FAKE JOB RECRUITMENT DETECTION

KARRI SAI SURESH REDDY

PG Scholar, Department of Computer Science,

SVKP & Dr K S Raju Arts & Science College,

Penugonda, W.G.Dt., A.P, India

sureshreddy645@gmail.com

KARRI LAKSHMANA REDDY

PG Scholar, Department of Computer Science,SVKP &

Dr K S Raju Arts & Science College, Penugonda,

W.G.Dt., A.P, India.

Klreddy29@yahoo.co.in

ABSTRACT

To avoid fraudulent post for job in the internet, an application using machine learning based classification techniques is proposed in the project. 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 single classifier and ensemble classifiers are considered for fraudulent job posts detection. However, experimental results indicate that ensemble classifiers are the best classification to detect scams over the single classifiers.

INTRODUCTION

1.1 Introduction

These days' recruitments are mainly done online through online portals such as naukri.com [1], monster.com [2]. 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 [8]. 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. Figure 1 shows some of the news snippets that are showing the damage caused by the ORF problem. Figure 1(a) shows a snipped from news media that job seekers have lost nearly 2 Crore rupees because of ORF [3]. Figure 1 (b) shows a warning sign issued by one of the MNC (ABB corporation) against ORF. (a) (b) Fig. 1. News snippets showing the ORF problem. a: the amount of loss Happened because of ORF, b: One of the MNC issuing waning 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 challenging task.

2. LITERATURE SURVEY

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

A. Review Spam Detection

People often post their reviews online forum regarding the products they purchase. It may guide other purchaser while choosing their products. In this context, spammers can manipulate reviews for gaining profit and hence it is required to develop techniques that detects these spam reviews. This can be implemented by extracting features from the reviews by extracting features using Natural Language Processing (NLP). Next, machine learning techniques are applied on these features. Lexicon based approaches may be one alternative to machine learning techniques that uses dictionary or corpus to eliminate spam reviews[11].

B. Email Spam Detection

Unwanted bulk mails, belong to the category of spam emails, often arrive to user mailbox. This may lead to unavoidable storage crisis as well as bandwidth consumption. To eradicate this problem, Gmail, Yahoo mail and Outlook service providers incorporate spam filters using Neural Networks. While addressing the problem of email spam detection, content-based filtering, case-based filtering, heuristic based filtering, memory or instance-based filtering, adaptive spam filtering approaches are taken into consideration [7].

C. Fake News Detection

Fake news in social media characterizes malicious user accounts, echo chamber effects. The fundamental study of fake news detection relies on three perspectives- how fake news is

written, how fake news spreads, how a user is related to fake news. Features related to news content and social context are extracted and a machine learning models are imposed to recognize fake news [12]..

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

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.

.

Disadvantages:

- 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.
- raudulent 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.

3.2 Proposed System:

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 Prediction

.

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.
- 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..

3.3 System Modules

Collecting Dataset:

In this step data set with are considered as input has many fields in that except last one all are called as features and last cell is called as labels which are in to 0 1 format.

Data set processing:

Given dataset is not in required format so we need to remove unwanted fields from the dataset and use them as features and process data for scalar format.

Pre-processing:

We convert data in to scalar format and then create new features which are passed to algorithm and features are saved in x and labels in y.

Algorithm Fit:

In this step train features and labels are fit to algorithm and model is saved to system which is used for prediction.

Prediction:

In this step details are fed as input in the form of csv of various profiles and prediction is performed.

Steps to implement Random Forest Classifier in Python

Algorithms

- Step 1 - Import the Libraries

  We will start by importing the necessary libraries required to implement the Algorithm in Python. We will import the numpy libraries for scientific calculation.

- **Step 2 - Fetch the Data**

  We will fetch the data from csv file using '**pandas_datareader**'. We store this in a data frame 'df'

- **Split the Dataset**

  we will split the dataset into **training dataset** and **test dataset**. We will use 70% of our data to train and the rest 30% to test. To do this, we will create a split parameter which will divide the data frame in a 70-30 ratio.

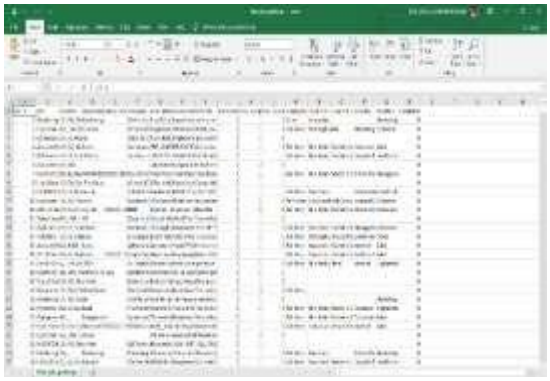
- **Instantiate RandomForestClassifier Model**

  After splitting the dataset into training and test dataset, we will instantiate RandomForestClassifier fit the train data by using '**fit**' function. Then we will store as model.

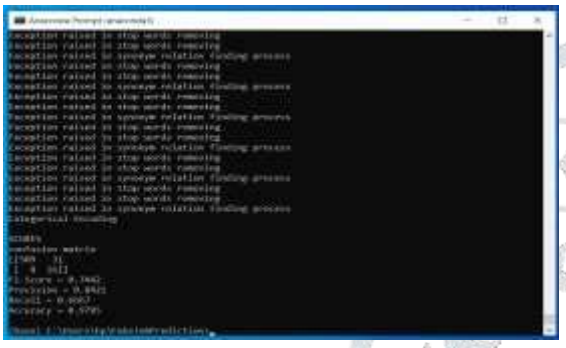
- **Prediction:**

  New input csv file with different job profile details is given as input and prediction is performed and details are stored in new csv with prediction results.

4. RESULTS



Dataset



Test data

Result

5. 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.

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.

REFERENCES

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 Naïve 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," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6713 LNCS, pp. 350–359, 2011, doi: 10.1007/978-3-642-21557-5_37.

ABOUT AUTHORS:

KARRI SAI SURESH REDDY is currently pursuing MCA in SVKP & Dr K S Raju Arts & Science College, affiliated, to Adikavi Nannaya University, Rajamahendravaram. His research interests include Data Structures Web Technologies, Operating Systems and Artificial Intelligent.

K. Lakshmana Reddy is working as an Associate professor in the Department of Computer Science in SVKP & Dr K S Raju Arts &Science College, Penugonda, A.P. He received MCA from Andhra University, 'C' level from DOEACC, New Delhi and M.Tech from Acharya Nagarjuna University, A.P. He attended and presented papers in conferences and seminars. He has done online certifications in several courses from NPTEL. His areas of interests includes Computer Networks ,Network Security and Cryptography, Formal Languages and Automata Theory and Object Oriented programming languages.