



Detection of Fake Job Recruitment Using Machine Learning (ML)

Ms. SUNKARA. SWARNA DURGA DEVI^{#1}, Mrs. B.SUSHMITHA ^{#2},

Mr. D.D.D.SURIBABU ^{#3}

^{#1}M.Tech Student, ^{#2} Assistant Professor, ^{#3} Associate Professor & HOD

Department of Computer Science & Engineering,

International School of Technology and Sciences (Women),

(Affiliated to JNTUK), East Gonagudem, Rajanagaram, Rajamahendravaram,

East Godavari District, Andhra Pradesh, India– 533294.

ABSTRACT

An application using machine learning-based categorization algorithms is presented in the project to prevent fraudulent job postings online. The outputs of various classifiers are evaluated in order to determine the best employment scam detection model. These classifiers are used to verify fraudulent posts on the web. It assists in identifying phoney job postings among a large number of postings. For the purpose of 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.

KEYWORDS:

Ensemble Classifiers, Machine Learning, Fraudulent Job, Fraud Detection.

1. INTRODUCTION

These days' recruitments are mainly done online through online portals such as naukri.com monster.com Organizations put their job advertisement with desired skills required on these portals. Job seekers or candidates put their resumes and skill details on these portals. Now, companies

can scan the profiles of desired candidates and contact the candidates as well as candidates can also apply to the job profiles in which they are interested. After first screening, companies contact the shortlisted candidates for further processing and recruit the suitable candidates. Online recruitment is beneficial for both candidates as well as the companies.

In Dec 2016, Naukri.com had a database of about 49.5 million registered users, 11000 resumes were getting added daily. This shows the impact that these online job portals have on users. The online recruitment is beneficial for both recruiter as well as candidates. However, in the recent years scammers have started this online recruitment industry which has given a new type of fraud, i.e., Online Recruitment Fraud (ORF). In ORF spammers give lucrative job offers to the candidates and steal their money and private information. ORF not only harms the users but it is also problematic for the companies. As, it damages the reputation of companies and leaves a negative impact in the mind of job seekers about the given company. Shows some of the news snippets that are showing the damage caused by the ORF problem. Shows a snippet from news media that job seekers have lost nearly 2 Crore rupees because of ORF shows a warning sign issued by

one of the MNC (ABB corporation) against ORF. News snippets showing the ORF problem. The amount of loss Happened because of ORF, One of the MNC issuing warning against ORF ORF 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. Detection of fraud job offers from a legitimate set of job is a technically challenging problem. The main challenge the class imbalance problem as the number of fraud jobs are relatively less as compared to the legitimate jobs. This makes learning the features of fraud jobs for automated prediction a challenge.

2. EXISTING SYSTEM AND ITS LIMITATIONS

Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF). In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this 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. Fraudulent job advertisements can be posted against a reputed company for violating their credibility. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs.

LIMITATION OF PRIMITIVE SYSTEM

1. In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers.
2. However, this 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.alerts the user.
3. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially.

4. A classifier maps input variable to target classes by considering training data. Classifiers addressed.

3. PROPOSED SYSTEM AND ITS ADVANTAGES

Machine learning approach is applied which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially. A classifier maps input variable to target classes by considering training data. Classifiers addressed in the paper for identifying fake job posts from the others are described briefly. These classifiers based prediction may be broadly categorized into - Single Classifier based Prediction and Ensemble Classifiers based Predict in this project we got best result with Naïve Bayes Algorithm.

ADVANTAGES OF THE PROPOSED SYSTEM

1. Machine learning approach is applied which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user.
2. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially.
3. A classifier maps input variable to target classes by considering training data. Classifiers addressed.

4. IMPLEMENTATION PHASE

Implementation is the stage where the theoretical design is converted into programmatically manner. In this stage we will divide the application into a number of modules and then coded for deployment. The front end of the application takes Google Collaboratory and as a Back-End Data base we took fake profile

recruitment as dataset. Here we are using Python as Programming Language to implement the current application. The application is divided mainly into following 5 modules. They are as follows:

1. Load Dataset Module
2. Generate Test and Train Data
3. Run Several Algorithms
4. Detect fake profile recruitment identification from Test Dataset
5. Comparative Graph

1) LOAD DATASET MODULE

In this module we try to load the dataset which is collected from Kaggle website and then try to give that excel file information as input to the next module.

Dataset URL: fake-fake-jobposting-prediction.zip

2) GENERATE TEST & TRAIN MODULE

Here we try to divide the data into test and train datasets and we used 70: 30 percent ratio for dividing the whole dataset into multiple parts. Here 70 percent data records are used for training the system and 30 percent data is used for testing the model.

3) RUN SEVERAL ALGORITHMS

Here we try to run the Several algorithm on the train dataset and try to check the probability of each and every attribute which is present in that record. Once all the records are processed now we try to find out which records are having fraud activity and which are having normal activities. Once we use K-Means, Gaussian Mixture, Isolation Forest on training dataset, we can get accuracy of each and every algorithm and finally we can tell Gaussian mixture is accurate more compared with all algorithms.

4) DETECT FAKE JOB RECRUITMENT

Here we try to apply several algorithms and check the model on test data. Once the test data is given as input we can see the data can be categorized into 2 categories where how many records are genuine and how many are not real recruitments.

5) COMPARITIVE ANALYSIS

In this current application we tested the dataset on Logistic Regression, Random Forest, SVM and Naïve Bayes and finally we got random forest is very good in finding the accuracy.

5. EXPERIMENTAL RESULTS

In this section we try to design our current model using Python as programming language and we used Google Collab as working environment for executing the application. Now we can check the performance of our proposed application as follows:

IMPORT LIBRARIES

```
[ ] from google.colab import files
files.upload()

[ ] !pip install -q kaggle

[ ] !mkdir -p ./kaggle
!cp kaggle.json -p ./kaggle
!chmod 600 ./kaggle/kaggle.json

[ ] !kaggle datasets download -d shivamb/real-or-fake-fake-jobposting-prediction

Downloading real-or-fake-fake-jobposting-prediction.zip to /content
87% 14.0M/16.1M [00:00:00.89, 67.7MB/s]
2022-12-01 12:12:18 [00:00:00.89, 66.9MB/s]
```

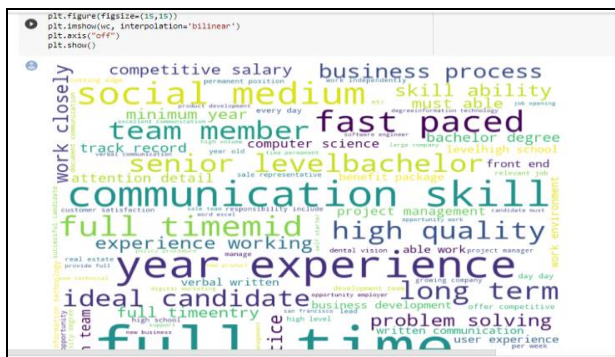
The above window clearly represents the list of several modules used in our application.

PRE-PROCESS THE DATA

job_id	title	department	company_profile	description	requirements	benefits	telecommuting	has_company_logo	has_questions	employment_type	required_experience	required_education
0	Marketing Intern	Marketing	Wish FoodCo. a growing, James Street a purpose-driven...	FoodCo. is hiring management systems & need am...	Experience with content management systems a...		0	1	0	Other	Internship	
1	Customer Service-Cloud Video Production	Success	6/5seconds the world's Direct Video Production	Organized, focused, assertive, awesome! Co you...	What we expect from you/our key responsibility through being part of...	What you will get from us	0	1	0	Full-time	Not applicable	
2	Commissioning (Mech) Engineer (CMA)	Engineering	Major Services provides innovative Solutions th...	Our client located in Houston, is actively se...	Implement pre-commissioning and commissioning...		0	1	0			
3	Account Executive- Washington DC	Sales	Our passion for THE COMPANY: ESPRIMO Systems Rese...	EDUCATION: Bachelor's or Master's in BUS, busi...	Our culture is anything but corporate—we have...		0	1	0	Full-time	Mid-Senior level	Bachelor's Degree
4	BI Review Manager	Sales/Source Solutions LLC is a Global Human Cap...	JOB TITLE: QUALIFICATION(S): license in the State of Texas	Full Benefits offered		0	1	1	Full-time	Mid-Senior level	Bachelor's Degree	

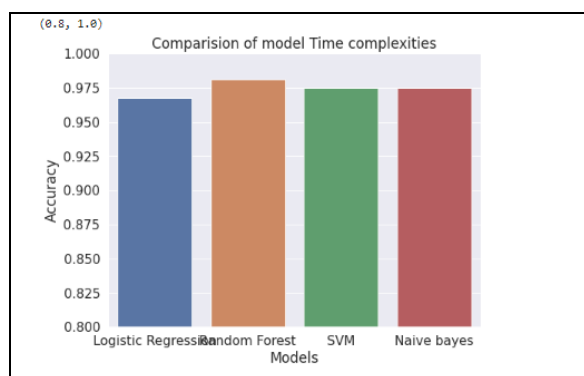
From the above window we can see DATA is pre-processed and we converted dataset into test and train.

WORD CLOUD



From the above window we can clearly see the word cloud.

PERFORMANCE ANALYSIS



From the above window we can clearly see the comparative graph between multiple algorithms and we can finally see random forest has best accuracy.

6. CONCLUSION

Employment scam detection will guide job-seekers to get only legitimate offers from companies. For tackling employment scam detection, several machine learning algorithms are proposed as countermeasures in this paper. Supervised mechanism is used to exemplify the use of several classifiers for employment scam detection. Experimental results indicate that Random Forest classifier outperforms over its peer classification tool. The proposed approach achieved accuracy 98.27% which is much higher than the existing methods.

7. REFERENCES

[1] B. Alghamdi and F. Alharby, —*An Intelligent Model for Online Recruitment Fraud Detection*,” J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.

[2] I. Rish, —*An Empirical Study of the Naive Bayes Classifier An empirical study of the naive Bayes classifier*, no. January 2001, pp. 41–46, 2014.

[3] D. E. Walters, —*Bayes’s Theorem and the Analysis of Binomial Random Variables*, Biometrical J., vol. 30, no. 7, pp. 817–825, 1988, doi: 10.1002/bimj.4710300710.

[4] F. Murtagh, —*Multilayer perceptrons for classification and regression*, Neurocomputing, vol. 2, no. 5–6, pp. 183–197, 1991, doi: 10.1016/0925-2312(91)90023-5.

[5] P. Cunningham and S. J. Delany, —*K -Nearest Neighbour Classifiers*, Mult. Classif. Syst., no. May, pp. 1–17, 2007, doi: 10.1016/S0031-3203(00)00099-6.

[6] H. Sharma and S. Kumar, —*A Survey on Decision Tree Algorithms of Classification in Data Mining*, Int. J. Sci. Res., vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.

[7] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, —*Machine learning for email spam filtering: review, approaches and open research problems*, Heliyon, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.

[8] L. Breiman, —*ST4 Method Random Forest*, Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1017/CBO9781107415324.004.

[9] B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, —*Bagging classifiers for fighting poisoning attacks in adversarial classification tasks*,