IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 7, May 2023

Predicting Fraudulent Job Ads with Machine Learning

Hitesh Ahire[1], **Aashish Kumar Singh**[2], **Arpit Bhorkar**[3], **Shrushti Daware**[4], **Prof. P. A. Deole**[5]

Students, CSE, Shrimati. Kashibai Navale College of Engineering Vadgaon (BK), Pune[1,2,3,4]

Assistant Professor, CSE, Shrimati. Kashibai Navale College of Engineering Vadgaon (BK), Pune[5]

Abstract: *Online Recruitment frauds are becoming an important issue in cyber-crime region. Companies find it easier to hire people with the help of the internet rather than the old traditional way. But it has greatly attracted scammers. In this project, we have proposed a solution on how to detect ORF. We have presented our results based on the previous model and the methodologies, to create the ORF detection model where we have used Jobs_Frauds.csv. We have selected this dataset from Kaggle. Furthermore, Dummy Classifier, Random Forest Classifier, Support Vector Machine, Gradient Boosting, Naïve Bayes Classifiers, XG Boost, SGD classifier, Passive Aggressive and KNN are the algorithms that have been used. We have found the accuracy of different prediction models, where Passive Aggressive (98.12%) and Gaussian Naïve Bayes (96.72%) give the highest accuracy. Through this project, we tried to create a precise way for detecting fraudulent hiring posts.*

Keywords: Passive Aggressive, Naïve Bayes, SVM, Job Ads.

I. INTRODUCTION

Most modern organizations use the web and social media platforms for employee recruitment and job opening advertisements. Online job posts are easily accessed by interested job-seekers. Unfortunately, this trend creates an opportunity for criminals to exploit job seekers with fake job offers.

The Job advertisement and recruitment process has been improved by using online advertisements. These ads are being used by criminals as a means of committing fraud. A system that can automatically identify fake job ads using their attributes will reduce the chances of job seekers falling a victim to scams. This project aims to develop a machine learning classifier to flag fake and real job advertisements.

II. PROBLEM STATEMENT

To avoid fraudulent post for job in the internet, an automated tool using machine learning based classification techniques is proposed in the paper. Different classifiers are used for checking fraudulent post in the web and the results of those classifiers are compared for identifying the best employment scam detection model. It helps in detecting fake job posts from an enormous number of posts. Two major types of classifiers, such as gradient boost and naïve bayes classifiers are considered for fraudulent job posts detection. However, experimental results indicate that naïve bayes and gradient boost classifiers are the best classification to detect scams over other classifiers.

III. MOTIVATION

- Most modern organisations use the web and social media platforms for employee recruitment and job opening advertisements.
- Online job posts are easily accessed by interested job-seekers, hence its popularity.
- Unfortunately, this trend creates an opportunity for criminals to exploit job seekers with fake job offers.
- Criminals extract personal information to be used in nefarious activities.

IV. SYSTEM REQUIREMENTS

Hardware Requirements:

A. Processor: Intel i3 minimum or more

DOI: 10.48175/IJARSCT-10203

B. Disk Space: At least 3GB for Python IDE
C. RAM: 4GB minimum or more

Software Requirements:
A. Python IDE
B. Stremlit (online tool for Deployment)

V. METHODOLOGY

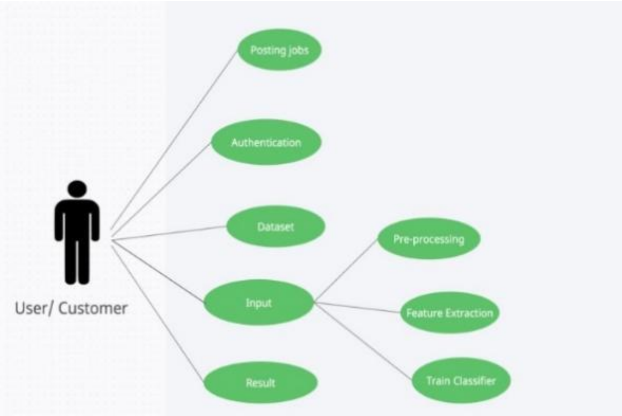
Module l - (Ads acquisition)

- We are going to use dataset from Kaggle for classification.
- Jobs will be taken from external source or online.
- Currently we are going to take a dataset around 17880 jobs froman open-source dataset platform(Kaggle).

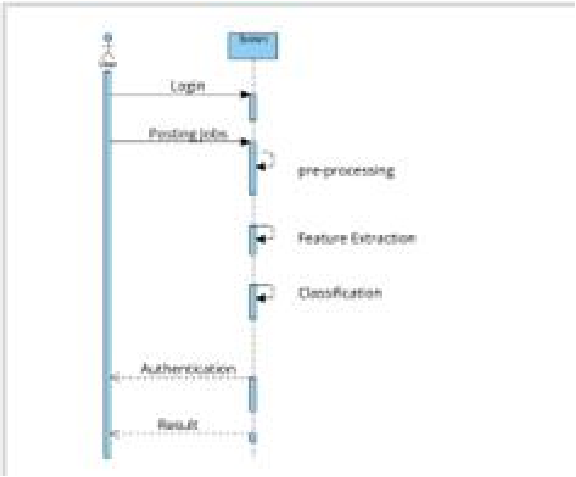
Module 2 - (Pre-processing)

- The dataset was made up of features with a lot of null values and somewith little or no null values. A 60% threshold for null was used to identify the columns that were dropped from the dataset.
- Null values were imputed with empty strings.

Use Case:



Sequence Diagram:

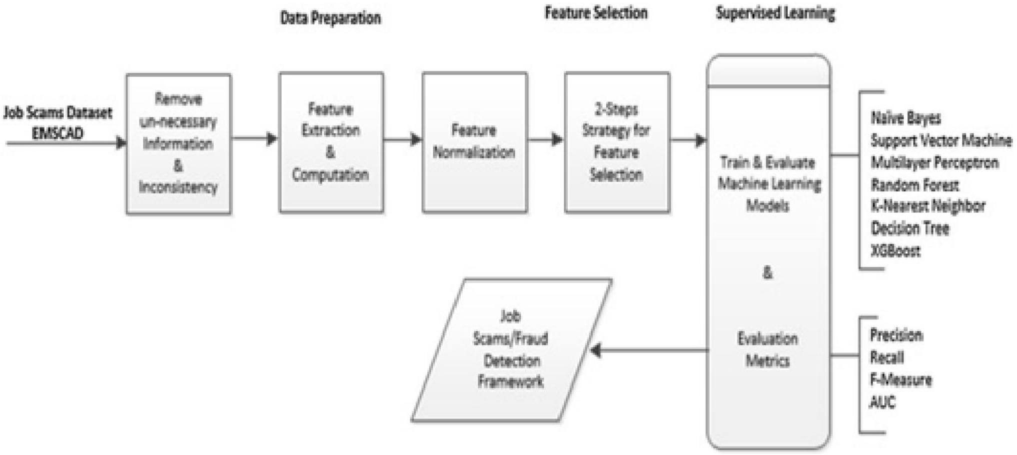


DOI: 10.48175/IJARSCT-10203

Proposed System:

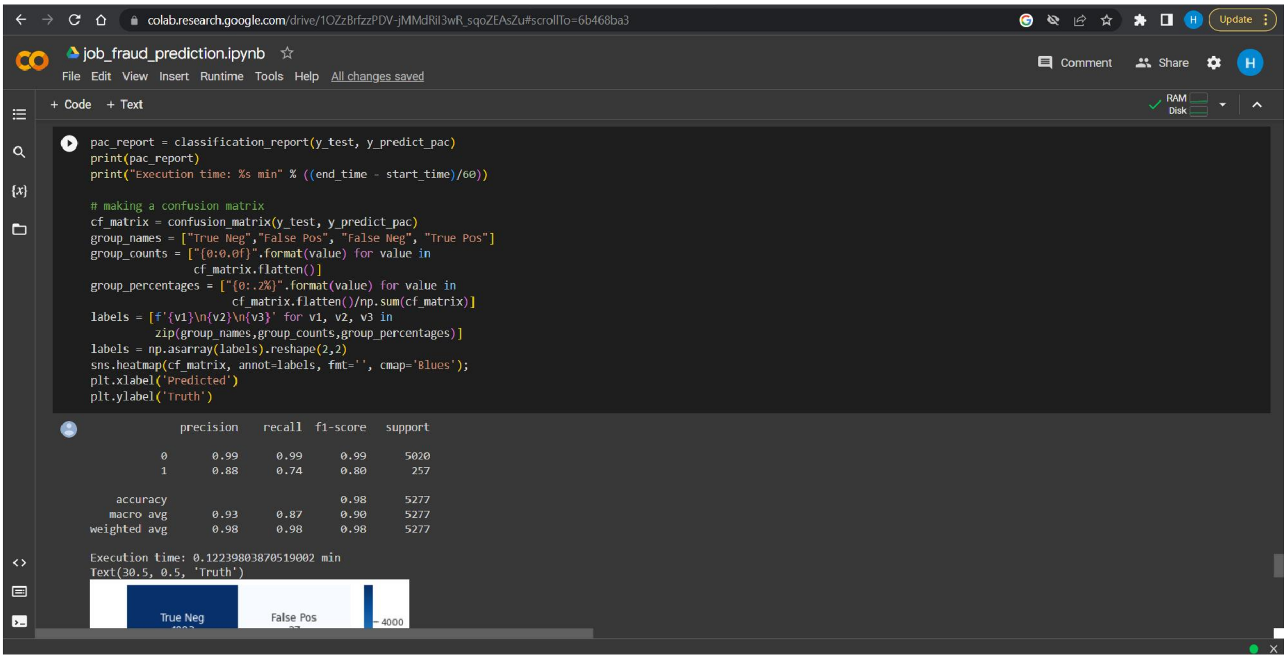
Algorithms used:

- Dummy Classifier
- Random Forest
- SVC
- Gaussian Naïve Bayes
- Passive Aggressive



VI. RESULT

Passive Aggressive Classifier Confusion Matrix:

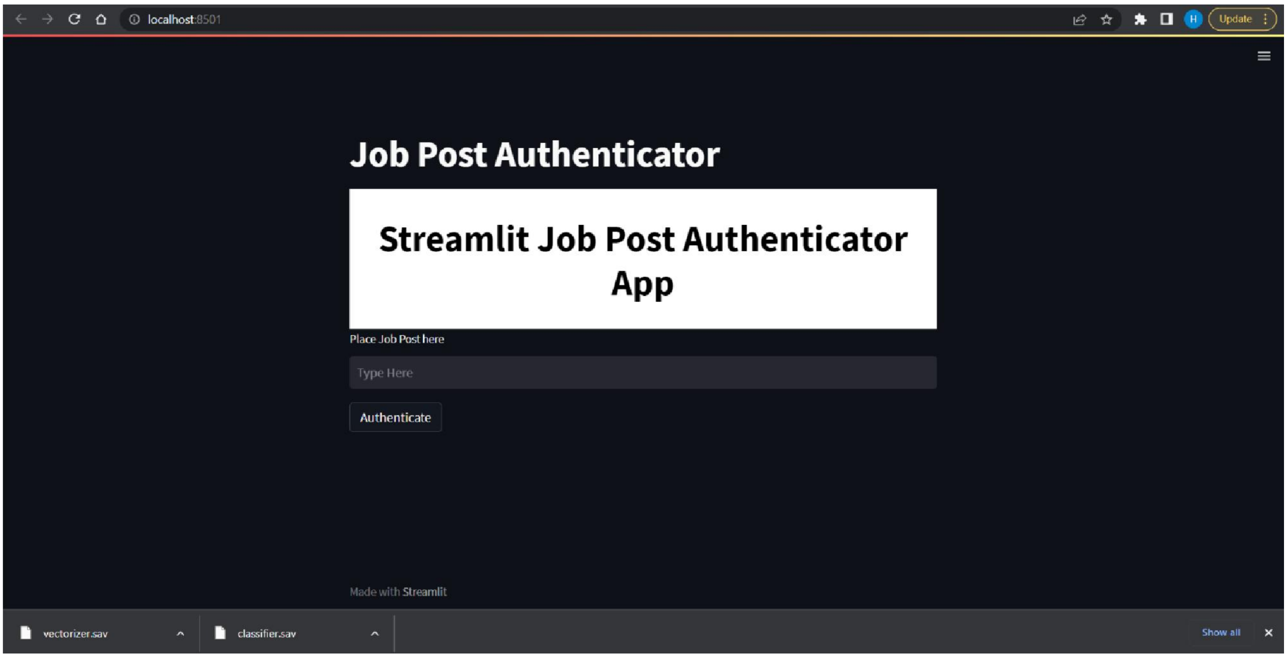

colab.research.google.com/drive/1OZzBrfzzPDV-jMMdRil3wR_sqoZEAsZu#scrollTo=6b468ba3
Update
job_fraud_prediction.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
True Neg
False Pos
4000

```
pac_report = classification_report(y_test, y_predict_pac)
print(pac_report)
print("Execution time: %s min" % ((end_time - start_time)/60))

# making a confusion matrix
cf_matrix = confusion_matrix(y_test, y_predict_pac)
group_names = ["True Neg","False Pos", "False Neg", "True Pos"]
group_counts = ["{0:0.0f}".format(value) for value in
                cf_matrix.flatten()]
group_percentages = ["{0:.2%}".format(value) for value in
                     cf_matrix.flatten()/np.sum(cf_matrix)]
labels = [f'{v1}\n{v2}\n{v3}' for v1, v2, v3 in
          zip(group_names,group_counts,group_percentages)]
labels = np.asarray(labels).reshape(2,2)
sns.heatmap(cf_matrix, annot=labels, fmt='', cmap='Blues');
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      5020
           1       0.88      0.74      0.80       257

    accuracy                           0.98      5277
   macro avg       0.93      0.87      0.90      5277
weighted avg       0.98      0.98      0.98      5277

Execution time: 0.12239803870519002 min
Text(30.5, 0.5, 'Truth')
```

DOI: 10.48175/IJARSCT-10203

Model Deployment:



Job Post Authenticator

Streamlit Job Post Authenticator App

Place Job Post here

Type Here

VII. CONCLUSION

In our research we have compared the performance of different machine learning algorithms like Dummy Classifier, Random Forest Classifier, Support Vector Machine, Gradient Boosting, Naïve Bayes Classifiers, XG Boost, SGD classifier, Passive Aggressive and KNN and determined which algorithm works best on our dataset. We did not only use the most common algorithms but also some latest ones so that we can determine how well do they work. We have seen that range of accuracy of all our algorithms lie between 96% to 98%. The main purpose for our model is to help the job seekers identify which jobs are fake and save themselves from fraudsters. Our model can also be used by different online job recruitment sites to detect fraudulent jobs. We have identified that some features play an important role in determining whether a job is fake or not. In future, we plan to incorporate these features in our dataset and thus try to further increase the accuracy range of the algorithms.

REFERENCES

[1] Ibrahim M. Nasser; Amjad H. Alzaanin; Ashraf Yunis Maghari, Online Recruitment Fraud Detection using ANN, 2021.

[2] Hridita Tabassum; Gitanjali Ghosh; Afra Atika; Amitabha Chakrabarty, Detecting Online Recruitment Fraud Using Machine Learning, 2021.

[3] Sangeeta Lal; Rishabh Jiaswal; Neetu Sardana; Ayushi Verma; Amanpreet Kaur; Rahul Mourya, ORFDetector: Ensemble Learning Based Online Recruitment Fraud Detection, 2019.

[4] Asad Mehboob & M. S. I. Malik, Smart Fraud Detection Framework for Job Recruitments, 2020.

[5] Bandar Alghamdi, Fahad Alharby, An Intelligent Model for Online Recruitment Fraud Detection, 2019.