Fake Job Recruitment Detection

ANUSHKA DHONDE-(A-04)
PRIYA KHANDAGALE-(A-12)
AKSHATA UTEKAR-(A-30)

CONTENT

Abstract

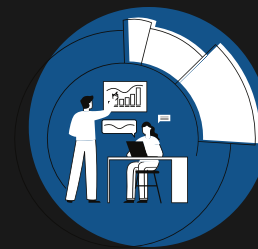- To avoid fraudulent post for job in the internet, an automated tool using machine learning based classification techniques is proposed .
- Many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers.
- intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them.
- Many people fall for this scam and lose a lot of money
- By performing an exploratory data analysis on the data, based on the insights we would gain, we can find out which job posting are fake and which are not.

Introduction

- The current market situation has led to high unemployment.
- Economic stress and the coronavirus's impact have significantly reduced job availability and job loss for many individuals.
- A case like this presents an appropriate opportunity for scammers.
- Fake job detection is an important problem to solve but it has not received much attention from the research community and it is currently a relatively unexplored area.
- This is a dangerous problem that can be addressed through Machine Learning techniques and Natural Language Processing (NLP).

Review of Literature

- According to several studies, Review spam detection, Email Spam detection, Fake news detection have drawn special attention in the domain of Online Fraud Detection.

- ORF detection is a relatively new field and there is not much work done in this area.

- To the best of our knowledge only Vidros . propose a method to detect the fraud jobs.

- However, they worked only with balanced dataset and the performance of prediction algorithms on imbalance dataset set is not known.

- Hence it is important to evaluate the prediction models on imbalanced dataset.

- The ORF Detector for online fraud detection proposed is an ensemble-based model. They have taken three baseline classifiers J48, Logistic Regression and Random Forest and applied average vote,Majority vote and Maximum vote on the classifers.
- But the main drawback of this approach is it only worked on balanced dataset and also yeilds less accuracy.

Metrics	Classifiers			
	Logistic Regression	J48	Random Forest	Avg.
Accuracy	95.3%	95.5%	95.5%	95.4%
Precision	93.9%	94.8%	94.8%	94.5%
Recall	95.3%	95.5%	95.6%	95.6%

Proposed Methodology

- This project aims to create a classifier that will have the capability to identify fake and real jobs.
- **Goal** : Verify Job authenticity due to numereous job scams and fake postings because of unemployment .

Model Deployment

- To make our model available for end users we are going to deploy our model using Flask or Heroku.
- Flask is a web application framework written in Python. It has multiple modules that make it easier for a web developer to write applications without having to worry about the details like protocol management, thread management, etc.
- Heroku is a cloud platform that supports several programming languages in which we can deploy our applications.

Experimental Results

- Logistic Regression

- Support Vector Machine(SVM)

- Naive Bayes Classifier

- Random Forest Algorithm

	Classifiers			
	Naïve Bayes	Support Vector Machine	Logistic Regression	Random Forest
Accuracy	84.2%	95.3%	96.1%	97%

Performance Analysis

- When performing classification predictions, there are four types of outcomes that could occur : TP(True Positive), TN(True Negative), FP(False Positive), FN(False Negative)
- We have used four metrics for evaluating the performance of Fake Job detection system which are:

	Logistic Regression	SVC	Random Forest	Best Score
Accuracy	0.957	0.957	0.968	Random Forest
Precision	0.762	0.781	0.992	Random Forest
Recall	0.184	0.155	0.360	Random Forest
F1 score	0.293	0.258	0.525	Random Forest

Conclusion

- Fake job recruitment detection will guide job-seekers to get only legitimate offers from companies
- For tackling fake job recruitment detection, several machine learning algorithms are proposed as countermeasures in this paper.
- Supervised mechanism is used to demonstrate the utilization of many mechanisms.
- The results of the experiments show that Random Forest is effective.
- The proposed method had a 97 percent accuracy rate
- Which is significantly greater than current approaches