International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

# Developing A Model to Detect Fraudulent Job Postings: Fake vs. Real

**T.Nandini[1],S.Gnana Chandrika[2], P.Mounika[3], V.Sandeeep Kumar[4]**
**UG students, Dept of CSE, Kallam Haranadhareddy Institute of Technology, AP, India.**

## Abstract

The study suggests an automated method based on categorization techniques based on machine learning to stop fake job ads online. These days, a lot of businesses like to post their vacant positions online so that job hunters may easily find them. Yet, this might just be a ruse used by con artists to get others to labour for them in exchange for money. This hoax deceives many people, who end up losing a lot of money. By performing an exploratory data analysis on the data and applying the insights obtained, we can distinguish between job ads that are fake and those that are not. A machine learning strategy that makes use of several categorization algorithms is employed to detect fake posts.In this project, we use various machine learning approaches to examine the bogus job postings on social networking sites like Linkedin, Naukari, etc. to forecast fraudulent employment based on data that includes job names, locations, departments, income ranges, employment types, etc. from the prior year. It analyses legitimate job ads in addition to focusing on foreseeing fake job postings. Based on historical data of fraudulent and legitimate job listings, the system would train the model to classify jobs as authentic or fraudulent. It will use two or more machine learning algorithms, choosing the one that predicts if a job advertisement headline is real or not with the highest accuracy score.

Keywords: Fake Job, Online Recruitment, Machine Learning, Real.

## 1.Introduction

Even on well-known job advertising websites, there are numerous job listings that never look fraudulent. But, following the selection, the so-called recruiters begin requesting payment and bank account information. Several candidates fall into their trap, losing a lot of money and occasionally their current jobs. So, it is preferable to determine whether a job posting on the website is genuine or fraudulent. Manually identifying it is extremely challenging and nearly impossible. We can train a model for fictitious job classification using machine learning. It can be taught from past real and fraudulent job adverts and effectively recognize a false job. Economic difficulty and the effects of the coronavirus have significantly decreased the availability of jobs and led to employment loss for many people. Fraudsters would love to exploit a circumstance like this. These con artists are relying on people's desperation as a result of an unprecedented incident, and many people are

falling victim to them. Most scammers do this in order to get personal information from the target of their fraud. Personal data includes things like addresses, bank account numbers, and social security numbers. Scammers entice victims with a great job offer before requesting money in return. As an alternative, they can require the job seeker to make a monetary investment in return for the assurance of employment.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Several websites can help recruiters identify qualified candidates. Sometimes, only to make money, fake recruiters may advertise a job on a job board. Numerous job boards experience this problem. Afterwards, those looking for legitimate jobs visit a new job portal; however, phoney recruiters also move to this portal. So, being able to tell the difference between real and fake employment possibilities is crucial. One of the most serious issues that has recently been addressed in the area of online recruitment frauds is employment fraud (ORF). Nowadays,

a lot of businesses like to post their job positions online so that candidates may easily and quickly find them. Yet, this might be a type of fraud being committed. Natural language processing and machine learning techniques can be used to overcome this hazardous issue (NLP). A machine learning strategy that makes use of several categorization algorithms is employed to detect fake posts. In this case, a classification method separates fake job postings from a larger pool of legitimate job postings and alerts the user. To begin with, research is being done on supervised learning algorithms as classification techniques to handle the problem of identifying fraudsters on job advertising.

Input variables are mapped to target classes using training data by a classifier. The classifiers used in the paper to separate fake job posts from the others are briefly described. Single-classifier predictions and ensemble-classifier predictions are the two categories into which these classifier-based predictions fall. The study suggests an automated method based on categorization techniques based on machine learning to stop fake job ads online. These days, a lot of businesses like to post their vacant positions online so that job hunters may easily find them. Yet, this might just be a ruse used by con artists to get others to labour for them in exchange for money. This hoax deceives many people, who end up losing a lot of money. By performing an exploratory data analysis on the data and applying the insights obtained, we can distinguish between job ads that are fake and those that are not. A machine learning strategy that makes use of several categorization algorithms is employed to detect fake posts.

## 2.Literature Survey

They typically do not correlate with other important features like robustness, fairness, interpretability, and so forth. They also do not account for the (sometimes considerable human and CPU/GPU) time required for hyperparameter fiddling. When the model user is not the model developer, a related problem, particularly in industrial-scale artificial intelligence (AI), occurs. The training data or the testing data might then not be available. Instead, a model that has already been trained—referred to as a pretrained model—may be provided, and the user may choose to utilise it as-is or modify, compress, and/or fine-tune it before using it.

Being naive, but in our experience typical among ML practitioners and ML theorists, one can say nothing about the quality of an ML model if they do not have access to training or testing data. This may be accurate in worst-case theory, but since models are employed in practise, a useful theory is required to direct that practise. In addition, if ML is to be applied in industry, it will need to be compartmentalised in order to scale: some groups will collect data, others will create models, and yet others will employ those models. Users of models cannot be expected to understand the minutiae of how models were created, the details of the data used to train the model, the values of the loss function or hyperparameters, the exact regularization method employed, etc.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

[1] In 2020, S Siddaraju, M Sivaranjani, V Sivasakthi, and S Tamilselvan recommended utilising the KNN algorithm to anticipate the Trends of Quality-Oriented Employment in order to assist and prepare for upcoming employment. [2] In their project, "A Machine Learning Approach for Predicting Execution Time of Spark Jobs," Sara Mustafa, Iman Elghandour, and Mohamed A. Ismail suggested three main approaches have been used to predict the execution time of queries. This project supports the execution of various types of workloads. [3] In their project "Student Placement Prediction Using Machine Learning," Shreyas Harinath, Aksha Prasad, Suma H S, Suraksha A, and Tojo Mathew in 2019 proposed predicting student placement status using two attributes, areas and CGPA outcomes. Utilizing advanced machine learning techniques, K. Sripath Roy, K. Roopkanth, V. Uday Teja, V. Bhavana, and J. Priyanka published a project titled "Student Career Prediction" in 2018. This project primarily focuses on the computer science domain candidates' predicted career areas. [5] The idea of "Automatic Student Analysis and Placement Prediction using Advanced Machine Learning Algorithms" was published in 2019 by Kachi Anvesh, B. Satya Prasad, V. Venkata Sai Rama Laxman, and B. Satya Narayana. It is an automatic prediction based on the student's qualifications. Students can assess their suitability for various job roles. Evaluation of a model's quality is a typical issue in machine learning (ML). To do this, a common strategy is to train a model and then assess the training/testing error. This strategy comes with a lot of issues. The training/testing curves provide very little information about the model's general characteristics.

both the horizontal and vertical directions. The most typical approach is using a 1D mask, [-1 0 1].

Binary robust independent elementary features (BRIEF) [30,57]: BRIEF is a binary descriptor that is simple and rapid to compute. This descriptor is based on the variations in the pixel intensity that are related to the family of binary descriptors such as binary robust invariant scalable (BRISK) and fast retina keypoint (FREAK) in terms of assessment. To decrease noise, the BRIEF description smoothens the image patches. After that, the disparities between the pixel intensity are employed to express the descriptor. This descriptor has attained the best performance and accuracy in pattern recognition.

## 3.Problem Statement

Hence, in order to help them effectively address whether a given job is fraudulent or real, we would like to examine the task of job prediction. The assignment is laid out as follows.

Input: given a job title and employment category that were obtained from online job-finding websites.

Output: determines whether a job title is false or not.

## 4.Existing System

Economic difficulty and the effects of the coronavirus have significantly decreased the availability of jobs and led to employment loss for many people. Fraudsters would love to exploit a circumstance like this. These con artists are relying on people's desperation as a result of an unprecedented incident, and many people are falling victim to them. A number of career categories have emerged as a result of the rapid development of information

technology (IT) in recent years. Finding a career that fits their knowledge and skills they have acquired at school or while working can be difficult because of diversity.

The rapid growth of information technology (IT) in recent years has produced a wide range of work categories as well as the qualifications needed for each type of IT position. It might be difficult for students or job searchers to find a position that utilises the knowledge and skills they have acquired at school or via employment due to variety. Also, the hiring organisation needs spend a lot of time personally screening candidate profiles to select those who are qualified for the position for which they are hiring. This is because there may be hundreds or even thousands of applications. So, in order to assist them in solving the aforementioned problems, we would want to investigate the challenge of IT job prediction. In order to anticipate a job based on job descriptions that include job criteria, knowledge, abilities, and interests, several machine learning and natural language processing techniques are used.

## 5.Proposed System

This demonstrates how predicting fake job posts using data gathered from several job sites or job portals, such as linked in, monster, first Naukri, etc., is possible. The collection includes information on up to 17880 entities. Based on the organisation and employment functions, all the data are categorised. to determine whether the position is genuine or false. With low, moderate, and high levels of openings and opportunities, it displays this information visually. Future pupils should be able to predict bogus jobs based on work titles and employment types.

This demonstrates how to forecast job openings and vacancies for the following years using data gathered from a variety of employment sites and job portals, including linked in, monster, first Naukri, and others. The dataset includes information on up to 10,000 companies. Based on the organization and employment functions, all the data are categorized. to forecast a job's scope after 5 years. With low, moderate, and high levels of openings and opportunities, it displays this information visually. the capacity to anticipate future employment opportunities for pupils based on their qualifications, qualifications, and graduation.

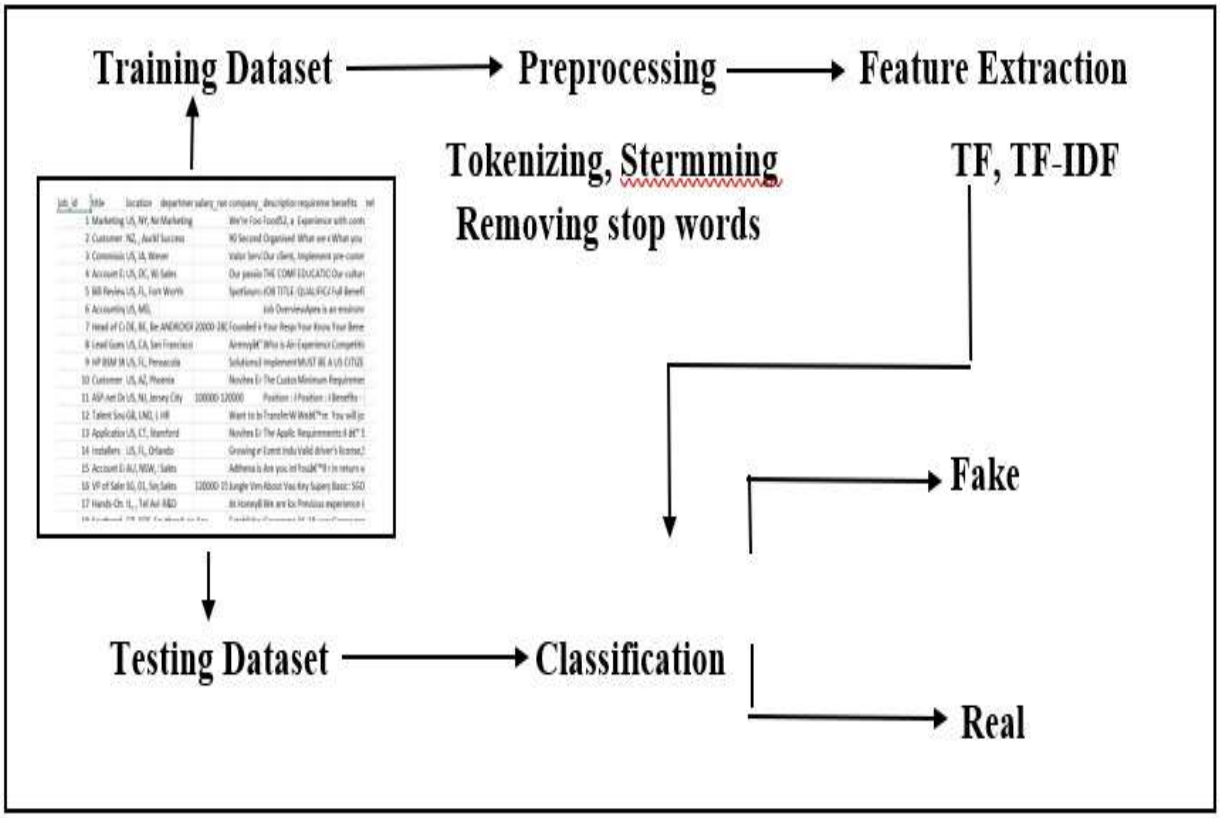
. Data Gathering For analysis and prediction, the system requires data from several employment portals. The system requires two primary datasets, which have been gathered and stored independently.

1. To determine the skills trends: Data from the Times Job portal's I.T department has been scraped using an internet web scraper application and saved as a.csv file (comma-separated-value).

2. To forecast salaries: A salary-activity model was trained using data gathered from different reliable sources, including Naukri, Monster, and others, and these wages were utilized to predict salaries.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

## Block Diagram




# 6.METHODOLOGY

## 6.1.TRAINING DATASET:

17,880 data entries for job postings make up this Kaggle dataset. Before fitting this data into any machine learning models or classifiers, we must first preprocess it to make it ready for prediction. This dataset is subjected to some pre-processing procedures before being fitted to a classifier. Other pre-processing techniques include missing values removal, stop-word removal, irrelevant attribute removal, and needless space removal. In order to create a feature vector, category encoding must first be applied to the dataset.

## 6.2.PRE PROCESSING OF DATASET:

Preprocessing involves converting unclean data into a clean data set. The dataset is preprocessed to look for missing values, noisy data, and other anomalies before the algorithm is performed. Stop words, unnecessary characteristics, and extra space are also removed from the text, along with noise and uninformative characters and phrases. The data set needed to be pre-processed due to its nature before being given into the classifier. The data is textual, so before we can make any predictions, we must convert it to a numerical representation. NLP is utilised here. NLP (Natural Language Processing) (Natural Language Processing) A computer program's capacity to comprehend natural language, or human language as it is spoken and written, is known as natural language processing (NLP).

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

## 6.3FEATURE EXTRACTION:

A step in the dimensionality reduction procedure, feature extraction shrinks and splits a big set of raw data into smaller groupings. The most important characteristic of these big data sets is the high number of variables. It takes a lot of processing resources to process these variables. Hence, feature extraction helps to extract the best feature from those enormous data sets, substantially reducing the amount of data, by choosing and combining variables into features. These features are easy to use while still accurately and uniquely describing the underlying data collection. In this study, features are extracted using TFI-DF.

## 6.4CLASSIFICATION:

In this part, classifiers are trained using the appropriate parameters. This framework employed Naive Bayes, SVM, and Logistic Regressor models for predictions. SVM has a number of distinctive qualities, which have helped it become well-known and produce encouraging experimental findings. In authentic input space, SVM creates a hyper level to divide the data points.

Whereas logistic regressors use a logistic regression equation to establish the association between the dependent variable and one or more independent variables, Naive Bayes predicts the probability of various classes depending on data.

The classification algorithm used was the Random forest ensemble classifier, which was constructed utilizing a number of tree structured classifiers.

# 7.Result

Job Prediction: Fraudulent or Real.

```
input_text = [ "Quality Improvement Manager - Full-time"]
#'Data Entry Admin/Clerical Positions - Work From Home '
#'Customer Service - Cloud Video Production-Full-time'
```

```
input_data = vect.transform(input_text)
prediction = dt.predict(input_data)

if (prediction[0] == 0):
  print("This advertisement belonging to fake job post category")

else:
  print("This advertisement belonging to real job post category")
```

```
This advertisement belonging to fake job post category
```

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

```
[ ] input_text = ["Quality Improvement Manager-Full-time"]

    input_data = vect.transform(input_text)

    prediction = dt.predict(input_data)

    if (prediction[0] == 1):
      print("This Prediction belonging to fake job post category")

    else:
      print("This Prediction belonging to real job post category")

    This Prediction belonging to real job post category
```

## 8.Conclusion

You will only receive offers from reliable companies. With the purpose of finding job frauds, several machine learning techniques are suggested. In this paper, we talk about defences. The employment of various mechanisms is demonstrated using supervised mechanisms. classifiers for spotting employment fraud. The outcomes of the trials demonstrate the efficiency of Random Forest. In classification, the classifier outperforms its competitors. The proposed method's accuracy rate was 97%. This is considerably more effective than current methods.

## 8.1.Future Scope

The idea of employing people by business enterprises through an online process is being carried out thanks to an upgrade in technology. This enables businesses to hire the right people for the job more quickly and immediately. That will be economical as well. One can quickly find a job that matches their skills and area of interest by browsing the internet. The people may not be aware that the advertised jobs could be false or real. We developed a new piece of software that anticipates job postings and determines whether they are real or false in order to eliminate issues of this nature. We're creating a system.

## References

[1] S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepakkumar, “Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms,” Rev. GEINTECGESTAO Inov. E Tecnol., vol. 11, no. 2, pp. 642–650, 2021.

[2] B. Alghamdi and F. Alharby, “An intelligent model for online recruitment fraud detection,” J.Inf. Secur., vol. 10, no. 03, p. 155, 2019.

[3] “Report | Cyber.gov.au.” https://www.cyber.gov.au/acsc/report (accessedJun. 19, 2021).

[4] A. Pagotto, “Text Classification with NoisyClass Labels.” Carleton University, 2020.

[5] “Employment Scam Aegean Dataset.” http://emscad.samos.aegean.gr/ (accessed Jun. 19, 2021).

[6] S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, "Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset," Futur. Internet, vol. 9, no. 1, p. 6, 2017.

[7] S. Lal, R. Jiaswal, N. Sardana, A. Verma, A. Kaur and R. Mourya, "ORFDetector: Ensemble Learning Based Online Recruitment Fraud Detection," 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-5, doi: 10.1109/IC3.2019.8844879..

[8] Bandyopadhyay, Samir & Dutta, Shawni. (2020). Fake Job Recruitment Detection Using Machine Learning Approach. International Journal of Engineering Trends and Technology. 68. 10.14445/22315381/IJETT-V68I4P209S