Fake Job Prediction using Machine Learning

V.Mahitha, T.Yasaswini Anjali, S.Thriveni, SK.Nannu Khaja

*(Under the guidance of **Mrs.Y.Swathi**, Assistant Professor, Department of ComputerScience and Engineering, Tirumala Engineering College)*

ABSTRACT

With increased number of data and privacy breaches(stolen) 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(tempt) the jobseekers 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 alertand take informed decisions, if required. The model will use NLP to analyse the sentiments and pattern in the job posting. The model will be trained as a Sequential Neural Network and using very popular Glove algorithm (Glove is an unsupervised learning algorithm for obtaining vector representations for words). To understand the accuracy in real world, we will use trained model to predict jobs posted. Then we worked on improving the model through various methods to make it robust and realistic.

I.   INTRODUCTION

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and adjust actions accordingly. Fake or real job prediction is an important task to identify whether the job is fake or real. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs.

Employment scams are on the rise. According to CNBC, the number of employment scams doubled in 2018 as compared to 2017. The current market situation has led to high unemployment. Economic stress and the coronavirus ' s impact have significantly reduced job availability and job loss for many individuals. A caselike this presents an appropriate opportunity for scammers. Many people are falling prey to these scammers using the desperation that is caused by an unprecedented incident. Most scammers do this to get personal information from the person they are scamming. Personal information can contain addresses, bank account details, social securitynumbers, etc. I am a university student, and I have received several such scam emails. The scammers provide users with a very lucrative job opportunity and later askfor money in return. Or they require investment from the job seeker with the promise of a job. This is a dangerous problem that can be addressed through Machine Learning techniques and Natural Language Processing (NLP).

II. LITERATURE SURVEY

Online recruiting fraud detection is a relatively new sector in which little research has been done.There are some indirect methods to solve Online recruitment fraud to a limited extent, such as Email Spam filtering, which prevents sending advertising-related emails to users, anti-phishing techniques to detect fake websites, and countermeasures against opinion fraud to detect the posting of deceptive and misleading fake reviews. Review spam detection, email spam detection,and fake news identification have all received a lot of attention in the realm of online fraud detection, according to many studies.

A. Review Spam Detection

People frequently share their opinions on the things they buy on online forums. It might be useful to other buyers while they're deciding what to buy. In this setting, spammers can modify reviews for financialadvantage, necessitating the development of algorithms to detect spam reviews. This can be done by extracting features from the reviews and using Natural Language Processing to do so (NLP). These features are subjected to machine learning algorithms.

B. Email Spam Detection

Unwanted bulk messages, sometimes known as spam emails, frequently occur in user inboxes. This could result in an inevitable storage shortage as well as increasedbandwidth usage. Spam filters based on Neural Networks are used by Gmail, Yahoo Mail, and Outlook to combat

this problem. Content-based filtering, case-based filtering, heuristic-based filtering, memory or instance- based filtering, and adaptive spam filtering approaches are all considered when tackling the problem of email spam detection.

C. Fake News Detection

In social media, fake news is defined by malicious user accounts and echo chamber effects. Fake news identification is based on three perspectives: how fake news is written, how fake news spreads, and how a user is connected to fake news. To identify fake news, features linked to news content and social context are retrieved.

D. Online Fraud Detection

However, the only used a balanced dataset, and the performance of prediction algorithms on an imbalanced dataset has yet to be determined. As a result, evaluating prediction models on an unbalanced dataset iscritical. The suggested ORF Detector is an ensemble- based model for detecting online fraud. They applied average vote, majority vote, and maximum vote to three baseline classifiers: J48, Logistic Regression, and Random Forest. However, the fundamental disadvantage of this strategy is that it only works on balanced datasets and produces lower accuracy.

III. EXISTING SYSTEM

Fig. 1. Fake Job Detection Framework

Fake or real job prediction predicts all job details. Even though, those records are not used in an efficient manner for prediction. To maintain the records in an efficient error free manner, the new proposed system is introduced.

IV. PROPOSD SYSTEM

We proposed to develop a system which will help to predict Fake or real job based on some attributes like location, salary_range, department and so on. So, there is a need for developing a decision system which will help to predict the job condition for employee in an easier way, which can offer prediction about the job so that further procedure can be made effectively. This proposed system not only accurately predicts fake jobs but also reduces time for prediction. The machine learning algorithms like decision tree, random forest, Naive Bayes, K Nearest Neighbours have proven to be most accurate & reliable and hence, used in this project.

Advantages

➢ Generates accurate and efficient results.

➢ Computation time is greatly reduced.

➢ Easy maintenance of employee details.

➢ Reduces

V. RESULTS

VI. CONCLUSION AND FUTURE SCOPE

We have used 5 algorithms like Decision Trees, Random Forests, Naive Bayes, KNN, Gradient Boost in- order to predict fake or real job prediction. The accuracy varies for different algorithms. The accuracy for Decision tree algorithm is 94.96. The accuracy for Naive Bayes algorithm is 59.96. The accuracy for KNN algorithm is 94.59. The highest accuracy is given when we have used Random Forest algorithm with an accuracy of 96.84 percent.

This project further can be developed as Android application to overcome the limitation of accessing the system by only desktop and suggest them about their stage in Fake or real job prediction. And suggest the required instructions to be taken while registering to any job.

VII. REFERENCES

[1] 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.

[2] 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.

[3] L. Breiman, ― ST4_Method_Random_Forest, ‖Mach. Learn., vol. 45, no. 1, pp. 5– 32, 2001, doi: 10.1017/CBO9781107415324.004.