



FAKE JOB POSTING DETECTION

¹Ch.Sathwika, ²P. Shirin, ³G.Meghana, ⁴A.Anokh, ⁵V.Prathima

¹Student, ²Student, ³Student, ⁴Student, ⁵Guide, Assistant Professor

¹Department of Computer Science and Engineering

¹Vidya Jyothi Institute of Technology, Aziznagar, Hyderabad – 500075, Telangana, India

Abstract - This project addresses the problem of fraudulent job postings, which is a significant concern in the realm of Online Recruitment Frauds (ORF). To tackle this issue, an automated system based on machine learning algorithms has been developed. Users provide details such as job title, location, requirements, description, benefits, company profile, and industry. Data mining techniques and classification algorithms like decision trees, support vector machines, naive Bayes classifiers, and random forest classifiers, are employed for this purpose. This project provides a practical solution to the problem of fraudulent job postings, contributing to a safer online job-seeking environment.

Keywords - Fraudulent job vacancy postings, Online recruitment Frauds (ORF), Job postings, Employment Scam Detection, Machine Learning Approaches for Job Fraud Detection.

I. INTRODUCTION

Through job portals and recruitment sites, individuals explore potential job opportunities based on their qualifications, experience, and preferences. Job scams involve offering enticing job offers to candidates, only to deceive them and obtain their money or personal information. Scammers often utilize fake websites and emails to lure job seekers, targeting platforms like LinkedIn to pose as legitimate recruiters. About 56% of jobs in India face scams during their job hunt process. Job scams rose 250 percent in the first quarter of 2023, nearly triple compared to the same period in 2022.

SYSTEM ARCHITECTURE

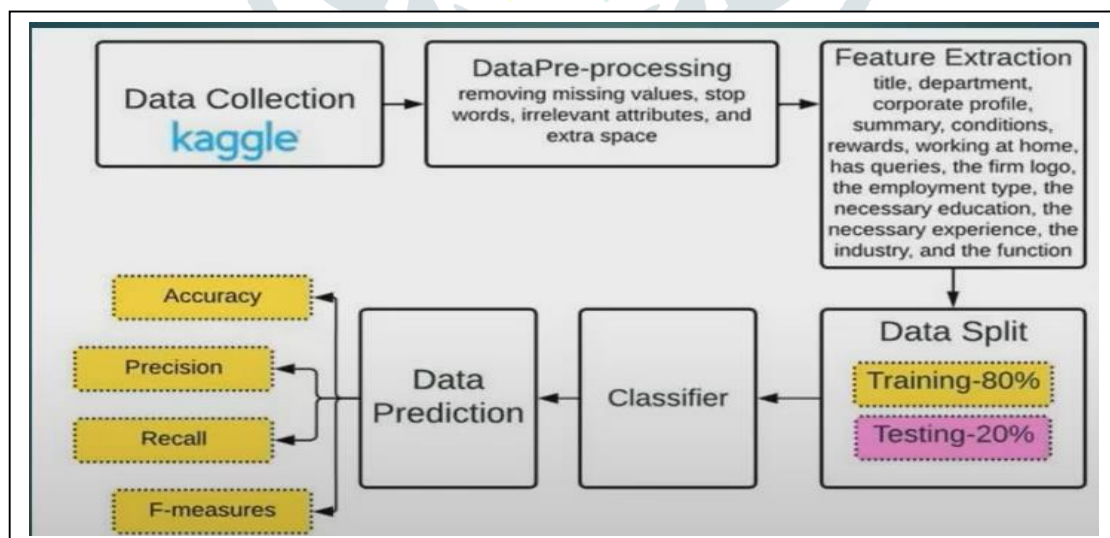


Fig:1 System Architecture

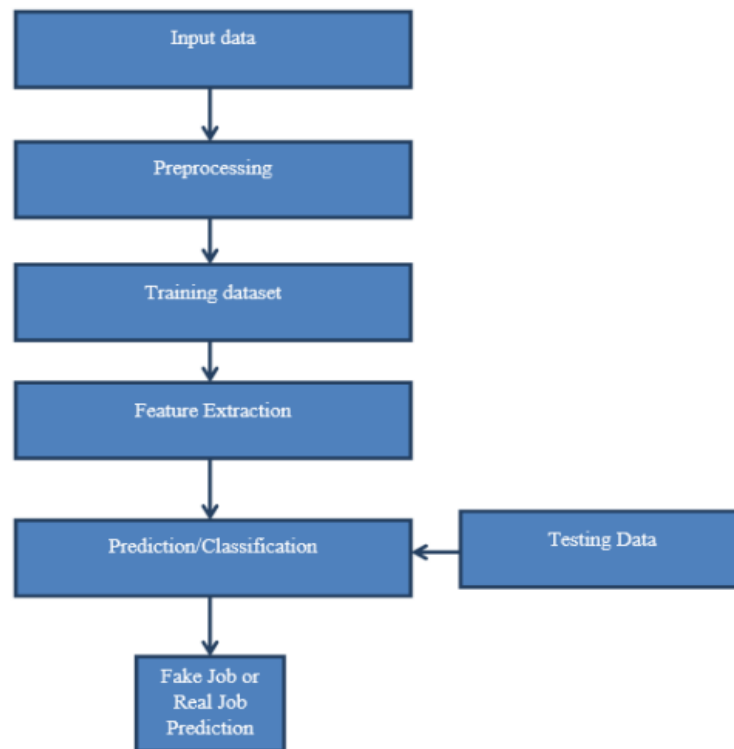


Fig:2 Methodology

II. SYSTEM ANALYSIS AND EXISTING SYSTEM

There exists a large body of research on machine learning methods for deception detection, but most of it is focused on classifying online reviews and news. Presently, there isn't a specific widely known website dedicated solely to detecting fake job listings. ORF detection has not received much attention and is currently a relatively unexplored area. The challenges of the existing system are that it is time – consuming, so job seekers can currently identify fake postings by verifying the contact details, searching the company's name, etc. This takes a lot of time and patience and it is not foolproof as scammers are evolving in their schemes and tactics to present a fake job as a real one, so it is not easy to identify a job posting. Another aspect is the dynamic nature of job Scams, wherein job scams are constantly evolving, requiring continuous updates and adaptations to detection strategies.

III. PROPOSED SYSTEM

This system uses ML algorithms and NLP approaches. A dataset that contains job posts, including attributes such as job id, title, location, and department is used to train the system. The data is cleaned and provided to the classifier for predictions to make it prediction ready. The classifier gets trained regularly as new training data is fed into classifier. A website is provided for administrator and users, allowing them to login and use the tool. The classifier gets trained regularly as new training data is feed into the classifier (keyword analysis, pattern recognition). A website (UI) is provided for administrators and users, allowing them to login, use the tool and see the prediction.

Features:

- User-Friendly Interface:** Developed a website using HTML and CSS web-based interface using a framework like Django. Users can easily register, login, post their job datasets, predict a job posting, view their profile and logout.
- Free:** This tool is absolutely free to use and has an easy to interact UI that non-technical users can use.
- Multiple classifiers:** It uses multiple algorithms like RF, NB, SVM and LR to compare their accuracies and to predict effectively.
- Database Integration:** Stores logs of user details like their names, emails, passwords, city, state and country in a database for persistence.

Advantages of the Proposed System:

- Improved accuracy:** Machine learning algorithms can be trained on large datasets to detect patterns that are indicative of fake job postings.
- Efficiency:** Machine learning models can process large volumes of data much more quickly and efficiently than humans.
- Scalability:** Machine learning models can be easily scaled up to handle large volumes of job postings.
- Adaptability:** Machine learning models can adapt to new fraud techniques and patterns as they emerge, making them more effective at identifying fake job postings.
- Cost-effective:** More cost-effective than employing a team of human moderators.
-

IV. IMPLEMENTATION MODULE DESCRIPTION

Random forest:

This consists of many decision trees. The random forest takes the prediction from each tree and based on the majority votes of predictions it predicts the final output. The ‘forest’ generated is trained through bagging or bootstrap aggregating. The steps are:

- I. Select a random sample from the dataset.
- II. A decision tree is constructed for every sample and produces a result of prediction for each sample.
- III. Each prediction result has been voted.
- IV. Choose the predicted result with the highest votes.

Random Forest is a popular machine learning algorithm used for both classification and regression tasks. It belongs to the ensemble learning techniques, which combine multiple individual models to improve predictive performance.

Naïve-Bayes:

This classification technique is based on Bayes theorem of conditional probability. Its primary use is text classification a huge training set. It assumes that the presence of a particular feature on a class (words) independent of the presence of any other feature, given the class label. It is used for binary classification like spam/not spam, positive/negative, yes/no. The Naive Bayes algorithm is a simple yet powerful classification algorithm used in machine learning. It's based on Bayes' theorem, which describes the probability of an event occurring given prior knowledge of conditions related to the event.

Logistic Regression:

This studies the association between a categorical dependent variable and a set of independent variables. It is used when the dependent variable has only two values, such as 0 and 1. It is specifically designed for predicting binary outcomes and has the highest accuracy in this case—98%. Logistic regression is widely used in various domains, including finance, healthcare, and marketing, for tasks such as churn prediction, credit scoring, and disease diagnosis.

Support Vector Machine (SVM):

This is powerful supervised learning algorithm used for classification and regression tasks. It is mainly a binary classification algorithm. It performs well with high dimensional data, making it suitable for text classification tasks, which helps in reducing overfitting and improving generalization performance, especially in cases where the dataset is limited or imbalanced. It's particularly effective in high-dimensional spaces and is widely used in various applications such as image classification, text classification, and bioinformatics.

V. RESULTS AND DISCUSSION

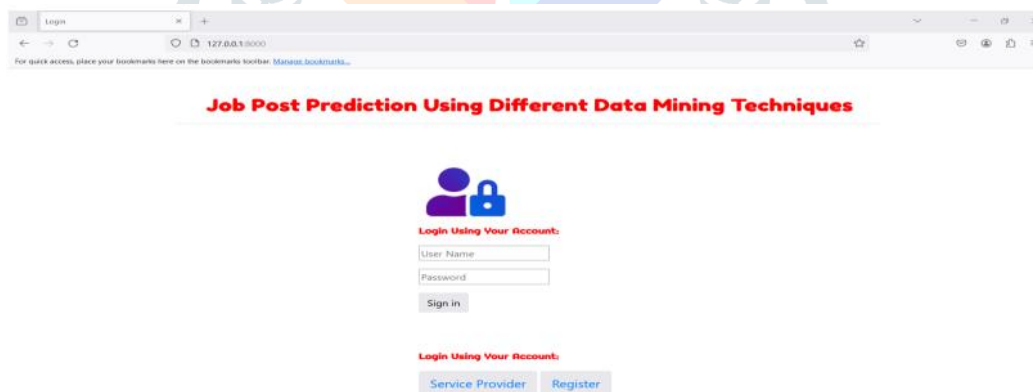


Fig: 3 Homepage/Login screen

1. The user can log in using their username and password.
2. If the user is new, they can register their details like their name, mobile, email, password, state, city, and country using the register option below. This successfully registers the user.
3. If the admin is trying to log in, they can click on the “Service Provider” option.

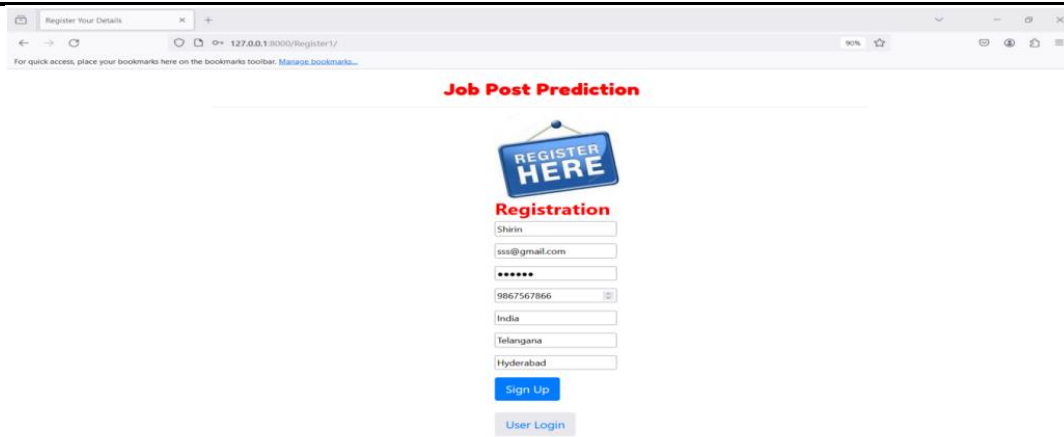


Fig: 4 New User Registration screen

The service provider refers to the admin. The admin can login to this page using his credentials.

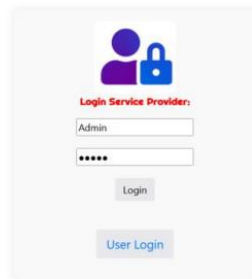
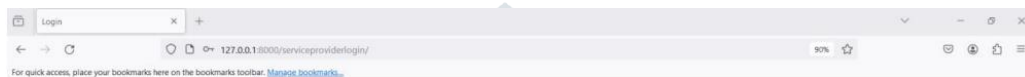


Fig: 5 Service Provider Login screen

OUTPUT:

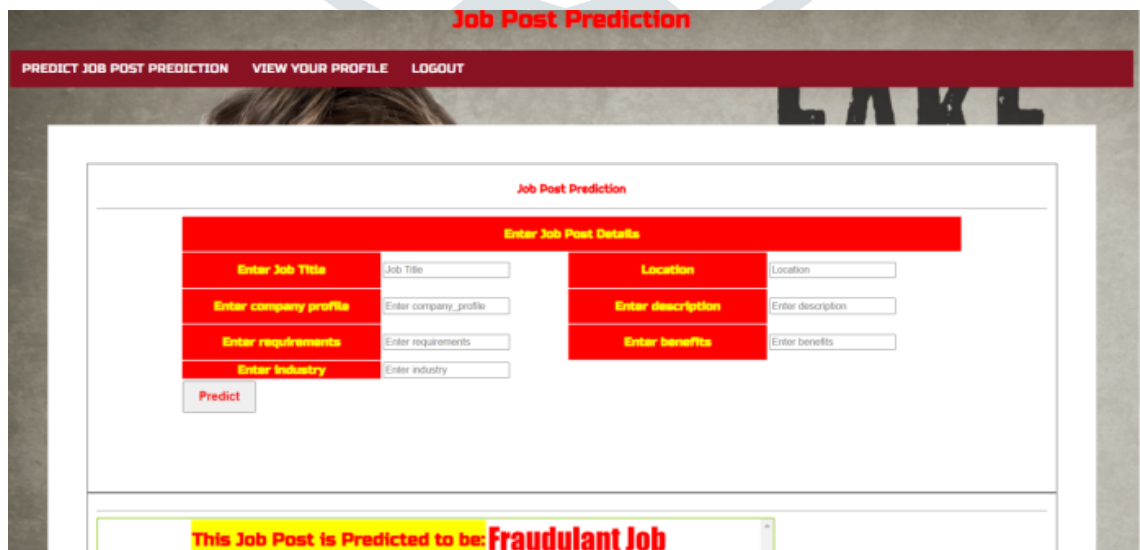


Fig:6 Real Job Post Prediction screen

Fig:7 Fraudulent Job Post Prediction screen

VI. CONCLUSION

Job scams have become a great concern all over the world at present. In this paper, we have analyzed the impacts of job scams which can be a very prosperous area in research field creating a lot of challenges to detect fraudulent job posts. The EMSCAD dataset which contains job posts from real situations, over the last ten years, has been experimented with. We have experimented with machine learning algorithms (SVM, Naïve Bayes, Random Forest and LR). This work shows a comparative study on the evaluation of traditional machine learning classifiers. Logistic Regression is seen to have the highest classification accuracy among the traditional machine learning algorithms. This is a versatile and practical tool in the fight against fraudulent jobs. It helps users avoid financial losses and protects their personal information. This will guide job-seekers to get only legitimate offers from companies. In essence, this tool project offers a valuable solution to the challenges posed by using online job recruitment portals, promoting a more efficient, secure, and focused digital environment for users across various contexts, whether fresher or senior.

REFERENCES

- [1] S Vidros, C. Kalias, 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, Alharby, "An Intelligent Model For Online Recruitment Fraud Detection" *Journal Of Information Security* 2019 Vol 10 Pp. 155176 <http://Doi.Org/10.4236/Iis.2019.103009>.
- [3] Scanlon J.R And Gerber MS "Automatic Detection Of Cyber Recruitment By Violent Extremists". *Security Informatics*. 3.4.2014, <http://Doi.Org/10.1186/S13388-014-0005-5>.
- [4] Y.Kim, "Conventional Neural Networks For Sentence Classification" *Arxiv Prepr. Arciv* 1408.5882, 2014.
- [5] I Rish, —An Empirical Study Of The Naïve Bayes Classifier An Empirical Study Of The Naïve Bayes Classifier, || No. January 2001, Pp. 41—46, 2014.
- [6] D. E. Walter, — Bayes's Theorem And The Analysis Of Binomial Random Variable, *Binometrical J.*, Vol 30, No. 7, Pp. 817 825, 1988, Doi:10.1002/Bimj.4710300710.
- [7] Tin Van Huynh, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Gia Tuan Nguyen, "Job Prediction: From Deep Neural Network Models to Applications", *RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020.