ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 8, Issue 4, 2020

FAKE JOB RECRUITMENT DETECTION USING MACHINE LEARNING APPROACH

[1]Dr.A. Avani,[2]T. Bharath Krishna
[1]Associate Professor, Department of Computer Science and Engineering
Anubose Institute of Technology, New Palvoncha-507115,
Bhadradri Kothagudem-Dist-TG
[2]Department of Computer Science and Engineering
Anubose Institute of Technology, New Palvoncha-507115,
Bhadradri Kothagudem-Dist-TG

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

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

II.LITERATURE SURVEY

1."Fake Job Detection Using Machine Learning: A Comprehensive Review", This literature survey offers an extensive review of the research on fake job post prediction using machine learning techniques. It covers various algorithms and approaches employed in detecting fraudulent job listings, discussing their advantages and limitations.The survey serves as a foundational resource for understanding the landscape of fake job detection.

2."Data Mining Techniques for Fake Job Prediction: A Comparative Analysis" In this survey, we focus on data mining techniques and their application in predicting fake job posts. It provides a comparative analysis of the effectiveness of different data mining approaches, including decision trees, clustering, and classification algorithms, in identifying deceptive job listings.

3."Natural Language Processing for Fake Job Detection: Challenges and Solutions" This survey delves into the application of natural language processing (NLP) in fake job post prediction. It discusses the challenges in analyzing textual data from job listings and reviews NLP-based techniques and tools used to detect anomalies and
inconsistencies in job descriptions.

Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches

Natasha Akram; Rabia Irfan; Ahmad Sami Al-Shamayleh; Adila Kousar; Abdul Qaddos; Muhammad Imran Most companies nowadays are using digital platforms for the recruitment of new employees to make the hiring process

easier. The rapid increase in the use of online platforms for job posting has resulted in fraudulent advertising. The scammers are making money through fraudulent job postings. Online recruitment fraud has emerged as an important issue in cybercrime. Therefore, it is necessary to detect fake job postings to get rid of online job scams. In recent studies, traditional machine learning and deep learning algorithms have been implemented to detect fake job postings; this research aims to use two transformer-based deep learning models, i.e., Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT-Pretraining Approach (RoBERTa) to detect fake job postings precisely. In this research, a novel dataset of fake job postings is proposed, formed by the combination of job postings from three different sources. Existing benchmark datasets are outdated and limited due to knowledge of specific job postings, which limits the existing models' capability in detecting fraudulent jobs. Hence, we extend it with the latest job postings. Exploratory Data Analysis (EDA)

highlights the class imbalance problem in detecting fake jobs, which tends the model to act aggressively toward the minority class. Responding to overcome this problem, the work at hand implements ten top-performing Synthetic Minority Oversampling Technique (SMOTE) variants. The models' performances balanced by each SMOTE variant are analyzed and compared. All implemented approaches are performed competitively. However, BERT+SMOBD SMOTE achieved the highest balanced accuracy and recall of about 90%.

Fake News Detection Using Machine Learning

The emergence of the World Wide Web and the quick uptake of social media platforms (like Facebook, Instagram and Whatsapp) have made it possible for information to be disseminated in ways that have never previously been seen in the history of humanity. Online hiring has changed the hiring pattern. In particular, putting job adverts on corporate websites and career portals includes looking for a sizable pool of qualified candidates worldwide. Unfortunately, it has become yet another forum for scammers, which threatens applicants' privacy and harms the reputation of companies. The topic of detecting recruiting fraud and scams is addressed in this case study. Three machine learning models are used in the construction of an effective recruitment

fraud detection model, which includes a number of significant organizational, job description, and kind of remuneration features. The proposed system uses different Three Machine learning approaches like

Support Vector Machine, Random Forest, Naive Bayes Classifier to determine whether a job posting is genuine or false. We extracted features from data using two techniques: Term Frequency-Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW). As a result, three models achieve decent results. An ensemble model is created by training three independent machine learning models using different segments of samples and then using a simple majority vote of the three models to determine the final predictions. As a result, three models achieve decent results and an accuracy of more than 98.18% have been achieved using Random Forest Model.

III.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. Vidrosetal. 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 dominant classifiers. 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 Text CNN 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 of 72.4%. 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.

IV.PROPOSED SYSTEM

The system has used EMSCAD to detect fake job post. This dataset contains 18000 samples and each row of the data has 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. telecommuting, has_company_logo, has_questions, employment_type, required experience, 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

ISSN 2321-2152
www.ijmece.com
Vol 8, Issue 4, 2020

'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

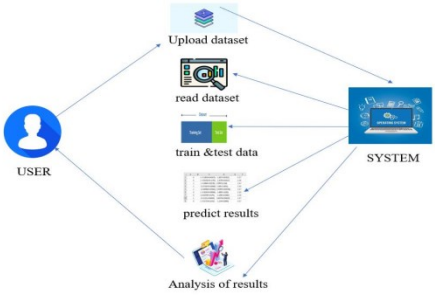

Fig.1 :System architecture

V.RESULT:

This website gives accuracy of Fake Job Post Detection by using three algorithms which are MultinomialNB algorithm with 86% of accuracy ,SVM algorithm with 96% of accuracy and Passive Aggressive Classifier with 86% of accuracy.in this interface we have to enter the URL from the dataset.

This figure display the job post is not a fake job post , it is legitimate job.

VI.CONCLUSION

Job scam detection has become a great concern all over the world at present. In this project, 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 project we have experimented machine learning algorithms (SVM, KNN, Naïve Bayes , Random Forest and MLP).Only reputable business offers will be sent to you.several machine learning methods are proposed for detecting employment scams.In this work,we discuss counter measures.Supervised mechanism is used to demonstrate the utilization of many mechanisms.Classifiers for detecting job scams.The results of the experiments show that MLP is effective. The classifier exceeds its peers in classification. The proposed method had a 98 percent

accuracy rate. Which is significantly greater than current approaches.

VII.FUTURE ENHANCEMENT

The outlined future work demonstrates a forward-thinking approach, integrating advanced models, explainability,multi modal features, and ethical considerations.The emphasis on user feedback, collaboration, and dynamic adaptation positions the project to effectively combat evolving challenges in fake job post detection.Future enhancements in fake job prediction using machine learning will likely focus on refining detection models to improve accuracy and reduce false positives. Advancements in natural language processing (NLP) techniques, such as transformer- based models, will help analyze job descriptions and detect subtle signs of fraudulent listings more effectively. Additionally, incorporating multi-modal data, including company reputation, user reviews, and job market trends, could provide a more holistic approach. Integrating real-time data sources, like social media and employment forums, will allow for dynamic and adaptive models that can detect emerging fake job tactics. Furthermore, explainable AI (XAI) can play a crucial role in making the decision-making process transparent, allowing users to better understand how predictions are made. Finally, collaboration between different sectors, including recruiters, job seekers, and tech companies, could lead to the development of more robust systems for combating job fraud, ensuring a safer and more reliable job market for all.38

VIII.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 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 S. Yu, “FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive

Neural Network", IEEE 36 th 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

[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.

[8] 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.

[9] 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.

[10] Yasin, A. and Abuhasan, A. (2016) An Intelligent ClassificationModel for Phishing Email Detection. International Journal of Network Security& Its Applications, 8, 55-72. https://doi.org/10.5121/imsa.2016.8405.