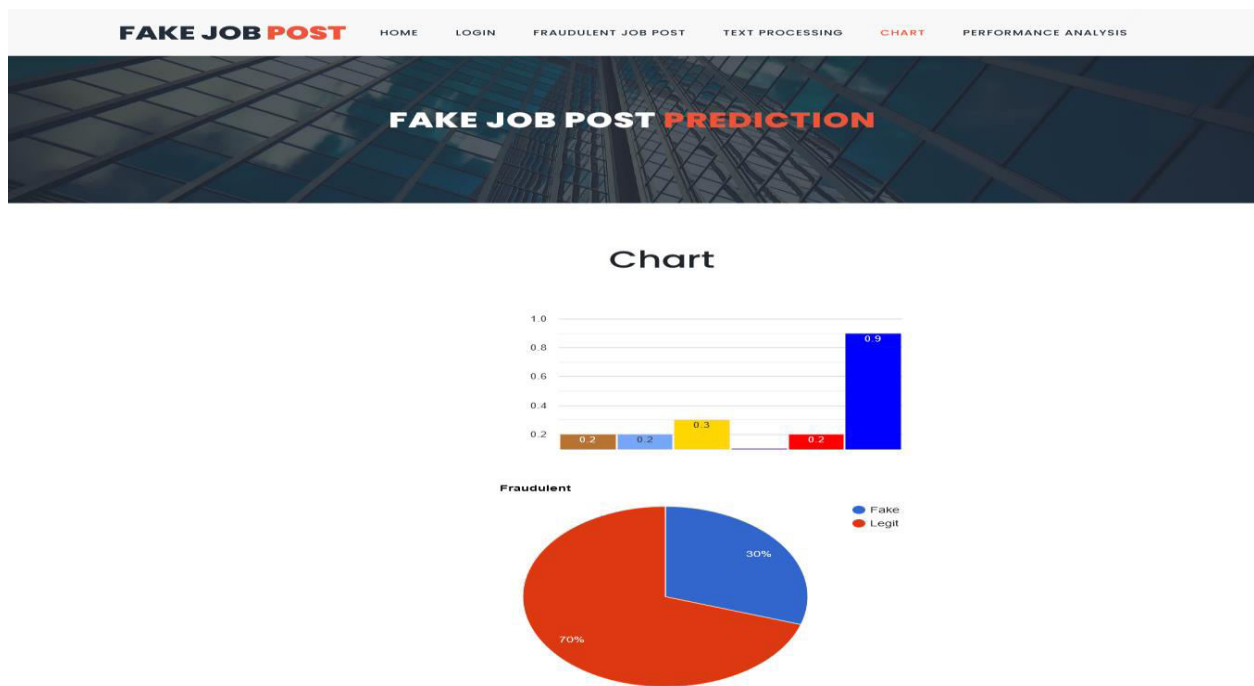


Fig 3 : Fake job post prediction details page. we need to enter the details of the job and it has seven attributes those are used in the fields after entering the fields we need to submit.



(f)

Fig. 4 : When we enter the information for any firm profile. The content includes a corporate profile as well as company logo. The machine will then predict whether or not the task is a forgery. This text depicts a bogus job posting.

### V. CONCLUSION

The detection of job scams has recently become a major problem worldwide. We have examined the effects of employment scams in this project since they might be a very lucrative topic of study and make it difficult to identify the posts of fake job. We made use of EMSCAD dataset, which includes real-time job postings. Random forest classifier gives 98 percent of accuracy then the previously used algorithms like SVM, Decision tree classifier, etc which gives 90 percent of accuracy. We are making the hiring procedure through online safer, by avoiding frauds and scams in the job. Therefore, you can go for applying the jobs through online process. Therefore, avoiding the financial losses of a person and protecting the personal information of a person.





|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207143 |

#### REFERENCES

- [1] S. Vidros, C. Koliass, 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. 155176, <https://doi.org/10.4236/iis.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)*, 2020T.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.423**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details