ISSN:2229-6107

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY

E-mail :
editor.ijpast@gmail.com
editor@ijpast.in

www.ijpast.in

INTERNATIONAL JOURNAL OF PURE AND APPLIED SCIENCE & TECHNOLOGY
IJPAST

ISSN 2229-6107
www.ijpast.in
Vol 13,Issuse3.july 2023

A COMPARATIVE STUDY ON FAKE JOB POST PREDICTION USING DIFFERENT DATA MINING TECHNIQUES

Dr.C.V.P.R.Prasad[1] , B.Amulya[2], D.Chandana[3],G.Soumya[4],G.Sushma[5]

ABSTRACT

In recent years, due to advancement in modern technology and social communication, advertising new job posts has become very common issue in the present world. So, fake job posting prediction task is going to be a great concern for all. Like many other classification tasks, fake job posing prediction leaves a lot of challenges to face. This paper proposed to use different data mining techniques and classification algorithm like KNN, decision tree, support vector machine, naive bayes classifier, random forest classifier, multilayer perceptron and deep neural network to predict a job post if it is real or fraudulent. We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. Deep neural network as a classifier, performs great for this classification task. We have used three dense layers for this deep neural network classifier. The trained classifiershows approximately 98% classification accuracy (DNN) to predict a fraudulent job post.

INTRODUCTION

In modern time, the development in the field of industry and technology has opened a huge opportunity for new anddiverse jobs for the job seekers. With the help of the advertisements of these job offers, job seekers find out their options depending on their time, qualification, experience, suitability etc. Recruitment process is now influenced by the power of internet and social media. Since the successful completion of a recruitment process is dependent on its advertisement, the impact of social media over this is tremendous. Social media and advertisements in electronic media have created newer and newer opportunity to share job details. Instead of this, rapid growth of opportunity to share job posts has increased the percentage of fraud job

[1]Professor and HOD, Department of CSE,
[2,3,4,5] UG Scholar, Department of CSE, Malla Reddy Engineering College for Women
Hyderabad, Telangana, India
mrecwhodcse@gmail.com
amulyabolisetti2003@gmail.com, chandanadevalla2009@gmail.com,
soumyagovindula24@gmail.com, sushmagudumalla@gmail.com

postings which causes harassment to the job seekers. So, people lacks in showing interest to new job postings due to preserve security and consistency of their personal, academic and professional information. Thus the true motive of valid job postings through social and electronic media faces an extremely hard challenge to attain people's belief and reliability. Technologies are around us to make our life easy and developed but not to create unsecured environment for professional life. If jobs posts can be filtered properly

predicting false job posts, this will be a great advancement for recruiting new employees. . Fake job posts create inconsistency for the job seeker to find their preferable jobs causing a huge waste of their time. An automated system to predict false job post opens a new window to face difficulties in the field of Human Resource Management.

A. Fake Job Posting: Job Scam

Online job advertisements which are fake and mostly willing to steal personal and professional information of job seekers instead of giving right jobs to them is known as job scam. Sometimes fraudulent people try to gather money illegally from job seekers. A recent survey by Action Fraud from UK has

shown that more than 67% people are at great risk who look for jobs through online advertisements but unaware of fake job posts or job scam [2]. In UK, almost 700000 job seekers

complained to lose over $500000 being a victim of job scam. The report showed almost 300% increase over the last two years in UK [2]. Students, fresh graduates are being mostly targeted by the frauds as they usually try to get a secured job for which they are willing to pay extra money. Cybercrime avoidance or protection techniques fail to decrease this offence since frauds change their way of job scam very frequently.

B. Common types of Job Scam

Fraudsters who want to gain other people's personal information like insurance details, bank details, income tax details, date of birth, national id create fake job advertisements. Advance fee scams occur when frauds ask for money showing reasons like admin charges, information security checking cost, management cost etc. Sometimes fraudsters act them- selves as employers and ask people about passport details, bank statements, driving license etc. as pre-employment check.

Illegal money mulling scams occur when they convince students to pay money into their accounts and then transfer it back

[2]. This 'cash in hand' technique causes to work cash in hand without paying any tax. Scammers usually create fake company websites, clone bank websites, clone official looking documents etc. to trap job seekers. Most of the job

scammers try to trap people through email rather than face to face communication. They usually target social media like Linkedln to prove themselves as recruitment agencies or headhunters. They usually try to represent their company profile or websites to the job seeker as realistic as possible. Whatever the type of job scam they use, the always target the job seeker to fall in their trap, collecting information and making benefit either earning money or any other things.

EXISTING SYSTEM

Many researches occurred to predict if a job post is real or fake. A good number of research works are to check online fraud job advertiser. Vidros [1] et al. identified job scammers as fake online job advertiser. They found statistics about many real and renowned companies and enterprises who produced fake job advertisements or vacancy posts with ill-motive. They experimented on EMSCAD dataset using several classification algorithms like naive bayes classifier, random forest classifier, Zero R, One R etc. Random Forest Classifier showed the best performance on the dataset with 89.5% classification accuracy. They found logistic regression performing very poor on the dataset. One R classifier performed well when they balanced the dataset and experimented on that. They tried in their work to find out the problems in ORF model (Online Recruitment Fraud) and to solve those problems using various dominantclassifiers.

Alghamdi [2] et al. proposed a model to detect fraud exposure in an online recruitment system. They experimented on EMSCAD dataset using machine learning algorithm. They worked on this dataset in three steps- data pre-processing, feature selection and fraud detection using classifier. In the preprocessing step, they removed noise and html tags from the data so that the general text pattern remained preserved. They applied feature selection technique to reduce the number of attributes effectively and efficiently. Support Vector Machine was used for feature selection and ensemble classifier using random forest was used to detect fake job posts from the test data. Random forest classifier seemed a tree structured classifier which worked as ensemble classifier with the help of majority voting technique. This classifier showed 97.4% classification accuracy to detect fake job posts.

Huynh [3] et al. proposed to use different deep neural network models like Text CNN, Bi-GRU-LSTM CNN and Bi-GRU CNN which are pre-trained with text dataset. They worked on classifying IT job dataset. They trained IT job dataset on TextCNN model consisting of convolution layer, pooling layer and fully connected layer. This model trained data

through convolution and pooling layers. Then the trained weights were flattened and passed to the fully connected layer. This model used softmax function for classification technique. They also used ensemble classifier (Bi-GRU CNN, Bi-GRULSTM CNN) using majority voting technique to increase classification accuracy. They found 66% classification accuracy using TextCNN and 70% accuracy for Bi- GRU- LSTM CNN individually. This classification task performed best with ensemble classifier having an accuracy of72.4%.

Zhang [4] et al. proposed an automatic fake detector model to distinguish between true and fake news (including articles, creators, subjects) using text processing. They had used a custom dataset of news or articles posted by PolitiFact website twitter account. This dataset was used to train the proposed GDU diffusive unit model. Receiving input from multiple sources simultaneously, this trained model performed well as an automatic fake detector model.

Disadvantages

1) The system is implemented byConventional Machine Learning.
2) The system doesn't implement for analyzing large data sets.

PROPOSED SYSTEM

The system has used EMSCAD to detectfake job post. This dataset contains18000 samples and each row of the datahas 18 attributes including the class label.The attributes are job_id, title, location,department, salary_range, company_profile,description,requirements,benefits,telecommunication,has_company_logo,has_questions,employment_type,required_experience,required_education, industry, function, fraudulent (class label). Among these 18 attribute, we have used only 7 attributes which are converted into categorical attribute. T elecommuting,has_company_logo,has_questions,employment_type,requiredexperience,required_education and fraudulent are changed into categorical value from text value. For example, "employment_type" values are replaced like this- 0 for "none", 1 for 'full-time", 2 for "part-time" and 3 for "others", 4 for "contract' and 5 for "temporary". The main goal to convert these attributes into categorical form is to classify fraudulent job advertisements without doing any text processing and natural language processing. In this work, we have used only those categorical attributes.

Advantages

1) The proposed has been implemented EMSCAD technique which is very accurate and fast.
2) The system is very effective due to accurate detection of Fake job posts which creates inconsistency for the job seeker to find their preferable jobs causing a huge waste of their time.

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Train and Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, Predict Job Post Type Details, Find Job Post Type Prediction Ratio, Download Trained Data Sets, View Job Post Type Prediction Ratio Results, View All Remote Users.

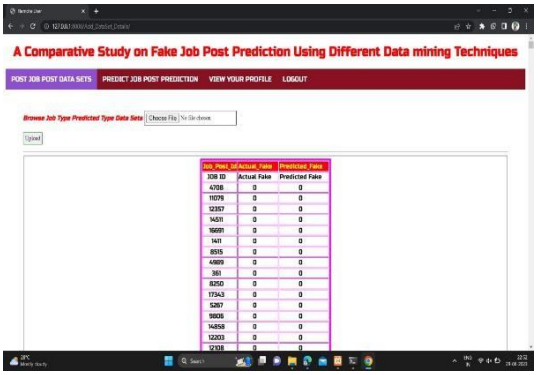
View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like register and login, post job post data sets, predict job post prediction, view your profile.

Fig.1. Login page



.Fig.2. Upload page details.

Fig.3. Upload page details.

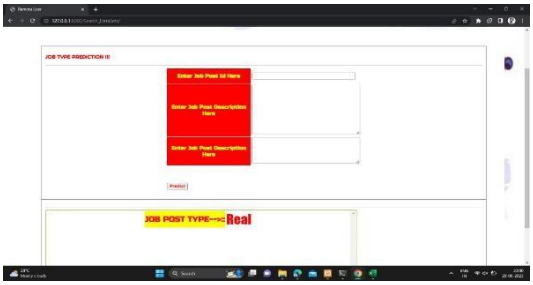


Fig.4. Output results.

CONCLUSION

Job scam detection has become a great concern all over the world at present. In this paper, we have analyzed the impacts of job scam which can be a very prosperous area in research filed creating a lot of challenges to detect fraudulent job posts. We have experimented with

EMSCAD dataset which contains real life fake job posts. In this paper we have experimented both machine learning algorithms (SVM, KNN, Naïve Bayes , Random Forest and MLP) and deep learning model (Deep Neural Network). This work shows a comparative study on the evaluation of traditional machine learning and deep learning based classifiers. We have found highest classification accuracy for Random Forest Classifier among traditional machine learning algorithms and 99 % accuracy for DNN (fold 9) and 97.7% classification accuracy on average for Deep Neural Network.

REFERENCES

[1] S. Vidros, C. Kolias , G. Kambourakis ,and L. Akoglu, "Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset", Future Internet 2017, 9, 6; doi:10.3390/fi9010006.

[2] B. Alghamdi, F. Alharby, "An Intell igent Model for OnlineRecruitment Fraud Detection", Journal of Information Security, 2019, Vol 10, pp.155176,https://doi.org/10.4236/iis.2019.103009 .

[3] Tin Van Huynh1, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen1, and Anh Gia-Tuan Nguyen, "Job Prediction: From Deep Neural Network Models to Applications", RIVF International Conference on Computing and Communication Technologies (RIVF), 2020.

[4] Jiawei Zhang, Bowen Dong, Philip

[5] S. Yu, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network", IEEE 36th

International Conference on Data Engineering (ICDE), 2020.

[6] Scanlon, J.R. and Gerber, M.S., "Automatic Detection of Cyber Recruitment by Violent Extremists", Security Informatics, 3, 5, 2014, https://doi.org/10.1186/s13388-014-0005-5

[7] Y. Kim, "Convolutional neural networks for sentence classification," arXiv Prepr. arXiv1408.5882, 2014.

[8] T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A.G.- T. Nguyen, "Hate Speech Detection on Vietnamese Social Media Text using the Bi-GRU-LSTM-CNN Model," arXiv Prepr. arXiv1911.03644, 2019.

[9] P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," Neurocomputing, vol. 174, pp. 806814,2016.

[10] C. Li, G. Zhan, and Z. Li, "News Text Classification Based on Improved BiLSTM-CNN," in 2018 9th International Conference on Information Technology in Medicine and Education (ITME), 2018, pp. 890-893.

[12] K. R. Remya and J. S. Ramya, "Using weighted majority voting classifier combination for relation classification in biomedical texts," International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014, pp. 1205-1209.