Fake Job Post Detection Using Machine Learning

K. Bindhu Sri[1], K. Durga Mounika[2], Ch. Anusha[3], E. Hima Sri[4] , B. Teja Sree[5]
Artificial Intelligence and Data Science, Department of Information Technology,
S.R.K.R Engineering College(A) [1,2,3,4]
Assistant Professor, Department of Information Technology,
S.R.K.R Engineering College(A), Bhimavaram, A.P[5]
kambhampatibindhusri@gmail.com[1], kasireddydurgamounika@gmail.com[2],
anushachunduri18@gmail.com[3], himasrieaday@gmail.com[4]

ABSTRACT: The rapid growth of online job platforms has revolutionized the job search process, offering both opportunities and challenges. One significant challenge is the presence of fake job postings, which not only waste job seekers time but also pose potential risks such as identity theft and financial fraud. This study proposes a predictive modeling approach leveraging machine learning techniques to identify fake job posts. We explore a dataset containing features extracted from job postings, including textual content, metadata, and contextual information. Various supervised learning algorithms, including decision trees, random forest, Stochastic gradient descent classifiers are employed to build predictive models.

The effectiveness of the proposed models is evaluated using standard performance metrics such as accuracy, precision and F1-score. Our experimental results demonstrate promising performance in detecting fake job posts, with our models achieving high accuracy and robustness across different evaluation metrics. The proposed approach offers a practical solution for online job platforms and recruitment agencies to automate the detection of fraudulent postings, thereby safeguarding job seekers and maintaining the integrity of the job market.

Keywords: Decision Trees, Random Forest, Stochastic Gradient Descent, Accuracy, F1-score

I. INTRODUCTION

Fake job post detection has emerged as a critical issue in recent times due to the surge in unemployment and the growing number of online job postings. Scammers take advantage of this situation by creating fraudulent job postings to extract personal information or demand payment from unsuspecting job seekers. To combat this issue, machine learning and natural language processing techniques can be used to detect and filter out these fake job postings. Machine learning algorithms play a crucial role in detecting fake job posts. This project focuses on the development of a reliable fake job post detection system, aiming to enhance trust and security in the online job market.

These algorithms classify job postings into two categories: real and fake. The classification process involves training a model on a labelled dataset of job postings, where each posting is marked as real or fake. Through the analysis of various features within job postings, including textual content, metadata, and user engagement patterns, we aim to identify key indicators of fraudulent activity. Once the model is trained, it can predict the

class of new, unseen job postings.The successful implementation of a fake job post detection system holds immense potential for reducing the effects of fraudulent activities in the job market. Not only will it protect job seekers from falling victim to scams, but it will also foster a more transparent and trustworthy environment for all stakeholders involved.

II. LITERATURE SURVEY

1. Title: "Detecting Fake Job Postings on the Internet" Authors: Ioannis Anagnostopoulos, Georgios K. Karamanis, Nicos Magoutis Published in: Data Mining and Knowledge Discovery, 2009 Summary: This study proposes a machine learning approach for detecting fake job postings online, utilizing textual and metadata features. The authors discuss the challenges of distinguishing between genuine and fraudulent job postings and evaluate various classification algorithms.

2. Title: "Automated Detection of Fake Job Advertisements" Authors: Himani Agrawal, Pushpendra Singh, Ruchi Agrawal Published in: International Journal of Engineering and Technology, 2019 Summary: This paper presents a framework for automated detection of fake job advertisements using machine learning techniques. The authors analyze textual features extracted from job postings and compare the performance of different classifiers.

3. Title: "Fraud Detection in Online Job Marketplaces using Machine Learning" Authors: Yash M. Pandya, Vishal R. Satish, Viraj S. Ghatge Published in: International Journal of Advanced Research in Computer Engineering & Technology, 2020 Summary: This research explores the application of machine learning algorithms for fraud detection in online job marketplaces. The authors investigate features such as job description, company profile, and location to identify fake job postings.

4. Title: "Detecting Fake Online Job Postings using Machine Learning Techniques" Authors: Shilpa Rathi, Priya Suryawanshi Published in: International Journal of Computer Applications, 2021 Summary: This study proposes a machine learning-based approach to detect fake online job postings. The authors analyze features extracted from job descriptions and evaluate the performance of classification algorithms.

5. Title: "Deep Learning for Detecting Fraudulent Job Advertisements" Authors: Zongruo Zhang, Chang Xu, Liang Wang Published in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019 Summary: This research investigates the use of deep learning techniques for detecting fraudulent job advertisements. The authors explore the effectiveness of deep neural networks in capturing complex patterns in textual data.

III. PROBLEM STATEMENT

The main purpose is to identify whether a job posting is genuine or not. This endeavour seeks to provide job seekers with a dependable tool to differentiate between genuine opportunities

and potential scams, thereby enhancing their security and confidence during the job search process. In this system, we plan to use a Kaggle dataset that contains information on the job, including attributes such as job id, title, location, department, etc.

The challenge lies in developing an automated system capable of accurately distinguishing genuine job posting from fraudulent ones, considering the diverse tactics used by scammers, including sophisticated language, counterfeit company affiliations, and misleading job descriptions. This problem necessitates a robust machine learning solution that can efficiently process large volumes of job postings, detect subtle indicators of fraud, and provide actionable insights to safeguard job seekers and uphold the integrity of online job markets.
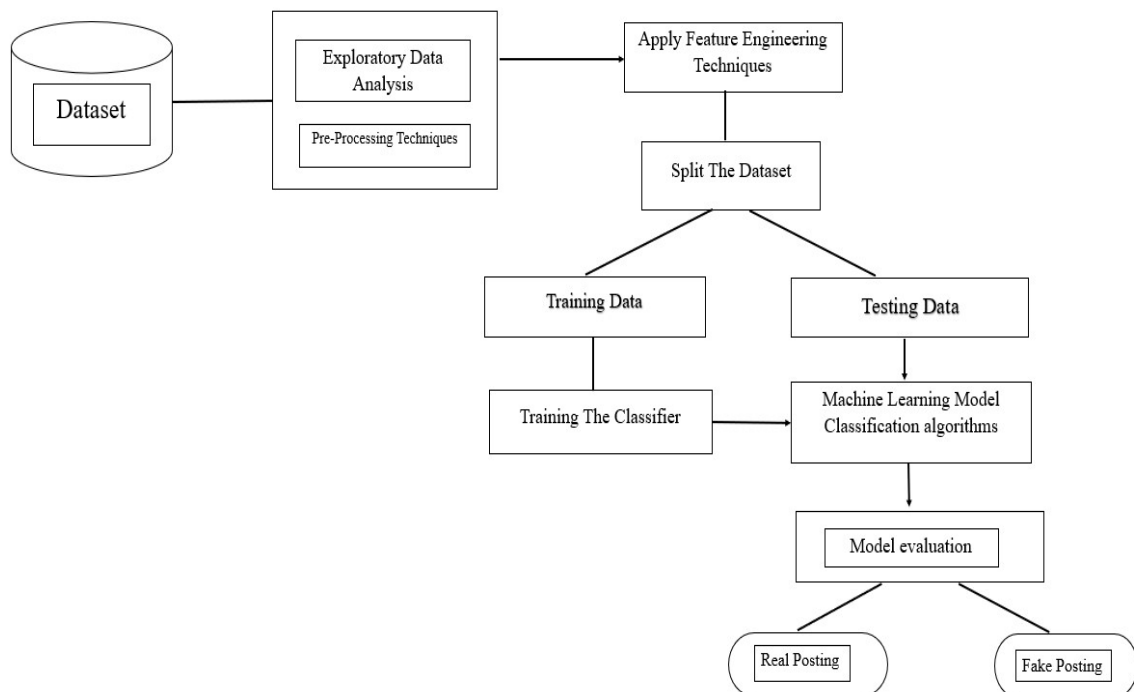
IV. ARCHITECTURE

Fig 1. System Architecture

ALGORITHMS:

1. **Random Forest:** The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement, called the bootstrap sample.

   Step-1: Select random K data points from the training set.

   Step-2: Build the decision trees associated with the selected data points (Subsets).

   Step-3: Choose the number N for decision trees that you want to build.

   Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

2. **SGD Classifier:** The class SGD Classifier implements a plain stochastic gradient descent learning routine which supports different loss functions and penalties for classification. SGD Classifier supports multi-class classification by combining multiple binary classifiers in a "one versus all" (OVA) scheme.

Step-1: Initialize the model parameters randomly or with some default values.

Step-2: Randomly shuffle the training data.

Step-3: Compute the gradient cost function with respect to the model parameters.

Step-4: Update the model parameters in the direction of negative gradient by a small step size known as learning rate.

Step-5: Repeat this process for a specified number of iterations(epochs).

V. IMPLEMENTATION

Input:
- Dataset of job posts with corresponding features
- Machine learning models (SGD Classifier, Random Forest, etc.)

Output:
- Trained model for fake job post detection

Steps:
1. Data Acquisition and Preparation:-
   i) Collect a diverse dataset- This involves gathering data relevant to the project's goal. The dataset contains 17,880 job posting data entries. The data consists of both textual information and meta-information about the jobs.
   ii) Preprocess the data rigorously- Raw data often needs cleaning and transformation. This might involve Handling missing values and Encoding categorical data

2. Feature Engineering :-
   i) Extract a rich set of informative features :
   Feature Selection- Choosing the most informative features can improve model performance and reduce training time. We can use techniques like correlation analysis or feature importance scores to identify the best features.

3. Model Selection and Training:-
   i) Address class imbalance- In some datasets, one class might have significantly more data points than others. This imbalance can affect model training. Techniques like oversampling (duplicating data points from the minority class) or undersampling (removing data points from the majority class) can help balance the data.

ii) Experiment with different classification algorithms-There are many classification algorithms, each with strengths and weaknesses. Some of them are Decision Tree, Random Forest, SGD Classifiers, etc…

4. Evaluation and Deployment:-

i) Use appropriate evaluation metrics- After training a model, you need to assess its performance. Different metrics are suitable for different tasks. For classification algorithms, accuracy, precision, recall, and F1-score are common choices.

ii) Flask- It is utilized under the deployment stage. It is a lightweight and versatile web framework for Python. It's designed to be simple, easy to use, and flexible, making it an excellent choice for building web applications, APIs, and prototypes.

VI. RESULTS AND DISCUSSION

```
1 print("Classification Accuracy:", accuracy_score(y_test, y_pred_class))
2 print("Classification Report\n")
3 print(classification_report(y_test, y_pred_class))
4 print("Confusion Matrix\n")
5 print(confusion_matrix(y_test, y_pred_class))
```

```
Classification Accuracy: 0.9057692307692308
Classification Report

              precision    recall  f1-score   support

           0       0.89      0.94      0.91       276
           1       0.93      0.86      0.90       244

    accuracy                           0.91       520
   macro avg       0.91      0.90      0.90       520
weighted avg       0.91      0.91      0.91       520


Confusion Matrix

[[260  16]
 [ 33 211]]
```
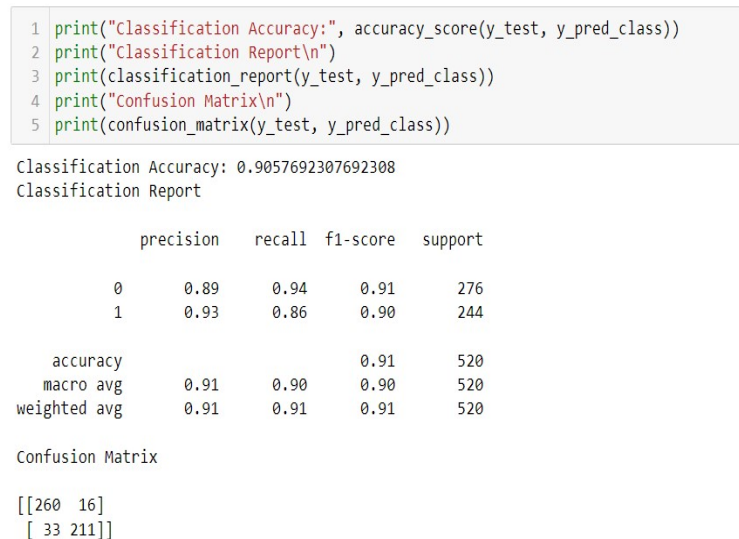
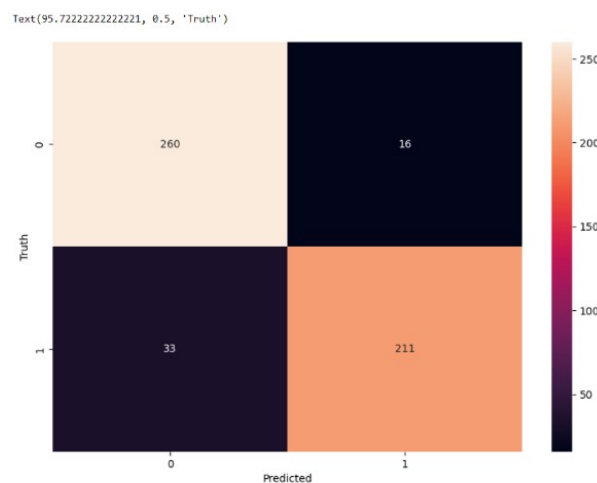Fig 2. Results of SGD classifier
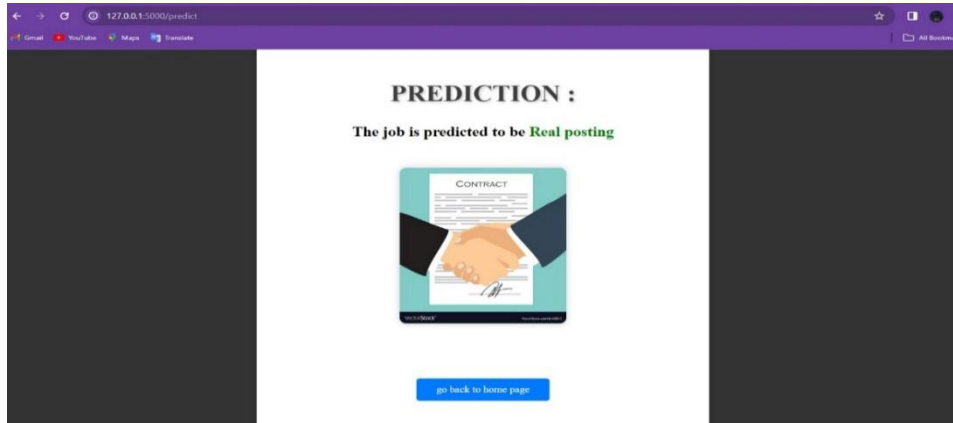

Fig 3. Confusion Matrix of SGD Classifier
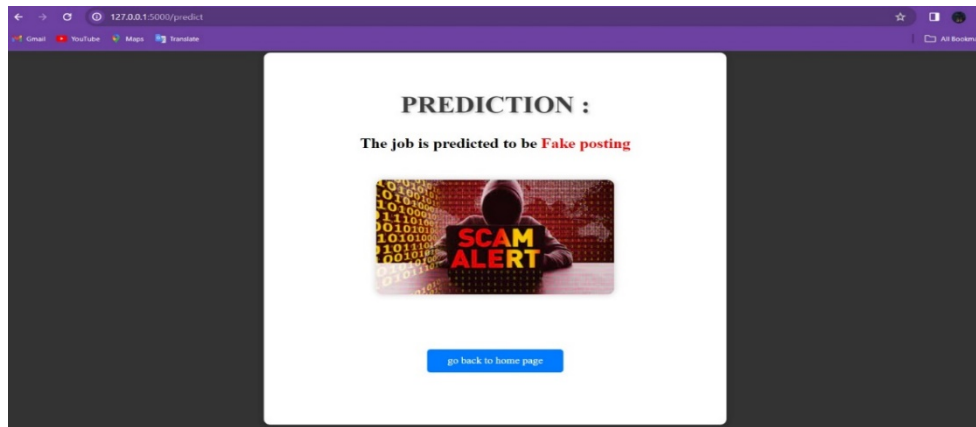
Fig 4. Detection of Real post

Fig 5. Detection of Fake Post

We had equipped both single classifier and ensemble classifiers in the training of classification process. And results the best performance with the ensemble classifier called stochastic gradient descent algorithm (SGD) based on accuracy, precision and f1 score. And deployed the model using flask.

VII. CONCLUSION

The project "Fake Job Post Detection Using Machine Learning" employing SGD and Random Forest classifiers has demonstrated promising results in accurately identifying fraudulent job postings. Through meticulous feature engineering and model fine-tuning, both classifiers have showcased their efficacy in distinguishing between legitimate and fake job advertisements. While SGD classifier offers efficiency in large-scale datasets, Random Forest classifier excels in capturing complex patterns and achieving high precision. The successful implementation of this project highlights the potential of machine learning in combating online job scams, thus safeguarding job seekers from fraudulent activities in the digital employment landscape.

VIII. REFERENCES

[1] k, swapna, A Machine Learning Approaches for Fake Job Identification (July 10, 2023). International Journal of Progressive Research in Engineering Management and Science, Vol. 03, Issue 07, July 2023, Pp : 149-154, Available at SSRN: https://ssrn.com/abstract=4706112

[2] T. Bhatia and J. Meena, "Detection of Fake Online Recruitment Using Machine Learning Techniques," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 300-304, doi: 10.1109/ICAC3N56670.2022.10074276.

[3] Papat, E. A., Gosavi, A., Mohite, S., Sundarkar, S., & Bhosle, S. (2023). "Fake Job Listing Detection System." International Journal of Computer Science and Information Technology Research, 11(5),Nutan Maharashtra Institute of Engineering and Technology, Talegaon Dabhade, Pune, India. ISSN: 2320-2882.

[4] Reddy, K. S. S., & Reddy, K. L. (2021). Fake Job Recruitment Detection. Journal of Emerging Technologies and Innovative Research (JETIR), 8(8), d443. SVKP & Dr K S Raju Arts & Science College, Penugonda, W.G.Dt., A.P, India. ISSN: 2349-5162.

[5] Dutta Shawani., & Bandyopadhyay, S. K. (2020). "Fake Job Recruitment Detection Using Machine Learning Approach." International Journal of Engineering Trends and Technology (IJETT), 68(4), 48. ISSN: 2231-5381.