International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Online Recruitment Fraud Detection Using Machine Learning

Dr.PSV. Srinivasa Rao[1], A.Prem kumar[2], Anas Abdullah[3] ,Ch.Rajeshwari[4] ,K.devika[5]

[1]Dr.PSV.Srinivasa Rao (professor)
[2]A.Prem Kumar Department of Computer Science and Engineering (Joginpally b.r. Engineering College) [3]Anas Abdullah Department of Computer Science and Engineering (Joginpally b.r. Engineering College)
[4]Ch.Rajeshwari Department of Computer Science and Engineering (Joginpally b.r. Engineering College)
[5]K.Devika Department of Computer Science and Engineering (Joginpally b.r. Engineering College)

Abstract - The research suggests using machine learning classification methods to identify and prevent fake job postings online. Nowadays, many companies prefer to advertise their job vacancies on the internet for easy access by job seekers. However, there are scammers who trick people into paying for non-existent jobs. Through data analysis and machine learning, we can distinguish between legitimate and fake job postings. Various algorithms are utilized to detect fraudulent posts and protect job seekers from falling victim to scams. The system is designed to teach the model how to distinguish between real and fake job listings using past data on fraudulent and legitimate job postings. Initially, supervised learning algorithms like classification techniques can be used to tackle the problem of identifying scammers in job ads. It will utilize multiple machine learning algorithms and choose the one with the best accuracy in predicting if a job posting is legitimate or not.

Key Words: Fraud Job, Job Seeker, Machine Learning, Internet Recruitment, Classification

1.INTRODUCTION

Employment scams have become a significant concern nowadays within the realm of online recruitment fraud. Nowadays, many companies opt to advertise their job openings online to reach job seekers easily and promptly. However, this practice can sometimes be exploited by fraudsters who deceive job seekers by offering fake job opportunities in exchange for money. These deceptive job postings can tarnish the reputation of reputable companies. Detecting and removing fraudulent job posts has become a pressing issue, highlighting the need for automated tools to identify and report fake jobs to prevent job seekers from falling into these traps. To tackle this issue, we use machine learning techniques that involve various classification algorithms to detect fake posts. A tool is used to separate fake job postings from a pool of job ads and notify the user. Initially, supervised learning algorithms were used to identify scams in job postings. These algorithms map input data to target categories using training data. The paper briefly describes the classifiers used to identify fake job posts, which can be broadly grouped into two categories: single classifierbased prediction and ensemble classifier-based prediction.

A. Single Classifier Based Prediction

Predictions for unknown test cases are made using previously learned classifiers. When identifying fake job postings, the following classifiers are utilized.

- **Naive Bayes**

Naive Bayes classifiers require a number of parameters that grows linearly with the number of variables in a learning problem, making them highly scalable. Instead of using costly iterative approximation like many other classifiers, training with maximum likelihood can be done by simply evaluating a closed-form expression in linear time. Naive Bayes is a simple method for constructing classifiers that assign class labels to problem cases represented as vectors of feature values, with the class labels chosen from a limited set. To assess the accuracy of this classifier, one must estimate the amount of information loss in the class due to the assumption of independence, rather than feature dependencies.

- **Support Vector Machine**

Support Vector Machine (SVM) is a type of supervised learning model that is versatile and can be used for tasks such as classification and regression. SVMs are capable of solving both linear and nonlinear problems, making them useful in various scenarios. In a classification problem, SVM works by drawing a line between different classes to maximize the distance between points on either side of the line. This helps the model accurately predict the target classes for new cases after the separation. This approach is beneficial for solving classification tasks effectively.

- **K-nearest Neighbor Classifier**

The K-nearest Neighbor Classifier, also known as lazy learners, is a method that identifies objects by looking at how close they are to the training examples in the feature space. This classifier looks at the k number of objects closest to determine the class. The main challenge with this method is selecting the right value for k.

3.Ensemble Approach Based Classifiers

The ensemble approach allows multiple machine learning algorithms to work together in order to achieve greater accuracy for the overall system. Random forest (RF) is an example of an ensemble learning technique that combines

regression and classification methods. This classifier uses multiple tree-like classifiers on different subsets of the data, with each tree contributing its vote for the most appropriate class for the input. Boosting is a smart method where multiple unreliable learners are combined into one learner to enhance the accuracy of classification. This technique involves using a classification algorithm on the weighted versions of the training data and selecting the weighted majority decision of a series of classifiers. AdaBoost is a great example of a boosting technique that generates better results even when weak learners underperform. Boosting algorithms are very effective in resolving spam classification issues. Another boosting technique known as the gradient boosting algorithm is a classifier that leverages decision tree principles and reduces prediction errors.

2. RELATED WORK

Detecting online recruiting fraud is a recent field with limited research. Some ways to address this issue include email spam filtering to stop advertising emails, anti-phishing techniques for spotting fake websites, and strategies to detect fake reviews and deceptive opinions. Studies have focused on identifying review spam, email spam, and fake news as key aspects of online fraud detection.

1. Review Spam Detection

Consumers frequently share their opinions on online forums about the products they buy. These reviews can help others make informed decisions when selecting products. However, spammers may try to distort reviews for financial gain. To address this issue, it is important to create methods that can identify and filter out spam reviews. One way to do this is by utilizing Natural Language Processing (NLP) to extract key features from the reviews and then applying machine learning techniques to analyze these features. Alternatively, lexicon-based approaches that rely on dictionaries or corpora can also be used to detect and remove spam reviews.

2. Email Spam Detection

Unwanted messages, commonly known as spam emails, frequently overwhelm user inboxes, causing storage limitations and increased bandwidth usage. Services like Gmail, Yahoo Mail, and Outlook utilize Neural Networks to prevent this issue with effective spam filters. Different techniques, including contentbased filtering, case-based filtering, heuristic based filtering, memory or instance based filtering, and adaptive spam filtering, are used to identify and combat email spam.

3. Fake News Detection

Fake news on social media is typically spread by malicious users and fueled by echo chamber effects. Detecting fake news involves understanding how it is created, shared, and consumed by users. By analyzing features within the news content and social context, machine learning models can be applied to identify fake news sources.

3. PROPOSED SYSTEM

An AI system is used to detect fake job posts by applying various classification algorithms. The system separates fake job ads from a large pool of job listings and notifies the user. Initially, supervised learning algorithms are used to identify scams in job postings. The algorithms map input data to target classes based on training data. The paper describes different classifiers used to identify fake job posts, which can be categorized into Single Classifier based Prediction and Ensemble Classifiers based Prediction. The best result was achieved with the Naive Bayes classifier.

3.1. Advantages

- A machine learning method is used to identify fake posts by employing multiple classification algorithms. The tool specifically targets fake job postings within a large pool of job ads and notifies the user.
- Using a classification tool, fake job postings are singled out from a wide range of advertisements and the user is promptly alerted. Supervised learning algorithms are initially explored as classification techniques to tackle the issue of detecting job scams.
- A classifier uses training data to map input variables to target classes. Various types of classifiers are discussed.

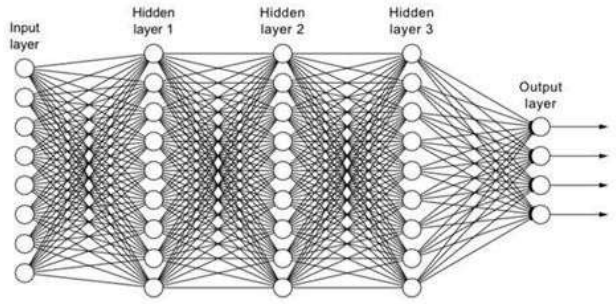
4. METHODOLOGY

We used different data mining methods to figure out if a job listing is legitimate. Once we prepared the data, we taught the EMSCAD models to classify it. The model we created is now able to detect fake job postings online.

- Neural Network

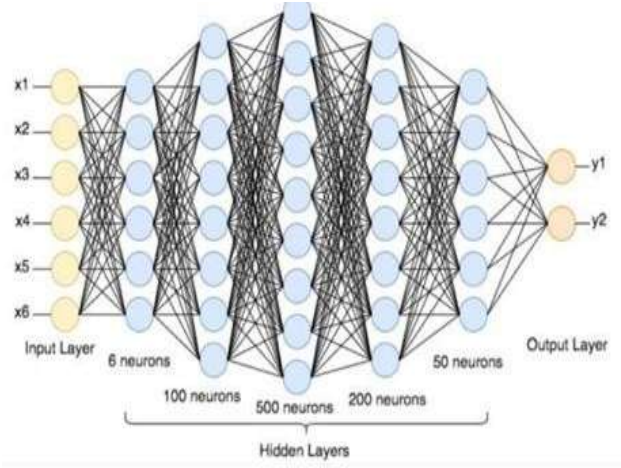
Neural networks are like the human brain - they compare patterns to see how similar or different they are. Neurons in a network categorize patterns by extracting their characteristics. The network has layers of nodes that are connected to each other, and each node acts like a linear regression.In simple terms, multiple linear regression results are processed by the perceptron and transformed into a non-linear activation function. These perceptrons are arranged in interconnected layers. In order to minimize errors, the hidden layers adjust the weights of the input layers. Ultimately, the neural network acts as a classifier in supervised learning scenarios.

● Deep neural networks, also known as DNNs, are a type of Artificial Neural Networks (ANNs) with multiple layers. The feed-forward algorithm powers



Deep Neural Network

DNNs, transferring data from the input layer to the output layer. DNNs create numerous virtual neurons with randomly initialized connection weights. These weights are multiplied by the input, producing an output between 0 and 1. The training process adjusts the weights to effectively categorize the output. However, the model may overfit when it learns obscure patterns from additional layers. Dropout layers help generalize the model by decreasing the number of trainable parameters.



➢ Here are the key steps for building a system for fake job detection using machine learning: Data Collection: Collect a variety of job ads, both real and fake, for the dataset.

Lastly, deploy the trained model into a production environment to handle incoming job postings in real - time.

Data Preprocessing:

Clean and prepare the text data by removing unnecessary details, splitting it into words, eliminating common words, and reducing words to their base form

Feature Extraction:

Transform the processed text data into numerical features using methods such as Bag-of Words (BoW) or TF-IDF. Model Training:

To train the model, select suitable classification algorithms such as Naive Bayes, SVM, and Random Forest, and then use the labeled data for training.

Model Evaluation:

After training, evaluate the models' performance by measuring metrics like accuracy, precision, recall, and F1 score through methods like cross-validation.

Model Deployment:

in deceptive practices, thus enabling more precise detection abilities that are specific to particular areas.

5. FUTURE ENHANCEMENTS

The project "Fake Job Detection Using Machine Learning" opens up exciting possibilities for future advancements. One area that could be further explored is the use of explainable AI (XAI) techniques. By incorporating XAI methods, the system's transparency can be improved, giving users a better understanding of how the machine learning models make decisions and building more trust in the detection system.

The integration of active learning mechanisms is a promising approach. This involves the system requesting feedback from users on unclear or difficult cases, which helps improve its understanding of evolving deceptive tactics and ultimately leads to increased accuracy over time. Ensuring consistency across different platforms is essential for wider effectiveness. Future studies could concentrate on developing a standardized model that can be applied to various job platforms, ensuring a consistent and dependable method for detecting fake job postings regardless of platform differences.

In addition, incorporating geospatial analysis could enhance the system's complexity. Taking into account the location of job listings could offer valuable information on regional differences

6. CONCLUSION

In conclusion, the project titled "Fake Job Detection Using Machine Learning" represents a significant advancement in addressing the problem of misleading job listings online. The study effectively shows how machine learning algorithms can differentiate between real and fake job ads, laying the foundation for a safer and more trustworthy job market. Its impact includes providing job seekers with a proactive tool, preventing financial losses, and enhancing the credibility of job platforms. This research highlights the potential of machine learning in making the job market a safer and more reliable place. In today's online world, job platforms are extremely important. It's clear that we need strong measures to prevent fraud. This research has given us a great starting point for improving online job markets. We want job seekers to feel safe and sure as they look for new opportunities.

7. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our guide, Mrs. J. Lakshmi, for her invaluable support and guidance throughout the project. Her expertise and encouragement were crucial to our success, and we are truly grateful for her unwavering dedication and mentorship. We also want to thank the faculty of the Computer Science and Engineering Department at Tirumala Engineering College for giving us the opportunity to work on this research project, which has been a valuable learning experience.

8. REFERENCES

1. B. Alghamdi and F. Alharby, ―An Intelligent Model for Online Recruitment Fraud Detection," J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.

2. Shawni Dutta and Prof.Samir Kumar Bandyopadhyay "Fake Job Recruitment Detection Using Machine Learning Approach " International Journal of Engineering Trends and Technology (IJETT) – Volume 68 Issue 4- April 2020

3. S.Vidros, C. Kolias , G. Kambourakis ,and L. Akoglu, "Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset",Future Internet 2017, 9, 6; doi:10.3390/fi9010006

4.https://www.spiceworks.com/tech/artificialintelligence/articles/w hat-is-a-neural-network/

5. https://www.researchgate.net/figure/Deep-NeuralNetwork architecture_fig1_330120030

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

7. I. Rish, ―An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive Bayes classifier,‖ no.January 2001, pp. 41–46, 2014.

8. Ms. Sunkara. Swarna Durga Devi, Mrs. B.sushmitha, Mr.D.D.D.suribabu "Detection of Fake Job Recruitment Using Machine Learning (ML)" © 2022 IJCRT | Volume 10, Issue 12 December 2022 | ISSN: 2320-2882

9. D. E. Walters, ―Bayes's Theorem and the Analysis of Binomial Random Variables,‖ Biometrical J., vol. 30, no.7, pp. 817–825, 1988, doi: 10.1002/bimj.4710300710.

10. 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.

11. H. M and S. M.N, ―A Review on Evaluation Metrics for Data Classification Evaluations,‖ Int. J. Data Min. Knowl. Manag. Process, vol. 5, no. 2, pp. 01–11, 2015, doi:10.5121/ijdkp.2015.5201.