

# An Ensemble Machine Learning Approach To Classifying Fake Jobs And Its Recommendations

Akilan.K<sup>1</sup>, Venkatesh.R<sup>2</sup>, Gokula Krishnan.S<sup>3</sup>, Aravind Kumar.R<sup>4</sup>, Mr.Nagarasan<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup> Dept of CSE

<sup>1, 2, 3, 4, 5</sup> Info Institute of Engineering

**Abstract-** To avoid fraudulent post for job in the internet, an automated tool using machine learning based classification techniques is proposed. Different classifiers are used for checking fraudulent post in the web and the results of those classifiers are compared for identifying the best employment scam detection model. This Project makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. Finally, we discuss how such statistical models of graphs can be combined with text-based information extraction methods for automatically constructing knowledge graphs from the Web. To this end, we also discuss Google's knowledge vault project as an example of such combination.

## I. INTRODUCTION

Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF) . In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them. Fraudulent job advertisements can be posted against a reputed company for violating their credibility. 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. For this purpose, machine learning approach is applied which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially. A classifier maps input variable to target classes by considering training data.

Classifiers addressed in the Project for identifying fake job posts from the others are described briefly. These

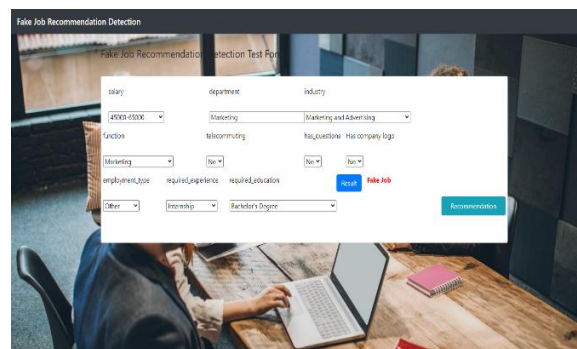
classifiers based prediction may be broadly categorized into – Single Classifier based Prediction and Ensemble Classifiers based Prediction. And real Job Recommendation system

## 1.1 PROBLEM DEFINITION

This project aims to create a classifier that will have the capability to identify fake and real jobs. The final result is evaluated based on two different models. Since the data provided has numeric and features, one model will be used on the text data and another on numeric data. The final output will be a combination of the two. The final model will take in any relevant job posting data and produce a final result determining whether the job is real or not. And Real Job Recommendation

## II. PROPOSED SYSTEM

Input variable to target classes by considering training data. Classifiers addressed in the Project for identifying fake job posts from the others are described briefly. These classifiers based prediction may be broadly categorized into –Single Classifier based Prediction and Ensemble Classifiers based Prediction and real Job Recommendation system.In this Project, we are using Random Forest classifier with Machine Learning.



## III. EXISTING SYSTEM

Previous model and the methodologies, to create the ORF detection model where we have used our own dataset. We have created our dataset based on the job field and by

using a publicly accessible dataset as a reference. Furthermore, Logistic Regression, AdaBoost detecting fraudulent

#### IV. FEATURE SELECTION

- Feature selection is the process of reducing the number of input variables when developing a predictive model.
- It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model.
- Statistical-based feature selection methods involve evaluating the relationship between each input variable and the target variable using statistics and selecting those input variables that have the strongest relationship with the target variable. These methods can be fast and effective, although the choice of statistical measures depends on the data type of both the input and output variables.
- As such, it can be challenging for a machine learning practitioner to select an appropriate statistical measure for a dataset when performing filter-based feature selection.

#### V. CONCLUSIONS

Employment scam detection will guide job-seekers to get only legitimate offers from companies. For tackling employment scam detection, several machine learning algorithms are proposed as countermeasures in this paper. Supervised mechanism is used to exemplify the use of several classifiers for employment scam detection. Experimental results indicate that Random Forest classifier outperforms over its peer classification tool. The proposed approach achieved accuracy 97.27% accurate. Based on the obtained results we recommended Real Job for different level of Domain.

The results show that the system algorithm has fast convergence speed and high accuracy and coverage. It can recommend products that meet the needs and interests of users and promote higher click-through rate and purchase rate. Therefore, in the following research work, we should take improving user satisfaction as the main goal. In addition, during the trial operation, the user survey should be carried out many times to make the data more representative. We also need to analyze the reasons for the lack of customer satisfaction in detail and get specific feedback from users.

The debate on fake job detection has been a challenging one due to the complex and dynamic nature of fake job. In this paper, we did an overview of fake job detection models taking into cognizance the various types of fake details. It is a reality that fake job has caused enormous damage hence detecting them become imperative. We

recommend that fake job can be verified based on source, author or publishers and experts can be able to distinguish between those genuine sources and fake sources. We can add it on cloud computing for future purpose

#### REFERENCES

- [1] Pham, Trung Tin” A Study on Deep Learning for Fake News Detection.” *Journal of Information Security*. (2019)
- [2] 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)
- [3] Abedalla, A., Al-Sadi, A., & Abdullah, M. (2019). A closer look at fake news detection: A deep learning perspective. *ACM International Conference Proceeding Series*, 24–28. [\(2019\)](#)
- [4] Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2020). Fake news types and detection models on social media a state-of-the-art survey. In *12th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2020 (Vol. 1, pp. 562–573)*.
- [5] Amjad, Maaza , Sidorov, Grigoria, Zhila, Alisaa, Gómez-Adorno, Helenab, Voronkov, Iliac, Gelbukh, Alexander.” Bend the truth” Special section: Selected papers of LKE 2019 In[7], Rami Mohawesh, Son Tran, Robert Ollington, Shuxiang Xu” Analysis of concept drift in fake reviews detection.” *Expert Systems with Applications*. (2021)
- [6] B. Faqihi, N. Daoudi, I. Hilal, and R. Ajhoun, “Pedagogical resources indexation based on ontology in intelligent recommendation system for contents production in d-learning environment,” *Journal of Computer Science*, vol. 16, no. 7, pp. 936–949, (2020)
- [7] E. L. Sachi Nandan Mohanty Lydia, M. Elhoseny, M. Majid, G. Al Otaibi, and K. Shankar, “Deep learning with LSTM based distributed data mining model for energy efficient wireless sensor networks,” *Physical Communication*, vol. 40, (2020)
- [8] Onan A. “Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach” *Comput Appl Eng Educ* (2020)