A Comparative Analysis of Fake Job Post Prediction Using Various Data Mining Techniques

#1V. TEJASWINI, *UG Student,* #2D.ROOPA, *UG Student,*

#3K. MAMATHA, *UG Student,* #4E.TEJASWINI, *UG Student,*

#5Mr. SYED MEER SUBAN ALI, Associate Professor

Department of Computer Science & Engineering,
TRINITY COLLEGE OF ENGINEERING & TECHNOLOGY, PEDDAPALLI, TS

.

ABSTRACT:

In recent years, social media and modern technologies made simpler to market new job opportunities. actual translations of job ads have become one of the most common issues facing contemporary civilization. Consequently, counterfeit job For all, posting the foretelling role will be a huge concern. For illustration False job posing foretelling trails a lot of challenges in many other characterization duties. This report recommended the use of copious data sources. To determine whether a job posting is real or deceptive, mining techniques and classification algorithms like decision random forest, ignorant bayes sorters, random forests, and backing vector machinery deep neural network, multilayer perceptron, and sorter are employed. We have executed experiments.in connection to the 18,000-example Employment Scam Aegean Dataset (EMSCAD). As sorters, deep neural nets play astonishingly well.

1. INTRODUCTION

Modern business and technologizing have advanced to the point that there is a abundant of frizzle and distinct career potentials available to job seekers. With assistances from these job offering advertisements, job searches discover their options based on their time, credentials, experiences, suitability, etc. In these eras, the efficacy of social media and the internet impact the hiring progress. Social media plays a main role in this because the effectiveness of the recruitment process largely depends on how well it is publicized. Chances are increasing as job details are spread through electronic media ads and social media. On the other hand, the rapid increment in the number of methods to broadcast job advertisements has raised the portion of fake job postings that False Employment Utilizements. Using advanced techniques like Random Forests (RF), Decision Trees (DT), and Support Vector Machines (SVM) to Examinate job ads and categorize them as real or counterfeit. These algomenas use machine learnings to recognize patters and characteristics that differentiate genuine job ads from fictional ones.

1.1 Motivation

The increasing frequency of fraudulent activities in online job portals is driving the Fake Job Post Prediction using Supporting Vector Machines (SVM), Decision Trees (DF), and Random Forests (RF). Job searchers regularly face untrue job advertisements that might lead them to lose money, have their identities stolen, or face other negative outcomes. It is important that these issues are resolved to protect the safety of job seekers and uphold the integrity of online job platforms. The driving force behind Fake Job Post Prediction using SVM, DT, and RF is the commitment to create a safer and more trustable online job environment. By

proactively identifying and mitigating the risks related to fake job postings, this approach aims to foster a constructive and secure experience for job seekers and contribute to the overall integrity of online job platforms.

2. LITERATURE SURVEY

Devsmit Ranparia; Shaily Kumari; Ashish Sahani," Fake Job Prediction using Sequential Network",[1] With increased number of data and privacy breaches day-by-day it becomes extremely difficult for one to stay safe online. Number of victims of fake job posting is increasing drastically day by day. The companies and fraudsters lure the job-seekers by various methods, majority coming from digital job-providing web sites. We target to minimize the number of such frauds by using Machine Learning to predict the chances of a job being fake so that the candidate can stay alert and take informed decisions, if required. The model will use NLP to analyze the sentiments and patterns in the job posting. The model will be trained as a Sequential Neural Network and using very popular GloVe algorithm. To understand the accuracy in real world, we will use trained model to predict jobs posted on Linked In. Then we worked on improving the model through various methods to make it robust and realistic.

Sultana Umme Habiba; Md. Khairul Islam; Farzana Tasnim," A Comparative Study on Fake Job Post Prediction Using Different Data mining Techniques ",[2] 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.

Ranga Chakradhar, Viji Amutha Mary, Ranga Avinash, Jesudoss, S. Prayla Shyry, "A Comparative Study on Fake Job Post Prediction Using Different Machine Learning Techniques"[6] Lately, with the progression of current innovation and advertising, the advancement of new positions has turned into a test in this day and age. Along these lines, crafted by distorting predictions will be of worry to all. Likewise with numerous order assignments, making bogus predictions requires a great deal. The development requires the utilization of an assortment of strategies for AI and sequencing calculations, for example, KNN, memory sequencing, swarm collectors, and profound organizations, to anticipate whether the innovation is valid or bogus. We tried the Employment Scam Aegean Dataset (EMSCAD), which contains 18,000 occasions. We utilized three complete segments for this profound pipeline. The classifications shrouded in the preparation represent around 98% of the classifications (DNN) for foreseeing extortion.

Authors: Priya Khandagale , Akshata Utekar, Anushka Dhonde, Prof. S. S. Karve "Fake Job Detection Using Machine Learning",[7] The research proposes an automated solution based on machine learning-based classification approaches to prevent fraudulent job postings on the internet. Many organizations these days like to list their job openings online so that job seekers may access them quickly and simply. However,

this could be a form of scam perpetrated by con artists who offer job seekers work in exchange for money. Many people are duped by this fraud and lose a lot of money as a result. We can determine which job postings are fraudulent and which are not by conducting an exploratory data analysis on the data and using the insights gained. In order to detect bogus posts, a machine learning approach is used, which employs numerous categorization algorithms. The system would train the model to classify jobs as authentic or false based on previous data of bogus and legitimate job postings. To start, supervised learning algorithms as classification techniques can be considered to handle the challenge of recognizing scammers on job postings.

3. ALGORITHM

3.1 SVM

- Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

- The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

- SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplan.

3.2 Decision Tree

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree- structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

- The decisions or the test are performed on the basis of features of the given dataset.

- Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

3.3 Random Forest

- As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final

output.

- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

4. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.A system architecture can consist of system components and the sub-systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture, collectively these are called architecture description languages.
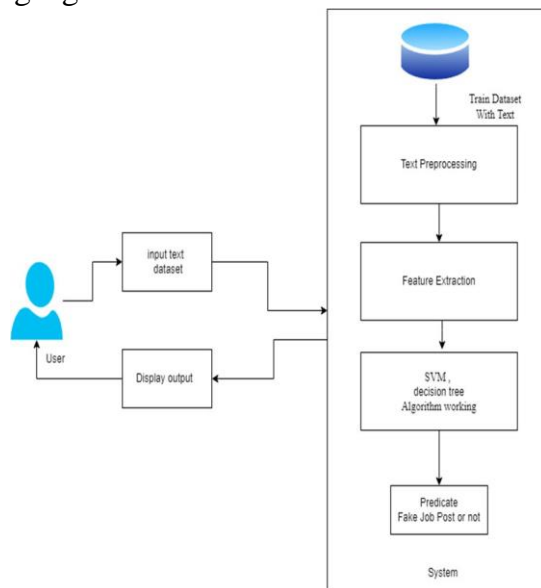
5. RESULTS

5. Conclusion

These times, finding job scums has become a large global concern. In this piece, we observed at the impacts of employment scums—a fast growing field of research that creates it hard to spot fraudulent job lists. In our researches, actual-life phony job lists

from the dataset were utilized. In this study, they have both been put to the test.machine learning techniques (SVM, decision tree) and deep learning models. The assessment of classifiers based on machine learning and deep learning is compared in this essay. We have observed that SVM and decision trees have the peak classifying accuracy.

REFERENCES

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

4) Philip S. Yu, Jiawei Zhang, and Bowen Dong, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network", IEEE 36th International Conference on Data Engineering (ICDE), 2020.

5) "Automatic Detection of Cyber Recruitment by Violent Extremists", Security Informatics, 3, 5, 2014, doi:10.1186/s1338-014-0005-5.

6) Y. Kim, "Convolutional neural networks for sentence classification," arXivPrepr.

7) "Hate Speech Detection on Vietnamese Social Media Text Using the Bi-GRULSTM-CNN Model," arXivprepr. arXiv1911.03644, 2019. T. Van Huynh, V. D. Nguyen, K. Van Nguyen, N. L.-T. Nguyen, and A.G.-T. Nguyen.

8) "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," Neurocomputing, vol. 174, no. 806814, 2016. P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao.

9) "News Text Classification Based on Improved BiLSTM-CNN," in 9th International Conference on Information Technology in Medicine and Education (ITME), 2018, pp. 890–893."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) K. R. Remya and J. S. Ramya.An Intelligent Classification Model for Phishing Email Detection, Yasi, A. and Abuhasan, A. (2016). 8, 55–72 of the International Journal of Network Security and Its Applications. https://doi.org/10.5121/imsa.2016.8