ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

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

S. MALLI BABU[1], EDIGI MANISHA[2], KUSHI GUPTA[3], SAMA SRAVIKA[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

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, naïve 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 classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job post.

Index Terms— Fake job posting prediction, Data mining techniques, Classification algorithms, K-nearest neighbors, Decision tree, Support vector machine, Random forest classifier

I. INTRODUCTION In modern time, the development in the field of industry and technology has opened a huge opportunity for new and diverse 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 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 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. 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[6], [7].

II. LITERATURE SURVEY

A) ORFPPrediction: Machine Learning Based Online Recruitment Fraud Probability Prediction Haiyan Zhang, Meifeng Wang, Yemeng Zhu Published (2023)- In response to the issue of inadequate public datasets concerning imbalances in Internet recruitment scams, this paper constructs an imbalanced experimental dataset, labels it, and proposes an ensemble learning method based on model stacking. This method enhances the prediction accuracy of the original model built with Random Forest. Initially, the method employs label encoding and TF-IDF (Term Frequency- Inverse Document Frequency) to preprocess structured and unstructured data respectively. Then, it utilizes Truncated SVD (Singular Value Decomposition) to reduce the dimensionality of the sparse matrix generated by feature extraction from TF-IDF. Subsequently, decision trees, random forests, and Light Gradient Boosting Machine are utilized as base models, with the random forest employed as the meta-model to integrate the predictions of the base models. This paper presents a comparison of the experimental results from seven machine learning models based on the self-constructed experimental dataset, with model stacking implemented on the random forest that yielded the best experimental result. Through model stacking enhancement of the machine learning models, compared to the random forest model that performed best on the self-constructed experimental dataset, an accuracy of 94.85% was achieved, marking a 7.33% improvement. The experimental results hold significant value and practical implications in understanding and resolving the probability risk issue of recruitment fraud.

B) Smart Fraud Detection Framework for Job Recruitments Asad Mehboob, M. S. I. Malik Published (2020)- Online recruitment has altered the hiring trend. Specifically, posting job ads on career portals and corporate sites involves seeking a large pool of professional applicants around the globe. Unfortunately, it has been established as another platform for fraudsters, which could lead to loss of privacy for applicants and damages organizations' reputation. This case study handles the recruitment fraud/scam detection problem. Several important features of organization, job description and type of compensation are proposed and an effective recruitment fraud detection model is constructed using extreme gradient boosting method. It develops an algorithm that extracts required features from job ads and is tested using three examples. The features are further considered for two-step feature selection strategy. The findings show that features of the type of organization are most effective as a stand-alone model. The hybrid composition of selected 13 features demonstrated 97.94% accuracy and outperformed three state-of-the-art baselines. Moreover, the study finds that the most effective indicators are "salary_range," "company_profile," "organization_type," "required education" and "has multiple jobs." The findings highlight the number of research implications and provide new insights for detecting online recruitment fraud.

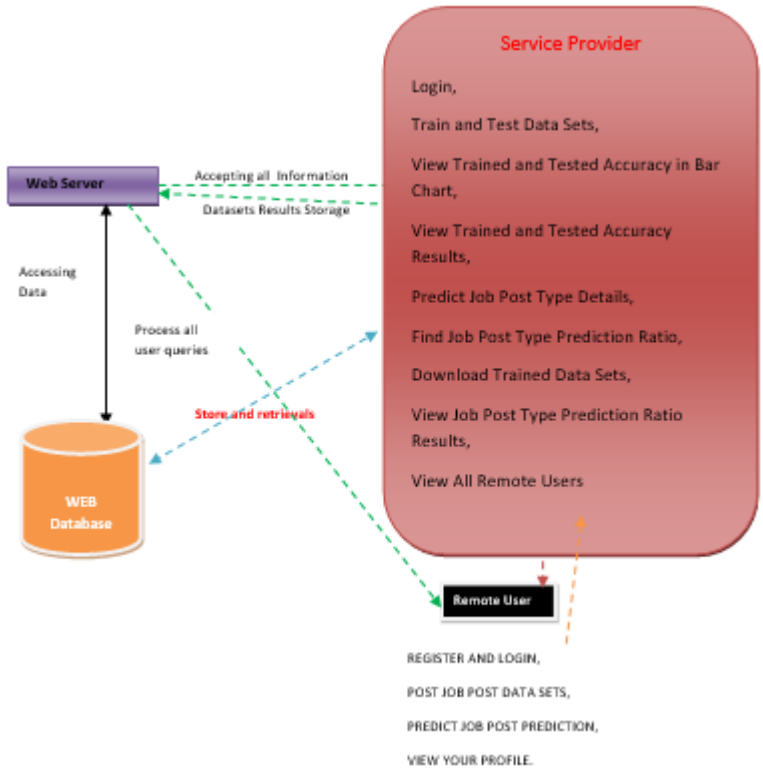
ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

C) Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms C. Anita, P. Nagarajan, G. Sairam, P. Ganesh, G. Deepakkumar Published (2021)- With the pandemic situation, there is a strong rise in the number of online jobs posted on the internet in various job portals. But some of the jobs being posted online are actually fake jobs which lead to a theft of personal information and vital information. Thus, these fake jobs can be precisely detected and classified from a pool of job posts of both fake and real jobs by using advanced deep learning as well as machine learning classification algorithms. In this paper, machine learning and deep learning algorithms are used so as to detect fake jobs and to differentiate them from real jobs. The data analysis part and data cleaning part are also proposed in this paper, so that the classification algorithm applied is highly precise and accurate. It has to be noted that the data cleaning step is a very important step in machine learning project because it actually determines the accuracy of the machine learning as well as deep learning algorithms. Hence a great importance is emphasized on data cleaning and pre-processing step in this paper. The classification and detection of fake jobs can be done with high accuracy and high precision. Hence the machine learning and deep learning algorithms have to be applied on cleaned and pre-processed data in order to achieve a better accuracy. Further, deep learning neural networks are used so as to achieve higher accuracy. Finally all these classification models are compared with each other to find the classification algorithm with highest accuracy and precision.

III. PROPOSED SYSTEM



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 Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, 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, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

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 Intelligent Model for Online Recruitment 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 GiaTuan 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 S. Yu, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network", IEEE 36th International Conference on Data Engineering (ICDE), 2020.

[5] 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

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

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

[7] 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.

[8] 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.

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

[10] 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.