

Fraudulent Job Post Recognition

L. Sai Lakshmi¹, V. Blessy Joy Helen², R. Thanuja³, SK. Noushin⁴

M. Tech, Computer Science & Engineering, Bapatla Women's Engineering College, Bapatla, INDIA¹

B. Tech, Computer Science & Engineering, Bapatla Women's Engineering College, Bapatla, INDIA²⁻⁴

Abstract: Modern technology has advanced to the point that employing staff through an online process is now possible for businesses. This enables businesses to hire workers for necessary positions more quickly and immediately. It will also be reasonably priced. One can quickly find a job that fits their skills and field of interest by searching the internet. People may not be aware that the jobs that are posted are false or authentic. We developed new technologies to forecast job posts and determine if they are legitimate or fraudulent in order to solve these kinds of issues. We are creating a fake job system. Utilising the idea of machine learning for post-prediction, we are employing the Random Forest classifier, which generates precise outcomes quickly. Comparing the designed algorithm to the previously utilised algorithms, the result is 98%. When students or users look for work, they can have trouble spotting phoney job postings and applying, inadvertently providing all of their personal information. In certain instances, people could fall victim to scams that include paying money in the form of application fees in order to obtain employment or receiving a guarantee of employment upon payment. The framework assists us in determining if the jobs listed are fraudulent or not.

Keywords: Fraudulent Job Post, support vector machine, Random Forest Classifier, Machine Learning, NLTK, Hyperparameter Tuning.

I. INTRODUCTION

Getting a job these days is hard. You must apply for a job, register, and then proceed to the interview before attending any interviews. The first and most important step is to apply for a job in accordance with the company's specifications and the field in which the user want to work. Several job posts may appear when you search the internet; these job postings could be fake or real. It could be difficult for the user to determine if the job posting is authentic or fraudulent. Consequently, in order to help many individuals avoid disclosing their sensitive data to anyone by being aware of the bogus job marketing purposes, we must develop software that can differentiate from genuine and phoney jobs.

In an effort to streamline and expedite the hiring process, the companies offer job details. Various data mining approaches are being employed to address the issue of fraudulent job postings. When the Random Forest Classifier is employed, it performs better than the previous method in recognising fake job ads. This aids them in Preventing financial losses: you might be asked to pay an application fee in order to register, or you might be asked to pay money in various ways as part of the hiring procedure, amongst additional factors. Every organisation implements an online hiring technique. After releasing the job details, they post the job details and bring in the student or user whose profile fits the work particulars. People's need for work may lead them to recklessly explore the internet trust nobody and provide their personal information to any phoney job postings that could lead to abuse, such as banking information, etc. When searching for jobs, job seekers ought to utilise caution because they might fall victim to fraudulent individuals offering fictitious positions, which could be exploited for other purposes.

We are utilising a random forest classifier, which produces significantly better results than the techniques that were previously in use. The task that was finished offers enhanced outcomes in terms of expenditure, time, accuracy, and efficiency. As a result of frauds and scams that involve the exploitation of personal information and ruining a company's reputation, the online hiring process has grown progressively unsafe.

II. LITERATURE SURVEY

Among the literature reviews are the following: Vidros et al. [1] significantly assisted in accurately recognising online process frauds. Online hiring scammers employ a technique called Random Forest Classifier. Online hiring frauds are not the same as electronic scams. Random Forest Classifier is used for detection and classification, and Support Vector Machines (SVM) are utilised for feature selection. The hundreds of data in the EMSCAD dataset, which is publicly available, were used by Alghamdi and Alharby, et al. [2]. We ended up with a 97.41% rate. The two primary websites of attention are a corporation's corporate logo and several additional significant characteristics.

In their proposed model, Tin Van Huynh et al. [3] stated that one has to consider an employee's knowledge and abilities into account before hiring someone. Corporations ought to select an individual or student who fits the job specification. Using pretrained data, we apply several kinds of neural networks, like Text CNN, BI-GRU-LSTM, and others as well. An effective generation of 72.71 percent of the f1-score is expected through this. In the research of Jiawei Zhang et al. [4], there is currently an ongoing increase in online social networking in terms of both governmental and economic development. Users might have been deceived by the counterfeits news articles. Being able to determine whether news is fictitious or not is vital. We employ algorithms based on machine learning to explore the origins of the news producers and the topics they decided on from social media platforms in order to deal with the underlying issue of fake news. Our objective is to deliver high-calibre news. et al., Thin Van Dang [5].

Virtual neurons are created using DNN, and the default weight values are random numbers. After multiplying the weight by the input, the outcomes lie between the range of 0 and 1. Weights are adjusted during training for the purpose to sort the outcomes into different categories. The less accomplished patterns are the result of an overfitting issue with a few more layers. The model uses dense layers for data training. By reducing the number of layers for the few parameters that need to be trained, a generic model can be generated. The Adam is the optimizer, and the relu is the activation function. Adam assesses each trainee's rate of learning according to particular constraints as part of the technique for training.

According to P. Wang et al. [6], the model's tenets are the building blocks of neural networks that behave similarly to a human brain. This makes it possible for a computer to compare two patterns and determine whether they are similar or distinct. A neuron is the structure that has certain properties and group categories. A neural network is an arrangement of many nodes arranged in several layers. Jihadists [7] claim that perceptron's are linked to one another and arranged in layers. By adjusting the input layer weight through hidden layers, the error rate can be lowered. Changing the weight of the input layers through hidden layers reduces the calculation error rate. By adjusting the input layer weight through hidden layers, the error rate can be lowered.

III. METHODOLOGY

MACHINE LEARNING: A collection of computer algorithms known as "machine learning" may learn from examples and get better over time without needing to be explicitly coded by a programmer. Creating suggestions is a typical machine learning challenge. A variety of jobs also make use of machine learning.

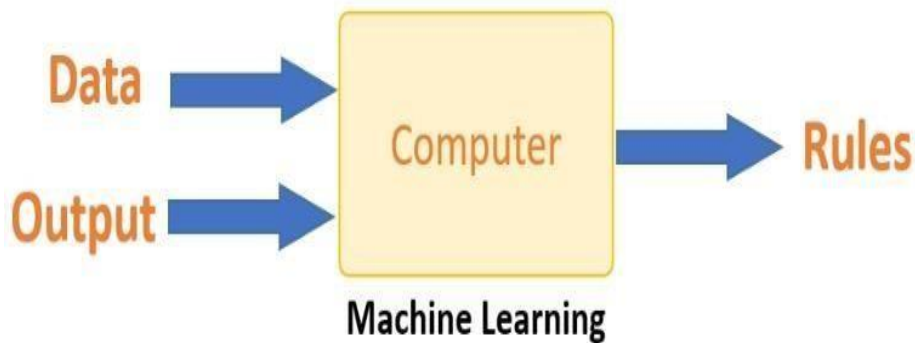


Figure 1: Machine Learning

1. **Random Forest Classifier:** The term refers to a collection of decision tree classifiers. Based on the voting process, we obtain the results based on the majority. Here are the steps to follow:

1. Choose a random sample from the provided dataset.
2. For each sample that is there, a decision tree is generated and each sample's prediction is produced.
3. Every forecast outcome has been submitted to a vote.
4. Select the outcome that is most likely to occur based on the number of responses.

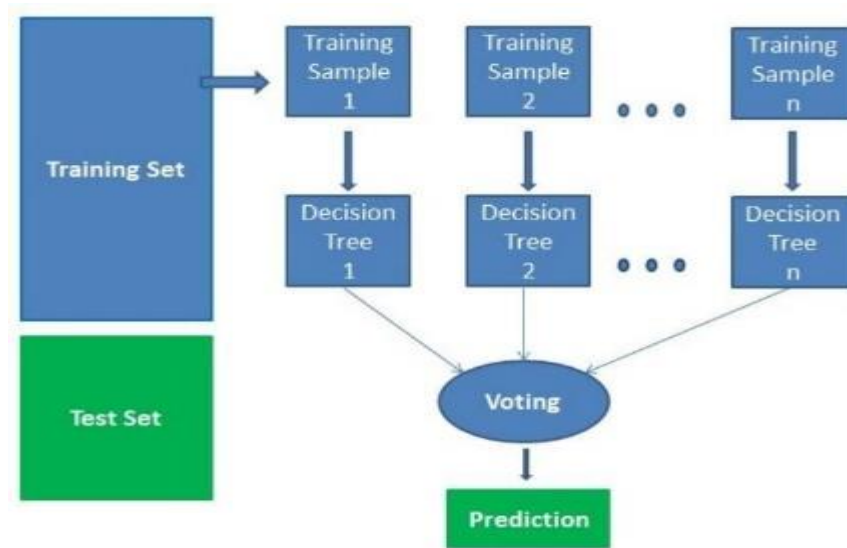


Figure 2: Random Forest Classifier

2. Hyperparameter Tuning: Hyperparameter tuning helps a lot to improve the accuracy of different models. Naive bayes model has better accuracy for fraudulent jobs and it classify more fraudulent jobs correctly, but it also classifies a lot of non-fraudulent jobs as fraudulent which is major drawback of naive bayes model for this project. The SVM model is very balanced model and it gives best roc_auc score. It is able to correctly classify most of the fraudulent jobs and non-fraudulent and the random forest model is good for classifying the non-fraudulent jobs. So, we can say that the svm model is the best fit for this problem.

3. Support Vector Machine:

Support Vector Machines (SVMs) are supervised learning models that can be used to solve tasks like classification and regression. They can solve both linear and nonlinear problems and are useful in a variety of situations. The concept behind Support Vector Machines is straightforward: In a classification problem, for example, the method draws a line between the classes. The line's purpose is to maximise the distance between points on either side of the so-called decision line. After the separation, the model may readily guess the target classes (labels) for new cases, which is a benefit of this procedure.

Construction of a Predictive Model

Machine learning needs data gathering have lot of past data ‘s. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can’t be used directly. It’s used to preprocess then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

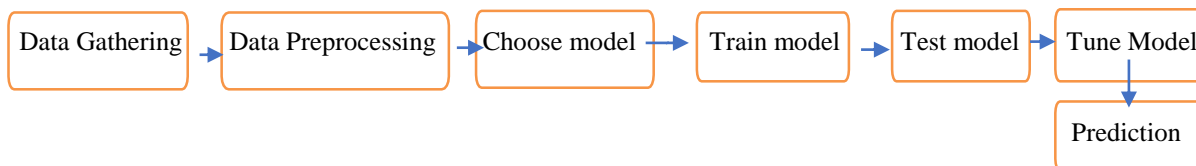


Figure 3: Process of dataflow diagram

IV. MODELLING AND ANALYSIS

The aim of the project is to identify fraudulent employment so that people don't fall for scammers. This guarantees that the information they give throughout the hiring process won't be exploited. Using several methods, we are trying to get better outcomes on an EMSCAD dataset. The pre-processed dataset for the fake job post has been gathered. The process of picking a few key features from the data that are needed for analysis and producing an appropriate result is known as feature selection. To determine if the job is real or fraudulent, we are using the Random Forest Classifier.

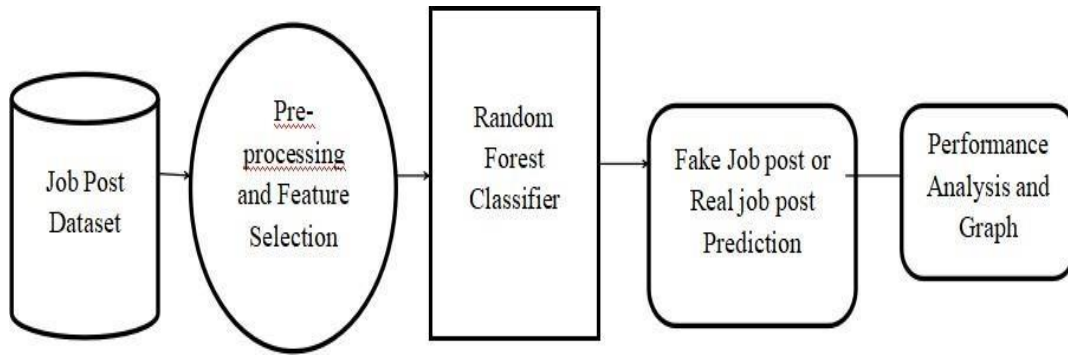


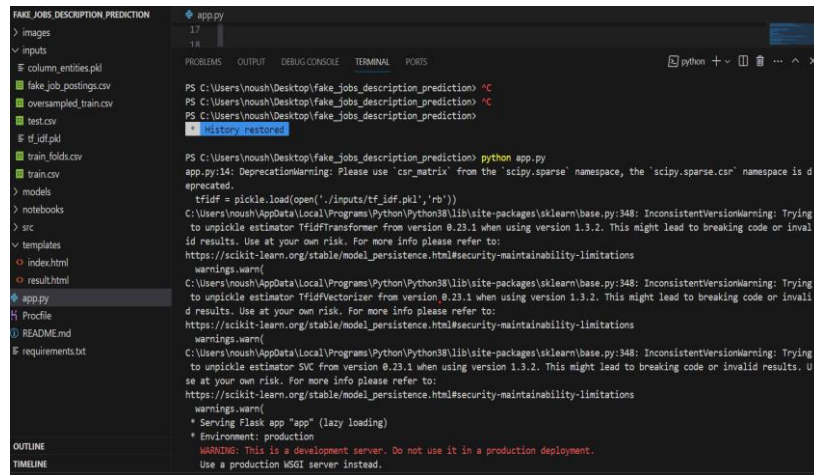
Figure 4: System Architecture

Certain tasks must be completed, such as passing the input and performing preprocessing, after which the data must be trained and the classifier applied. Predictions will come from it. The result will either be a fake or real job.

SNAPSHOTS:

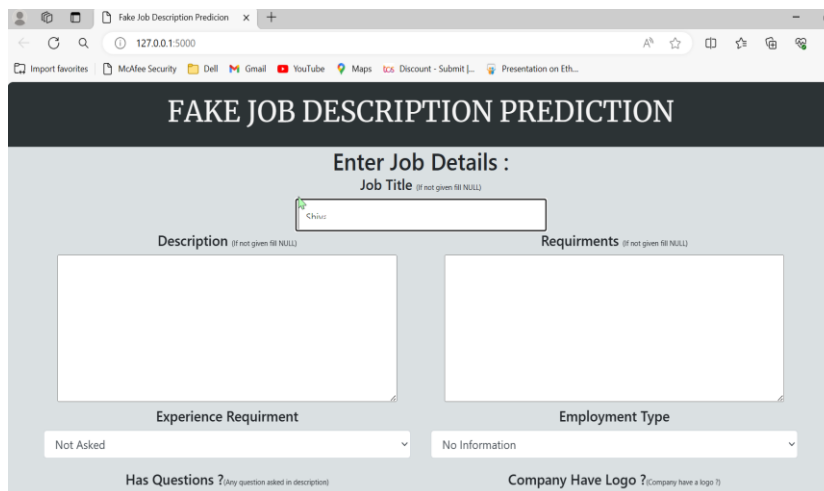
It is the result of fake job posting indicating that the details are fake about a job. The details you entered about a job like experience, qualifications, employment, company logo, etc. are false.

To get output we have to type python app.py in terminal



```
PS C:\Users\noush\Desktop\fake_jobs_description_prediction> python app.py
app.py:14: DeprecationWarning: Please use 'csr_matrix' from the 'scipy.sparse' namespace, the 'scipy.sparse.csr' namespace is deprecated.
  tfidf = pickle.load(open('./inputs/tf_idf.pkl', 'rb'))
C:\Users\noush\AppData\Local\Programs\Python\Python38\lib\site-packages\sklearn\base.py:348: InconsistentVersionWarning: Trying to unpickle estimator TfidfTransformer from version 0.23.1 when using version 1.3.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
C:\Users\noush\AppData\Local\Programs\Python\Python38\lib\site-packages\sklearn\base.py:348: InconsistentVersionWarning: Trying to unpickle estimator TfidfVectorizer from version 0.23.1 when using version 1.3.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
C:\Users\noush\AppData\Local\Programs\Python\Python38\lib\site-packages\sklearn\base.py:348: InconsistentVersionWarning: Trying to unpickle estimator SVC from version 0.23.1 when using version 1.3.2. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Serving Flask app "app" (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
```

By typing or clicking on the link(127.0.0.1:5000)the below page will be opened



FAKE JOB DESCRIPTION PREDICTION

Enter Job Details :

Job Title (if not given fill NULL)

Description (if not given fill NULL)

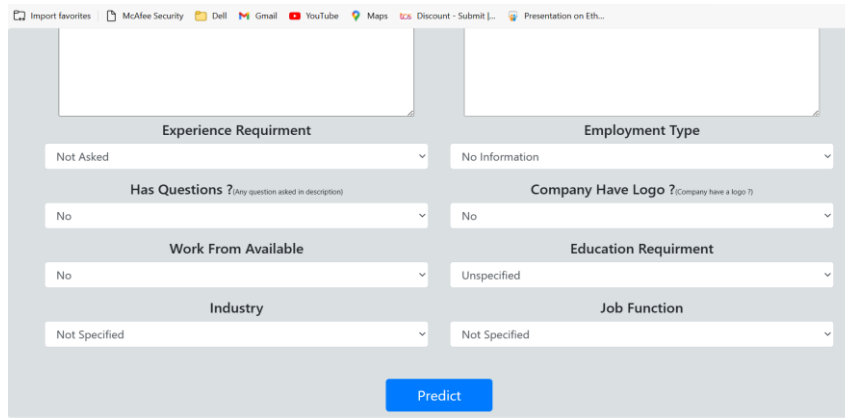
Requirements (if not given fill NULL)

Experience Requirement: Not Asked

Employment Type: No Information

Has Questions ? (Any question asked in description)

Company Have Logo ? (Company have a logo ?)

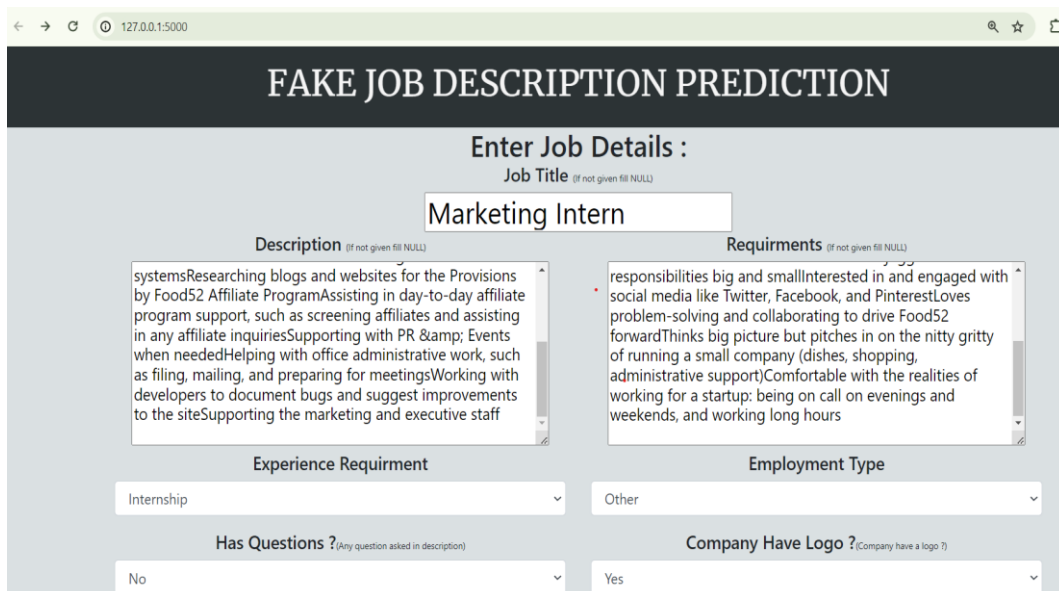


Import favorites McAfee Security Dell Gmail YouTube Maps tcs Discount - Submit ... Presentation on Eth...

Experience Requirement	Employment Type
Not Asked	No Information
Has Questions ? (Any question asked in description)	Company Have Logo ? (Company have a logo ?)
No	No
Work From Available	Education Requirement
No	Unspecified
Industry	Job Function
Not Specified	Not Specified

Predict

Fill the details below



127.0.0.1:5000

FAKE JOB DESCRIPTION PREDICTION

Enter Job Details :

Job Title (If not given fill NULL)
Marketing Intern

Description (If not given fill NULL)
systemsResearching blogs and websites for the Provisions by Food52 Affiliate ProgramAssisting in day-to-day affiliate program support, such as screening affiliates and assisting in any affiliate inquiriesSupporting with PR & Events when neededHelping with office administrative work, such as filing, mailing, and preparing for meetingsWorking with developers to document bugs and suggest improvements to the siteSupporting the marketing and executive staff

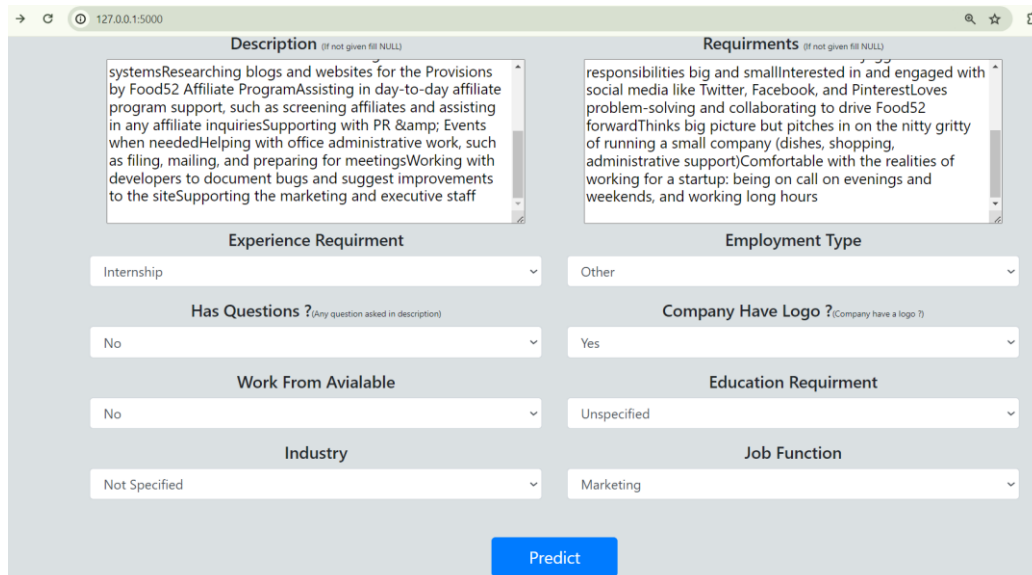
Requirments (If not given fill NULL)
responsibilities big and smallInterested in and engaged with social media like Twitter, Facebook, and PinterestLoves problem-solving and collaborating to drive Food52 forwardThinks big picture but pitches in on the nitty gritty of running a small company (dishes, shopping, administrative support)Comfortable with the realities of working for a startup: being on call on evenings and weekends, and working long hours

Experience Requirement: Internship

Employment Type: Other

Has Questions ? (Any question asked in description): No

Company Have Logo ? (Company have a logo ?): Yes



127.0.0.1:5000

Description (If not given fill NULL)
systemsResearching blogs and websites for the Provisions by Food52 Affiliate ProgramAssisting in day-to-day affiliate program support, such as screening affiliates and assisting in any affiliate inquiriesSupporting with PR & Events when neededHelping with office administrative work, such as filing, mailing, and preparing for meetingsWorking with developers to document bugs and suggest improvements to the siteSupporting the marketing and executive staff

Requirments (If not given fill NULL)
responsibilities big and smallInterested in and engaged with social media like Twitter, Facebook, and PinterestLoves problem-solving and collaborating to drive Food52 forwardThinks big picture but pitches in on the nitty gritty of running a small company (dishes, shopping, administrative support)Comfortable with the realities of working for a startup: being on call on evenings and weekends, and working long hours

Experience Requirement: Internship

Employment Type: Other

Has Questions ? (Any question asked in description): No

Company Have Logo ? (Company have a logo ?): Yes

Work From Avialable: No

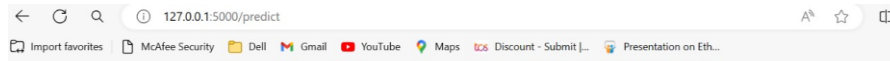
Education Requirement: Unspecified

Industry: Not Specified

Job Function: Marketing

Predict

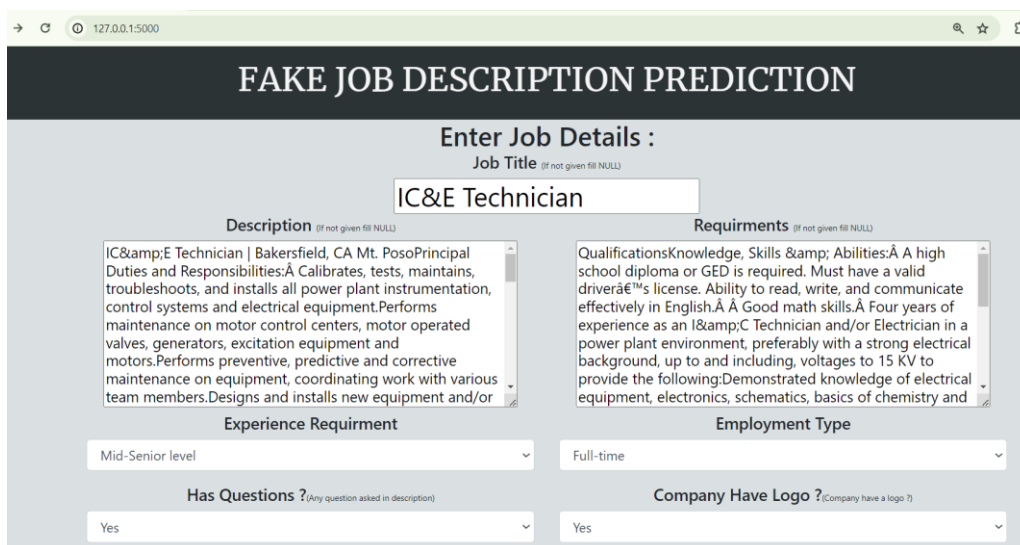
The output will be displayed as not a Fake Job description



Result :

Not a Fake Job Description

Fill the details below



FAKE JOB DESCRIPTION PREDICTION

Enter Job Details :

Job Title (if not given fill NULL)
IC&E Technician

Description (if not given fill NULL)
IC&E Technician | Bakersfield, CA Mt. Poso Principal Duties and Responsibilities: Calibrates, tests, maintains, troubleshoots, and installs all power plant instrumentation, control systems and electrical equipment. Performs maintenance on motor control centers, motor operated valves, generators, excitation equipment and motors. Performs preventive, predictive and corrective maintenance on equipment, coordinating work with various team members. Designs and installs new equipment and/or

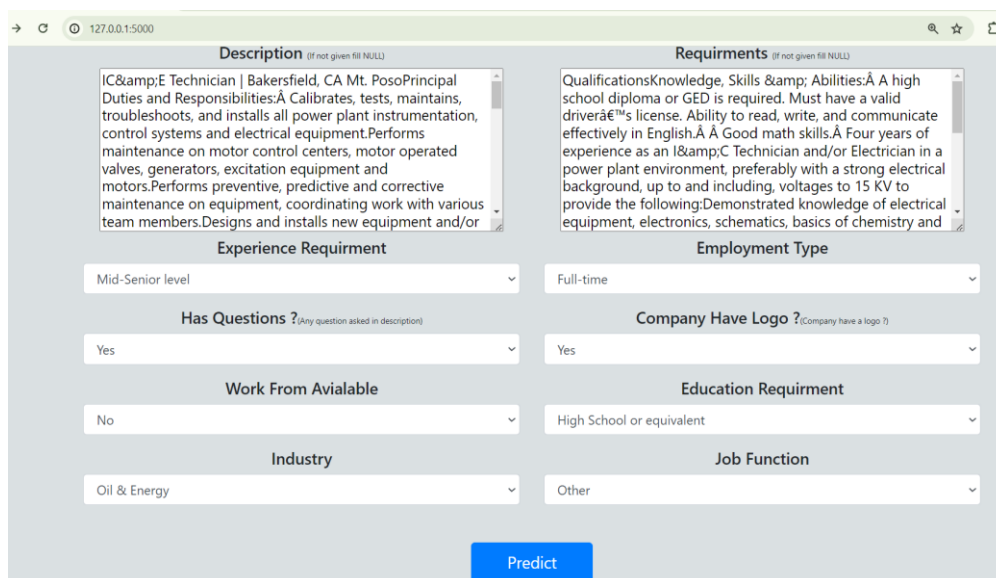
Requirements (if not given fill NULL)
Qualifications Knowledge, Skills & Abilities: A high school diploma or GED is required. Must have a valid driver's license. Ability to read, write, and communicate effectively in English. Good math skills. Four years of experience as an IC&E Technician and/or Electrician in a power plant environment, preferably with a strong electrical background, up to and including, voltages to 15 KV to provide the following: Demonstrated knowledge of electrical equipment, electronics, schematics, basics of chemistry and

Experience Requirement
Mid-Senior level

Employment Type
Full-time

Has Questions? (Any question asked in description)
Yes

Company Have Logo? (Company have a logo ?)
Yes



FAKE JOB DESCRIPTION PREDICTION

Enter Job Details :

Job Title (if not given fill NULL)
IC&E Technician

Description (if not given fill NULL)
IC&E Technician | Bakersfield, CA Mt. Poso Principal Duties and Responsibilities: Calibrates, tests, maintains, troubleshoots, and installs all power plant instrumentation, control systems and electrical equipment. Performs maintenance on motor control centers, motor operated valves, generators, excitation equipment and motors. Performs preventive, predictive and corrective maintenance on equipment, coordinating work with various team members. Designs and installs new equipment and/or

Requirements (if not given fill NULL)
Qualifications Knowledge, Skills & Abilities: A high school diploma or GED is required. Must have a valid driver's license. Ability to read, write, and communicate effectively in English. Good math skills. Four years of experience as an IC&E Technician and/or Electrician in a power plant environment, preferably with a strong electrical background, up to and including, voltages to 15 KV to provide the following: Demonstrated knowledge of electrical equipment, electronics, schematics, basics of chemistry and

Experience Requirement
Mid-Senior level

Employment Type
Full-time

Has Questions? (Any question asked in description)
Yes

Company Have Logo? (Company have a logo ?)
Yes

Work From Available
No

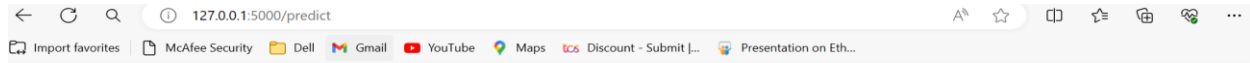
Education Requirement
High School or equivalent

Industry
Oil & Energy

Job Function
Other

Predict

The output will be displayed as Fake Job Description

**Result :****Fake Job Description**

V. CONCLUSION

Recently, detecting employment scams has grown to be a significant global issue. In this project, we have looked at the effects of employment scams because they may be a very profitable area of research and can make it challenging to spot false job postings. We used the EMSCAD dataset, which has up-to-date job listings.

Compared to previously utilized algorithms like SVM, Decision tree classifier, etc., which deliver 90 percent accuracy, the Random Forest Classifier provides 98 percent accuracy. By preventing fraud and scams in the employment market, we are making the online hiring process safer.

Consequently, you have the option to apply for employment online. Consequently, preventing a person's financial losses and protecting their personal data.

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 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)*, 2020T.
- [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]. T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A.G.-T. Nguyen, "Hate Speech Detection on Vietnamese Social Media Text using the Bi-GRU-LSTM-CNN Model," *arXiv Prepr. arXiv1911.03644*, 2019
- [6]. Thin Van Dang, Vu Duc Nguyen, Kiet Van Nguyen and Ngan Luu-Thuy Nguyen, "Deep learning for aspect detection on Vietnamese reviews" in *In Proceeding of the 2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, 2018, pp. 104-109.
- [7]. P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," *Neurocomputing*, vol. 174, pp. 806814, 2016.
- [8]. Scanlon, J.R. and Gerber, M.S., "Automatic Detection of Cyber Recruitment by Violent Extremists", *Security Informatics*, 3, 5, 2014, <https://doi.org/10.1186/s1338-014-0005-5>

BIOGRAPHY

Mrs. L. Sai Lakshmi working as Assistant Professor in Department of CSE, Bapatla Women's Engineering College, Bapatla. Pursuing Ph.D. in Bhartiya University, Ananthapur. Completed M. Tech (CSE) in Prakasam Engineering College, Kandukur. 7+ years' experience (6 years in teaching & 1 year in industry) .



V. Blessy Joy Helen B. Tech with Specialization of Computer Science and Engineering in Bapatla Women's Engineering college, Bapatla.



R. Thanuja B. Tech with Specialization of Computer Science and Engineering in Bapatla Women's Engineering college, Bapatla.



SK. Noushin B. Tech with Specialization of Computer Science and Engineering in Bapatla Women's Engineering college, Bapatla.