Fake Job Listing Analysis Domain: Machine Learning

SRUJANA VUPPALA
Department of CSE,
Sreyas Institute of Engineering
and Technology, Telangana, India
srujanavuppala25@gmail.com

I.MEGHANA REDDY
Department of CSE,
Sreyas Institute of Engineering and Technology, Telangana, India
inole.meghanareddy19@gmail.com

Dr. V.BIKSHAM
Professor, CSE.
Sreyas Institute of Engineering and Technology, Telangana, India
biksham.v@sreyas.ac.in

B.HARSHATH
Department of CSE,
Sreyas Institute of Engineering
and Technology, Telangana, India
harshathbabu222@gmail.com

G.TARUN
Department of CSE,
Sreyas Institute of Engineering and Technology, Telangana, India
tarunkumar251100@gmail.com

INTRODUCTION

Employment fraud is one of the most severe concerns addressed in the sphere of Online Recruitment Frauds in recent years. 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. For damaging a reputable company's credibility, fraudulent job adverts can be posted. These fraudulent job posting detections have sparked interest in developing an automated method for detecting phoney jobs and reporting them to the appropriate authorities. In order to det1ect bogus posts, a machine learning approach is used, which employs numerous categorization algorithms. A classification tool is used to separate bogus job postings from a wider set of job adverts in this scenario. To start, supervised learning algorithms as classification techniques are being studied to address the challenge of recognizing scammers on job postings. A classifier uses training data to map input variables to target classes. The paper's classifiers for distinguishing phone job postings from the others are briefly presented. These classifiers based prediction is classified into two types – Single classifier based prediction and ensemble classifier based prediction.

LITERATURE SURVEY

The whole concept that is used in the project is all about the machine learning and a great computer scientist Tom Mitchell stated that always a computer program learns from the experiences and also from some tasks and also from some experiences and hence performance increases and also he says that machine learning is totally concerned about

the relationship between the datasets and then uses these relationships to find the interesting patterns in data and hence this is used for the prediction purpose. This machine contains many and many algorithms and they are as follows Regression, Linear Regression, Naive Bayes classifier, KNN, Decision Tree, ID3,SVM, Random Forest and many more.

Arthur Samuel was the one who has coined the name which is Machine Learning. This machine learning is all about the leaning and getting with various machine learning algorithms and then making the predictions and giving the conclusions. This machine learning is also used in solving many complex problems and this is also called as predictive analytics.

PAPER-1

Richard J. Bolton.David J. Hand."Statistical Fraud Detection: A Review."Statist. Sci.17(3)235 - 255,August 2002.

This paper presents about the Statistical Fraud detection. The main theme of the paper is Fraud is increasing dramatically with the expansion of modern technology and the global superhighways of communication, resulting in the loss of billions of dollars worldwide each year. Although prevention technologies are the best way to reduce fraud, fraudsters are adaptive and, given time, will usually find ways to circumvent such measures. Methodologies for the detection of fraud are essential if we are to catch fraudsters once fraud prevention has failed. Statistics and machine learning provide effective technologies for fraud detection and have been applied successfully to detect activities such as money laundering, e-commerce credit card fraud, telecommunications fraud and computer intrusion, to name but a few. We describe the tools available for statistical fraud detection.

PAPER-2

L. Alekya, L. Lakshmi, G. Susmitha, and S. Hemanth, "A survey on fake news detection in social media using deep neural networks," *International Journal of Scientific & Technology Research*, **vol. 9, no. 3, 2020.**

The paper presents about the fake news detection. This we have taken as reference because there will be some similarity is finding out the fake jobs things like the process, methodology etc.Fake news, or fabric which appeared to be untrue with point of deceiving the open, has developed in ubiquity in current a long time. Spreading this kind of data undermines societal cohesiveness and well by cultivating political division and doubt in government. Since of the sheer volume of news being disseminated through social media, human confirmation has ended up incomprehensible, driving to the improvement and arrangement of robotized strategies for the recognizable proof of wrong news. Fake news publishers use a variety of stylistic techniques to boost the popularity of their works, one of which is to arouse the readers' emotions. Due to this, text analytics' sentiment analysis, which determines the polarity and intensity of feelings conveyed in a text, is now being utilized in false news detection methods, as either the system's foundation or as a supplementary component. This assessment analyzes the full explanation of false news identification. The study also emphasizes characteristics, features, taxonomy, different sorts of data in the news, categories of false news, and detection approaches for spotting fake news. This research recognized fake news using the probabilistic latent semantic analysis approach. In particular, the research describes the fundamental theory of the related work to provide a deep comparative analysis of various literature works that has contributed to this topic. Besides this, a comparison of different machine learning and deep learning techniques is done to assess the performance for fake news detection. For this purpose, three datasets have been used.

EXISTING SYSTEM

In existing approach, just a mathematical model is used; it does not classify the fake job post in an efficient way. The available systems are not sufficient to deal with the huge dataset. Machine learning algorithms that give great accuracy are not used.

Disadvantages

1. No proper user interface to enter the details correctly.

2. Less accuracy.

PROPOSED SYSTEM

The target of this project is to detect whether a job post is fraudulent or not. Identifying and removing bogus job adverts will allow job seekers to focus solely on authentic job postings. In Proposed approach we are using we are using 5 machine learning algorithms like Random forest, Ada boost, Linear SVC, Decision Tree and Gradient boost to predict the accuracy. We will consider the classifier with more accuracy. We are also using Natural language processing techniques like removing stop   words, lemmatization, tokenizing

Advantages

1. The biggest benefit of using various algorithms is to get more accuracy and more Precise result.

2. This model in a future scope can be useful to many youngsters seeking for a job.

SYSTEM DESIGN

The purpose of the design phase is to plan a solution of the problem specified by the requirement document. It is the process of defining software methods, functions, objects and overall structure and interaction of your code so that the resulting functionality will satisfy your users requirements. It allows you to do the best abstraction, to understand the requirements better and meet them better. This prevents redundancy and increases re-usability. This phase is the first step in moving from the problem domain to the solution domain. In other words, starting with what is needed, design takes us towards how to satisfy the needs. The design of a system is perhaps the most critical factor affection the quality of the software; it has a major impact on the later phase, particularly testing, maintenance. The output of this phase is the design document. This document is similar to a blueprint for the solution and is used later during implementation, testing and maintenance. The design activity is often divided into two separate phases System Design and Detailed Design.

System Design also called top-level design sign aims to identify the modules that should be in the system, the specifications of these modules, and how they interact with each other to produce the desired results. During, Detailed Design, the internal logic of each of the modules specification in system design is decided. During this phase, the details of the data is usually specified in a high-level design description language, which is independent of the target language in which the software will eventually be implemented.

In system design the focus is on identifying the modules, whereas during detailed design the focus is on designing the logic for each of the modules. During the system design activities, Developers bridge the gap between the requirements specification, produced during requirements elicitation and analysis, and the system that is delivered to the user.

RESULTS

1. The above figure shows us the data set which is being used the project Fake Job Listing Analysis.
2. The attributes in the data set are Company name, location, department, description, benefits etc.
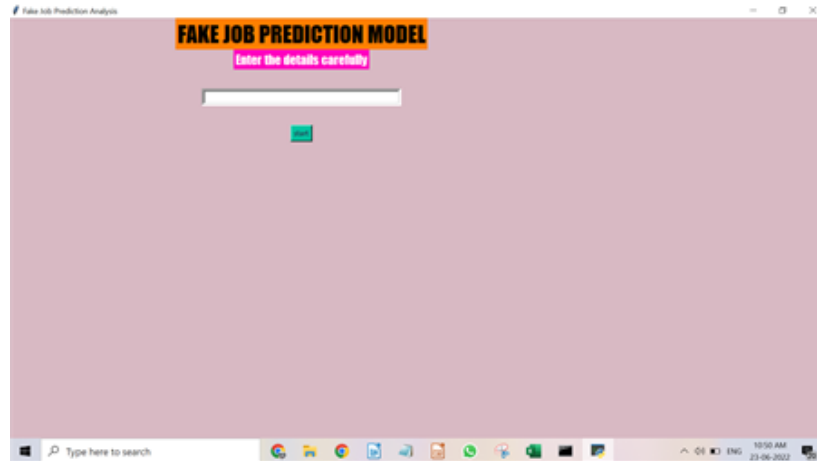
1. This figure shows the welcome page of the Fake Job Listing user interface.
2. Here in the text box the job advertisement should be entered and start the search to check whether the advertisement is genuine or not.

1. The final result window figure shows us about whether the job advertisement is genuine or not.
2. The above window figure shows us that the job advertisement entered in the text box is not genuine and that it is a fake advertisement and give a warning to the user about the advertisement.
3. It also gives a message to make an awareness of these type of job advertisements.

CONCLUSION

As the technology getting improved day by day lot of people are getting adapted to these technology even sometimes without having proper on it and hence that leads to the many scams in the market. Today the unemployment is a huge and huge problem to the youth so as the youth are so much eager of getting into a job and this is a small loop hole to the scammers and they started doing the scams on this and our this model is a small help towards finding pout the fake jobs but there are many and many different kind of fake jobs which cannot be trained and tested in our training model and hence the best solution is to be so aware of the fraud companies and using some official known websites and the more best thing is that always we need to surf the things safely safe surfing is one the best habit.

FUTURE SCOPE

The goal of any society shouldn't be to just catch fake job recruiters but to prevent fake job recruiters from posting job on portal in the first place. we can get more accurate results by considering dataset with no null values.

In future we can perform training the model using different hybrid algorithms to predict the fake job. And best among all algorithms is taken for predicting Fake job post.

Since the data has only 800 fake jobs out of 17,880, next step will be to reduce the number of real jobs or increase the number of fake jobs and then run a model and check the accuracy.

Even the future scope of the project is whenever the fake job is been predicted then an alert need to be created and circulated among all the official websites.

REFERENCES

1. T. Fawcett, "An Introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861– 874, 2006.
2. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion et al., "Scikit-learn: Machine learning in python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
3. A. Natekin and A. Knoll, ―Gradient boosting machines, a tutorial,‖ Front. Neurorobot., vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.
4. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, ―Fake News Detection on Social Media,‖ ACM SIGKDD Explor. Newsl., vol. 19, no. 1, pp. 22–36, 2017, doi: 10.1145/3137597.3137600
5. Shivam Bansal (2020, February). [Real or Fake] Fake JobPosting Prediction,Version 1.Retrieved March 29,2020 from https://www.kaggle.com/shivamb/real-or-fake-fakejobposting-prediction
6. H. Sharma and S. Kumar, ―A Survey on Decision Tree Algorithms of Classification in Data Mining,‖ Int. J. Sci. Res., vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.
7. E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems,‖ Heliyon, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.
8. https://www.machinelearningplus.com/nlp/lemmatization-examples-python/