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

Shreeja.V, Ashitha.P
R.Raj Kumar, Assistant Professor
VISHWA VISHWANI INSTITUTIONS
Survey No. 128, Boston House, Thumkunta Post, Shamirpet Road, Hakimpet (via), Thumkunta, Telangana 500078

Submitted: 02-09-2025 *Accepted: 03-10-2025* *Published: 10-10-2025*

Abstract: The study suggests an automated way of preventing bogus job postings online that uses categorisation techniques based on machine learning. To determine the most effective model for identifying job scams, the output of multiple classifiers was compared. In order to verify false internet postings, these classifiers are used. In the midst of several other ads, it aids in identifying fraudulent job listings. The two fundamental categories of classifiers considered for the purpose.

INTRODUCTION

Employment scams are one of the major issues that have lately been addressed in the area of internet recruiting fraud (ORF). Nowadays, many companies choose to post their vacant positions online so that job seekers may easily and quickly find them. But given that con artists promise work to job searchers in return for cash, this might be one of their ruses. If fake job postings damage a company's reputation, they may come under fire. The creation of an automated system for identifying phoney job advertising and warning individuals about them so they won't apply is of great interest.This is accomplished via a machine learning approach that makes use of several categorization techniques. In this instance, a categorization tool alerts the user after removing fraudulent job ads from a broader group of job postings.

Initially, supervised learning algorithms are viewed as classification techniques to address the issue of recognising fraudsters on job advertisements.

LITERATURE SURVEY

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.

The critical process of hiring has relatively recently been ported to the cloud. Specifically, the automated systems responsible for completing the recruitment of new employees in an online fashion, aim to make the hiring process more immediate, accurate and cost-efficient. However, the online exposure of such traditional business procedures has introduced new points of failure that may lead to privacy loss for applicants and harm the reputation of organizations. So far, the most common case of Online Recruitment Frauds (ORF), is employment scam. Unlike relevant online fraud problems, the tackling of ORF has not yet received the proper attention, remaining largely unexplored until now. Responding to this need, the work at hand defines and describes the characteristics of this severe and timely novel cyber security research topic. At the same time, it contributes and evaluates the first to our knowledge publicly available dataset of 17,880 annotated job ads, retrieved from the use of a real-life system.

2. An Intelligent Model for Online Recruitment Fraud Detection

This study research attempts to prohibit privacy and loss of money for individuals and organization by creating a reliable model which can detect the fraud exposure in the online recruitment environments. This research presents a major contribution represented in a reliable detection model using ensemble approach based on Random forest classifier to detect Online Recruitment Fraud (ORF). The detection of Online Recruitment Fraud is characterized by other types of electronic fraud detection by its modern and the scarcity of studies on this concept. The researcher proposed the detection model to achieve the objectives of this study. For feature selection, support vector machine method is used and for classification and detection, ensemble classifier using Random Forest is employed. A freely available dataset called Employment Scam Aegean Dataset (EMSCAD) is used to apply the model. Pre-processing step had been applied before the selection and classification adoptions. The results showed an obtained accuracy of 97.41%. Further, the

findings presented the main features and important factors in detection purpose include having a company profile feature, having a company logo feature and an industry feature.

3. Yasin, A. and Abuhasan, A. (2016) An Intelligent Classification Model for Phishing Email Detection. International Journal of Network Security & Its Applications, 8, 55-72.

an intelligent classification model to detect phishing emails using knowledge discovery, data mining and text processing techniques. A model based on knowledge discovery (KD) was proposed to build an intelligent email classifier to classify a new email message into legitimate or spam. The knowledge discovery model achieved high accuracy rates in classification of phishing emails that outperformed other schemes. Using the Random Forest algorithm and J48, 99.1% and 98.4% accuracy was achieved respectively. Using MLP classifier, TP rate and FP rate were 0.977 and of 0.026 respectively, while MLP achieved ROC area of 0.987. The results of this study confirmed that the proposed model achieves high rates of accuracy in the classification of phishing e-mail

4. Al-garadi, M.A., Varathan, K.D. and Ravana, S.D. (2016) Cybercrime Detection in Online Communications: The Experimental Case of Cyberbullying Detection in the Twitter Network. Computers in Human Behavi

Al-garadi, et al. (2016) introduced a study that has investigated cybercrime detection in online communications especially cyber bullying in Twitter. The main aim was to develop a number of unique features derived from Twitter. They included network, activity, user, and tweet content. A model to detect cyber bullying in Twitter was proposed using engineering features. The number of friends (followers), the number of users being followed (following), the following and followers ratio, and account verification status were collected through a survey. Users' activity features were also employed to measure the online communication activity of a user. The features implemented and included personality, gender and age. Naïve Bayes (NB), Support vector machine (SVM), Random forest and KNN were applied. Random forest showed f-measure 93%. The results of this study indicate that the proposed model contributes to providing a suitable solution for the detection of cyberbullying in online communication environments

BACKGROUND STUDY

The fields of review spam identification, email spam detection, and false news detection have attracted the most interest in the field of online fraud detection, according to studies. Consumer product reviews are frequently posted on online forums. When other consumers are making product choices, it may be helpful counsel. Due to the possibility that spammers may change reviews for financial gain, algorithms to identify these fake reviews must be developed. Natural Language Processing can be used to do this by extracting attributes from the reviews (NLP). After that, machine learning algorithms are used to these features. Machine learning methods that use dictionaries or corpora to identify bogus reviews may be replaced by lexicon-based strategies.

METHODOLOGY

The recommended approach merely makes use of the Random Forest algorithm to distinguish between honest and dishonest job adverts. The dataset from a double-blind study is used to train the model to be as accurate as feasible while also accounting for the varied methods of posting jobs on both professional and non-professional websites.

As a result, looking for a job is much more successful than it formerly was, and clients may feel at ease searching for positions online.

Because it has undergone significant study, the dataset used is quite useful. Users may utilise the front-end to predict job descriptions. The suggested approach builds a user-friendly web interface for non-technical users using flask and python. Furthermore, we want to create a very precise method for identifying real from bogus job postings. In order to avoid overfitting, the dataset was assembled from a variety of trustworthy sources and points of view, which contributes to its integrity.

1.Data Preprocessing
2. Training
3.Module Creation using Random Forest
4. Prediction

CONCLUSION

Recently, the global problem of detecting job scams has become more important. Since employment scams could be a very lucrative topic of study and make it difficult to recognise dishonest job advertisements, we

examined their consequences in this work. This study's main contribution is to provide evidence in favour of the hypothesis that deep learning can be used to identify bogus employment in particular circumstances. Our findings demonstrate that a standard ANN classification is capable of identifying a variety of potentially subtle linguistic patterns that a human may use after rigorous pre-processing of a small dataset (or may not be able to perceive). Given how easily people may acquire many of these language characteristics, they are helpful for categorising fake work. Exaggerations, slang terms, and generalisations are just a few of the innate traits that our algorithm has detected to designate bogus jobs.

BIBLIOGRAPHY

[1] Gulshan Shrivastava, Member, IEEE, Prabhat Kumar, Senior Member, IEEE, Rudra Pratap Ojha , PramodKumarSrivastava,SenthilKumarMohan"DefensiveModelingofFakeNewsThroughOnlineSocialNetworks", —Online social networks (OSNs) IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS

[2] deBeer,Dylan&Matthee,Machdel,"Approachesto IdentifyFakeNews:ASystematicLiterature Review." : Integrated Science in Digital Age2020.

[3] BandarAlghamdi,FahadAlharby,"AnIntelligentModelforOnlineRecruitmentFraudDetection". Journal of Information Security. (2019)

[4] Pham, Trung Tin" A Study on Deep Learning for Fake News Detection." Journal of Information Security. (2019) n [5], Manoj Kumar Balwant." Bidirectional LSTM Based on POS tags and CNN Architecture for Fake News" 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). (2019)

[6] Amjad, Maaza , Sidorov, Grigoria, Zhila, Alisaa, Gómez-Adorno, Helenab, Voronkov, Iliac, Gelbukh, Alexander." Bend the truth" Special section: Selected papers of LKE 2019

[7] Rami Mohawesh, Son Tran, Robert Ollington, Shuxiang Xu" Analysis of concept drift in fake reviews detection."Expert Systems with Applications. (2021)

[8] Joma George; Shintu Mariam Skariah; T. Aleena Xavier." Role of Contextual Features in Fake News Detection: AReview" International Conference on Innovative Trendsin Information Technology (ICITIIT). (2021) In[9], Sultana Umme Habiba; Md. Khairul Islam; Farzana Tasnim." A Comparative Study on Fake Job Post Prediction Using Different Data mining Techniques."2ndInternationalConferenceonRobotics, Electrical and Signal Processing Techniques (ICREST) (2021)

AUTHORS

Shreeja.V, Ashitha.P
R.Raj Kumar, Assistant Professor