# FAKE JOB DETECTION USING MACHINE LEARNING

**Prof. Sanjivni R. Kale[1], Advait A. Manohar[2], Pranav R. Vitankar[3], Sahil S. Hadke[4]**

[1]Assistant Professor at P R Pote College of Engineering and Management Amravati.

[2,3,4]Student at P. R. Pote College of Engineering and Management, Amravati.

**Abstract—** The technology has been updated to one level up, and the idea of hiring the employees the business companies, through online procedure is carried out. This makes the companies to get the employees of required post more immediate and in a faster way. It will be cost-effective as well. By exploring the internet, one can get the job easily of their qualifications and the field they wish to work in it. The posted jobs may be fake or legitimate, which are unaware by the people. To get rid of these kind of problems we come up with a new software which is designed to predict the job posts, as a result producing whether it is fake or legit one. We are designing a system as Fake job Post prediction using the concept of machine learning, in that we are using Random Forest classifier that produces accurate results in an efficient manner. The designed algorithm achieves the result of 98% as compared to the previously used algorithms. The students or users who search for a job may find difficulties in identifying the job posts that are fake and apply for the jobs, entering all the personal information without knowing about it. In some case they may get into the scams like paying money in the form of application fees in the need of job or the assurance of getting job after paying the money. The framework helps us to detect the posted jobs are fake not.

**Keywords—** Fraud Classification, Fraud Detection Techniques, Machine learning, Decision tree, Random forest Logistic regression, Fraud detection and prediction.

## I.INTRODUCTION

In recent years, the proliferation of online job portals has revolutionized the job search process, providing individuals with unprecedented access to a plethora of employment opportunities. However, amidst this convenience lies a growing concern: the prevalence of fake job postings. These deceptive listings not only waste job seekers' time and effort but also pose potential risks such as identity theft and financial scams. As such, the need for robust mechanisms to identify and combat fake job posts has become increasingly urgent.

This project aims to address this pressing issue by leveraging the power of machine learning algorithms. By harnessing the vast amounts of data available on job portals, social media platforms, and other online sources, machine learning models can be trained to distinguish between legitimate job postings and fraudulent ones. Through the analysis of various textual, visual, and contextual features, these models can learn to detect patterns indicative of deceptive practices, thereby enabling the automated identification of fake job posts with high accuracy.

The significance of this research extends beyond the realm of job seekers and recruiters. Businesses operating online job platforms stand to benefit significantly from the implementation of effective fake job post detection systems. By enhancing the integrity and credibility of their platforms, they can foster greater trust among users and safeguard their reputation in the competitive job market landscape. Moreover, the societal implications of combating fake job postings cannot be overstated. By reducing the prevalence of fraudulent employment opportunities, this research contributes to the protection of vulnerable individuals from exploitation and financial harm. Furthermore, it facilitates a more efficient and transparent labor market, wherein job seekers can make informed

decisions based on reliable information.

The digital age has transformed the landscape of employment, offering unprecedented opportunities for job seekers to connect with potential employers through online platforms. However, this convenience comes with a lurking threat: the proliferation of fake job postings. These deceptive listings not only undermine the trust and integrity of online job markets but also pose significant risks to unsuspecting applicants, ranging from financial scams to identity theft.

## II. LITERATUREREVIEW

In recent years, substantial research efforts have been directed towards developing and refining classification algorithms for the effective analysis of large datasets, particularly in the context of loan approval systems. Several noteworthy works in this domain have been consulted for the development of the proposed automation system.

- Performance comparison of two class boosted decision tree and two class decision forest algorithms in predicting fake job postings

Authors: Shibly, F., Sharma, U. & Naleer, H

The two-class boosted decision tree algorithm was more effective than the two-class decision forest algorithm in detecting fake job postings.

- Fake job recruitment detection using a machine learning approach. International Journal of Engineering Trends and Technology

Authors: Dutta, S., & Bandyopadhyay, S. K.

The paper suggests using machine learning, with the Random Forest classifier, to detect employment scams. It achieved an accuracy of 98.27%, outperforming existing methods

- Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence
Authors: Cook J, Lewandowsky S, Ecker

Inoculation neutralizes misinformation's impact on consensus perception and moderates polarization on climate change beliefs and policy support. Inoculations, focusing on general misinformation techniques, effectively complement consensus messaging in combating misinformation

- Fake job detection and analysis using machine learning and deep learning algorithms

Authors: Anita, C., Nagarajan, P., Sairam, G. A., Ganesh, P., & Deepakkumar

Machine learning and deep learning algorithms to distinguish between fake and real job postings. It finds that Bi Directional LSTM yields the most accurate results compared to other classification algorithms.Credit Data PredictionUsing Min-Max Normalization and K Nearest Neighbor (K-NN) Classifier.

This paper presents a study on detecting fake job postings using machine learning and deep learning techniques. It employs algorithms like logistic regression, KNN, random forest, and Bi-Directional LSTM. Data cleaning is highlighted as crucial for improving accuracy. The results show that Bi-Directional LSTM performs best in detecting fake jobs compared to other methods..

## III. SYSTEM DESIGN AND PROCESS

System Design also called top-level design signs 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 module's specifications in system design is decided. During this phase, the details of the data are 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.

1. A data set is being used for the project Fake Job Listing Analysis.

2. The attributes in the data set are Company name, location, department, description, benefits etc.

3. The welcome page of the Fake Job Listing user interface shows up.

4. In the text box, the job advertisement should be entered, and start the search to check whether the advertisement is genuine or not.

5. The final result window figure shows us whether the job advertisement is genuine or not.

6. 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 gives a warning to the user about the advertisement.

7. It also gives a message to make awareness of these types of job advertisements Architecture Diagram dimensional datasets. Naive Bayes classifiers are simple yet powerful, and they perform well in various scenarios.
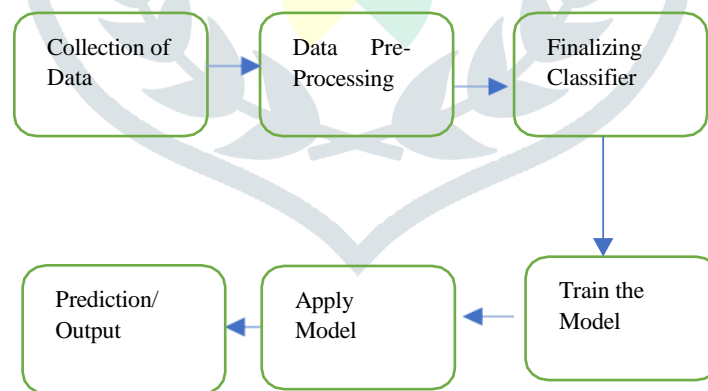


## IV. Naive Bayes

Naive Bayes is a supervised learning algorithm Commonly used for classification tasks. It is based on Bayes' theorem and assumes independence between features, making it particularly effective for high- Naive Bayes, a classification algorithm, assumes feature independence given the class label. Despite potential real-world deviations from this assumption, the algorithm often performs well. It calculates probabilities for features within each class based on the training dataset, estimating the likelihood of feature occurrence given a class. The algorithm determines prior class probabilities and utilizes Bayes' theorem to compute posterior probabilities based on observed features. During classification of a new data point, it calculates these posterior probabilities and assigns the class with the highest probability as the predicted class. Naive Bayes excels in simplicity, efficiency, and handling missing data, making it advantageous for various applications, particularly in text classification and spam filtering.

## V. CONCLUSION

In conclusion, fake job post detection using machine learning presents a critical solution to combat the proliferation of fraudulent activities in online job markets. By leveraging machine learning algorithms and advanced text analysis techniques, these systems can identify suspicious job postings and protect job seekers from falling victim to scams and deceitful practices. Through this project, we have explored various machine learning models, feature engineering methods, and evaluation metrics to develop effective detection systems. While the journey has been insightful, there remain numerous opportunities for further research and innovation to advance the field. Throughout this endeavor, we have delved into various machine learning models, such as logistic regression, support vector machines, decision trees, and ensemble methods, to construct robust detection frameworks. Furthermore, feature engineering methodologies, including TF-IDF vectorization and word embeddings, have enhanced the systems' ability to capture subtle indicators of fraudulent behavior within job postings.

## VI.REFERENCES

[1] Shibly, F., Sharma, U., & Naleer, H. (2021). Performance comparison of two class boosted decision tree and two class decision forest algorithms in predicting fake job postings. Annals of the Romanian Society for Cell Biology, 25(4), 2462– 2472.

[2] Dutta, S., & Bandyopadhyay, S. K. (2020). Fake job recruitment detection using a machine learning approach. International Journal of Engineering Trends and Technology, 68(4), 48–53.

[3] Cook J, Lewandowsky S, Ecker UK (2017) Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. PLoS One 12(5):e0175799.

[4] Anita, C., Nagarajan, P., Sairam, G. A., Ganesh, P., & Deepak kumar, G. (2021). Fake job detection and analysis using machine learning and deep learning algorithms. Revista GEINTEC-Gestao, Inovacao e Tecnologias, 11(2), 642–650

[5] Aljedaani, W., Javed, Y., & Alenezi, M. (2020). LDA categorization of security bug reports in chromium projects. In Proceedings of the 2020 European Symposium on Software Engineering (pp. 154–161).

[6] Aljedaani, W., Nagappan, M., Adams, B., &Godfrey, M. (2019). A comparison of bugs acrosstheiOSandAndroidplatformsoftwoopen-sourcecross-platformbrowserapps. In 2019 IEEE/ACM 6th International Conference on Mobile Software Engineering and Systems (MOBILESoft) (pp. 76–86). IEEE.

[7] Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759. [8] Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2019). Tweets classificationonthebaseofsentimentsforUSairlinecompanies.Entropy,21(11),1078.

[9] Sugumar, R. (2018). Improved performance of stemming using an efficient stemmer algorithm for information retrieval. Journal of Global Research in Computer Science, 9(5), 01–05.