FAKE JOB LISTING DETECTION SYSTEM

[1]Esha Ajay Papat, [2]Aryan Gosavi, [3]Sakshi Mohite, [4]Shreyal Sundarkar, [5]Prof. Sushma Bhosle

[1]Final Year BE Student, [2]Final Year BE Student, [3]Final Year BE Student, [4]Final Year BE Student, [5]Assistant Professor

[1,2,3,4]Department Of Information Technology, [5]Department of Electronics and Telecommunication

[1,2,3,4,5]PCET's Nutan Maharashtra Institute of Engineering and Technology, Talegaon Dabhade, Pune, India.

Abstract: The advancement in modern technology and social communication has been increased in recent times. Because of this, there are many job opportunities available online for job seekers which do include freshers as well. Sometimes, this job opportunities can be fraud. As many job seekers are freshers, so they don't have an idea about this fraud happening online. Hence, we propose a system which can help the job seekers to find out whether the job posting done online is fraud or a legit job posting. This system uses machine learning algorithm to train the detection model. The user has to insert the URL of the job in the search box, with the help of feature extraction, the model will classify the job posting as legit or fraud job posting.

Index Terms – **Job seekers, fraud, legit, machine learning, feature extraction.**

I. INTRODUCTION

Fake job news can be simply explained as a piece of article which is usually written for economic and personal gains. Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF). Fake job news is negative impact on individuals and society; Hence the detection of fake job news is becoming a bigger field of interest for data scientist. This fake job listing detection system will get a major boost and would be viral among the people, around the country and globe.

Fake Job Listing is one of the most recurring problems in the job sector, especially for freshers. Online Job Enrollment has gained popularity as job searches have gained more skills to find them on the internet. Extortionists may take advantage of this notion in order to get money from job searchers since they supply labor services in return for cash to individuals who are looking for work prospects. In order to identify fake posts, it is required to use a machine learning approach, which makes use of a variety of characterization calculations to do. As a result of this differentiation, consumers are alerted to the existence of fake job announcements, which are differentiation from the rest of the job announcements by the use of a website.

II. OBJECTIVE

1. The easy access and exponential growth of the information available on social media networks has made it intricate to distinguish between false and trye information.
2. To reduce the time gap.
3. The key purpose of this research is to protect individual from compromised privacy and loss of money through the fake job news.
4. To build a reliable model this helps to effectively detect fraud job news with highest accuracy.

III. LITERATURE SURVEY

Many researches have been done on this topic to find whether the job posting is legit or fraud. Some of the papers which we surveyed are given below.

Vuppala [1] et al. in this paper, one of the most serious issues addressed in the area of online recruitment frauds in recent years is employment fraud. These days, a lot of businesses like to post their vacant positions online so that job hunters may easily find them. However, this might just be a ruse used by con artists to get others to labor for them in exchange for money. False job postings are capable of harming the credibility of a legitimate organization. The discovery of bogus job postings has increased interest in creating an automated system for identifying fake jobs and notifying the proper authorities. A machine learning strategy that makes use of several categorization algorithms is employed to detect fake posts.

Pranay [2] et al. this research paper suggests that in order to eliminate misleading job listings on the internet a computerized apparatus is to be created that uses artificial intelligence-based organizing tactics. To check for false information on the internet, a variety of classifiers are utilized. The results of these classifiers are then analyzed in order to create the most effective business trick detection model that can be applied to the field of information security. This tool could be quite useful when shifting through a sea of genuine job adverts for bogus job postings. In order to identify fake job advertisements on the internet, two key types of classifiers are used: solitary classifiers and troupe classifiers, to name a few instances.

Chiraratanasopha [3] et al. in this paper, processing for term-frequency or patterns of terms is a common technique for text analysis and text-based classification. These characteristics, however might not be sufficient to distinguish between phony and real job postings. As a result, in this work, we presented a method to identify bogus job postings using a novel collection of attributes intended to mimic the actions of can artists who use false information. Missing details, hyperbole, and credibility were the characteristics. The feature were intended to be represented as a category and a readability score that could be generated automatically. A detection model for the detection of bogus job was trained using data from the EMSCAD dataset that had been modified in accordance with the predefined features.

Dutta [4] et al. proposed an automated tool using machine learning-based classification techniques to prevent fraudulent job postings on the internet paper. The results of various classifiers are compared in order to determine the best employment scam detection model. These classifiers are used to check fraudulent posts on the web. It assists in identifying fake job postings, two main categories of classifiers- single classifiers and ensemble classifiers- are taken into consideration. However, experimental findings show that ensemble classifiers are superior to single classifiers in their ability to detect fraud.

Habiba [5] et al. With the development of social media and modern technology, advertising new job openings has recently become a   very common problem in the modern world. Therefore, everyone will have a lot of reason to be concerned about fake job postings. Fake job posting prediction presents a variety of difficulties, just like many other classification tasks. In order to determine whether a job posting is legitimate or fraudulent, this paper proposed using various data mining techniques and classification algorithms like KNN, decision tress, support vector machine, naive bayes classifier, random forest classifier, multilayer perceptron, and deep neural network. 1800 samples from the Employment Scam Aegean Dataset (EMSCAD) were used in the experiment. For this classification task, a deep neural network classifier excels.

IV. PROBLEM STATEMENT

To detect Fake Job Listings across the internet and help people seeking jobs to differentiate between fake and real job listings. The target of this project is to detect whether a job post is fraudulent or not.
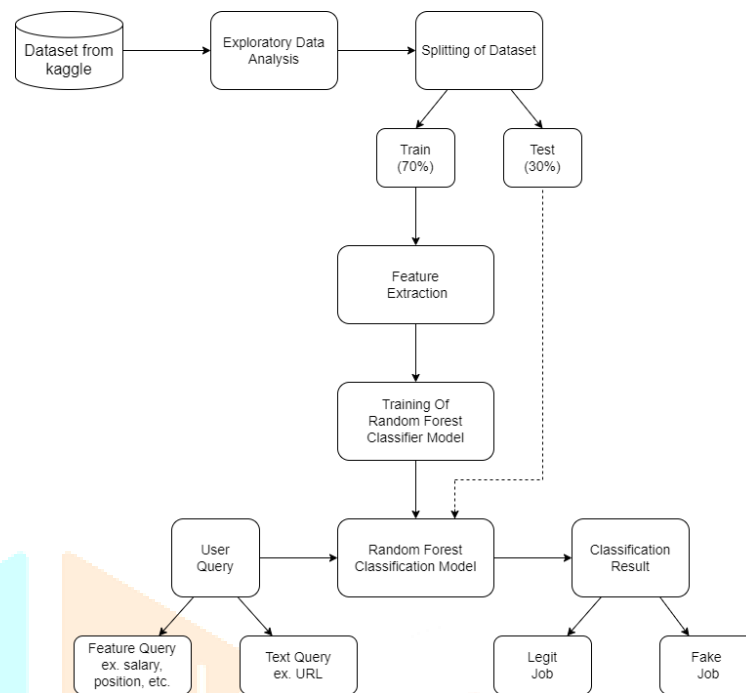
V. SYSTEM ANALYSIS

1.1 System Architecture

Fig 1. System Architecture

1.2 Proposed System

System Design also called top-level design signs aims to identify the modules that should be in the system, the specifications of these modules, and how they interact with each other to produce the desired results. During, Detailed Design, the internal logic of each of the module's specifications in system design is decided. During this phase, the details of the data are usually specified in a high-level design description language, which is independent of the target language in which the software will eventually be implemented.

In system design, the focus is on identifying the modules, whereas during detailed design the focus is on designing the logic for each of the modules. During the system design activities, Developers bridge the gap between the requirements specification, produced during requirements elicitation and analysis, and the system that is delivered to the user.

1. A data set is being used for the project Fake Job Listing Analysis.
2. The attributes in the data set are Company name, location, department, description, benefits, etc.
3. The welcome page of the Fake Job Listing user interface shows up.
4. In the text box, the job advertisement should be entered, and start the search to check whether the advertisement is genuine or not.
5. The final result window figure shows us whether the job advertisement is genuine or not.
6. The above window figure shows us that the job advertisement entered in the text box is not genuine and that it is a fake advertisement and gives a warning to the user about the advertisement.
7. It also gives a message to make awareness of these types of job advertisements.

VI. ADVANTAGES

1. Guides job-seekers to get only legitimate offers from companies or HR teams.
2. Provides users an automated system which classifies false or fake jobs and gives user an effective data.
3. Reduction of Cybercrime and unprotected techniques.
4. Saves time and money of young students as well as human resource management team.
5. Provides a secured way to share information only to best and secured companies.
6. Preserve security and consistency of their personal, academic and professional information. Provide a great advancement for recruiting new employees.

VII. DISADVANTAGES

1. For the training of model, huge amount of data is required, so that the model can be trained in an effective way.

VIII. APPLICATIONS

1. Can be used in applications like LinkedIn, Naukri.com, Job post, etc. to provide job seeker an effective way to apply jobs.
2. Verification of Job post and documents can also be done with the help of cloning.

IX. SUMMARY POINTS

1. We can use multiple machine learning algorithm to build this detection system but by doing literature survey of 5 research papers random forest algorithm gives the highest accuracy as compared to other algorithms.
2. We are going to use random forest algorithm so based on random parameters the model is trained and it is used to find out if the job posting is fake or not.

X. ACKNOWLEDGMENT

We are profoundly grateful to Prof. Sushma Bhosle, for her expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. We would like to express deepest appreciation towards Dr. Vilas Deotare, Principal, Nutan Maharashtra Institute of Engineering and Technology, Dr. Nitin Dhawas, Head of Department of Information Technology and Prof. Dheeraj Patil, Project Coordinator whose invaluable guidance supported us in completing this project. At last, we must express our sincere heartfelt gratitude to all the staff members of Information Technology Department who helped me directly or indirectly during this course of work.

XI. REFERENCES

[1] Vemula Pranay, Konda Reddy Sunil, Palakollu Divya, "Fake Job Recruitment Detection Using Machine Learning".

[2] Dr. V. Biksham, G. Tarun, B. Harshath, "Fake Job Listing Analysis Domain: Machine Learning".

[3] Boonthida Chiraratanasopha and Thodsaporn Chayintr, "Detecting Fraud Job Recruitment Using Features Reflecting from Real-world Knowledge of Fraud".

[4] Shawni Dutta, Prof. Samir Kumar Bandyopadhay, "Fake Job Recruitment Detection Using Machine Learning Approach".

[5] Sultana Umme Habiba, Md. Khairul Islam, Farzana Tasnim, "A Comparative Study on Fake Job Post Prediction Using Different Data Mining Techniques."