International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol.1, Issue. 2 (April – June) , 2024 , Pages: 14 - 16
Received: 23 April 2024 ; *Accepted:* 30 April 2024; *Published:* 01 May 2024
https://doi.org/10.63328/IJCSER-V1RI2P3
Original Paper : Action Research – Report Paper

Design and Implementation of Fake Job Post Detection using Machine Learning Techniques

R Md Shafi [1*], Reddem Janardhan Reddy [2] , Kathi Harsha Vardhan [3] , P Azharuddin [4] , M Manohar [5]

[1-5] Department of CSE , Aditya College of Engineering , Madanapalle , Andhra Pradesh , India;

** Corresponding Author : R.Ms.Shafi; mdshafi.r@acem.ac.in*

Abstract: To avoid fraudulent post for job in the internet, it is a one of the machine learning technique based on classification techniques is proposed in the project. 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. It helps in detecting fake job posts from an enormous number of posts. The major classifier, such as ensemble classifiers are considered for fraudulent job posts detection. However, experimental results indicate that ensemble classifiers are the best classification to detect scams over the other classifiers.

Keywords: ML, Ensemble Classifier, Fake Job Post Detection , Random Forest.

1. 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.

1.1. Problem Statement:

Design and implement a machine learning model to detect fake job postings from legitimate ones. With the rise of online job portals, the prevalence of fake job postings has become a significant concern for both job seekers and recruiters. Fake job postings can lead to various issues such as financial scams, identity theft, and misinformation. The goal of this project is to develop a robust and accurate machine learning solution that can automatically identify fake job postings, thereby improving the trustworthiness and reliability of online job platforms.

1.2. Key Objectives:

Data Collection: Gather a diverse dataset of job postings, including both legitimate and fake postings, from various sources such as online job portals, social media, and forums.

Data Preprocessing: Clean and preprocess the dataset to extract relevant features and normalize the data for machine learning model training.

Feature Engineering: Identify and engineer informative features from the job postings, including textual content, metadata, and contextual information.

Model Selection: Evaluate and compare different machine learning algorithms, such as logistic regression, random forest, support vector machines, and neural networks, to determine the most suitable model for fake job detection.

Model Training: Train the selected machine learning model on the preprocessed dataset using appropriate training techniques, such as cross-validation and hyperparameter tuning, to optimize its performance.

Evaluation Metrics: Define appropriate evaluation metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC, to assess the performance of the trained model and measure its effectiveness in distinguishing between fake and legitimate job postings.

Model Deployment: Deploy the trained machine learning model into a practical application or platform, such as a web service or API, to enable real-time fake job detection for users.

Performance Monitoring: Implement mechanisms to monitor the performance of the deployed model over time and incorporate feedback loops for continuous

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 2 , 2024 , https://doi.org/10.63328/IJCSER-V1RI2P3

improvement and adaptation to evolving fake job posting trends.

2. Relative Work

We begin by reviewing existing literature on fake job post detection using machine learning techniques. Several studies have explored various feature extraction methods to capture relevant information from job postings, including textual content, metadata, and contextual information. Common features used in these studies include keyword frequency, syntactic and semantic features, job description length, and posting metadata (e.g., posting date, location). Additionally, researchers have employed a wide range of classification algorithms to distinguish between legitimate and fake job postings. These algorithms include logistic regression, decision trees, random forests, support vector machines, and neural networks. Evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) are commonly used to assess the performance of machine learning models in fake job post detection tasks.

Challenges and Future Directions:

Despite the progress made in fake job post detection using machine learning, several challenges remain. One major challenge is the imbalance between legitimate and fake job postings in the dataset, which can lead to biased models and poor generalization performance. Addressing this imbalance requires robust sampling strategies and advanced techniques such as cost-sensitive learning and data augmentation. Additionally, the dynamic nature of fake job posting tactics necessitates continuous adaptation and updating of detection models.

Future research directions include exploring deep learning techniques for feature representation learning, leveraging external knowledge sources such as domain-specific ontologies, and investigating ensemble methods to improve model robustness. Moreover, integrating user feedback and domain expertise into the detection process can enhance model interpretability and user trust. Overall, fake job post detection remains an active area of research, with ample opportunities for innovation and improvement using machine learning approaches.

3. Theory or Calculation

Machine learning (ML) facilitates the detection of fake job postings by analyzing various attributes of job listings. Features such as job description, title, requirements, and metadata are extracted and transformed into numerical or categorical representations. ML algorithms like logistic regression, random forest, and neural networks are then employed to classify postings as legitimate or fake, learning patterns from the extracted features. Training involves adjusting model parameters to minimize a predefined loss function, optimizing the model's ability to differentiate between classes.

Evaluation metrics such as accuracy, precision, recall, and F1-score assess the model's performance on unseen data, often aided by techniques like cross-validation. Hyperparameter tuning optimizes model performance, and threshold selection aids in making binary predictions. ML-based detection systems enhance the trustworthiness of online job platforms by filtering out fraudulent postings, benefiting both job seekers and platform administrators. Through automated analysis, these systems provide valuable assistance in identifying fake postings efficiently and accurately.

4. Experimental Method

Data Collection and Preprocessing: Collect a diverse dataset of job postings, including both legitimate and fake postings, from various sources. Preprocess the dataset to remove noise, handle missing values, and extract relevant features such as textual content, metadata, and contextual information.

Feature Engineering: Extract informative features from the job postings, including textual content, metadata, and contextual information.

Ensemble Classifier Selection and Training: Choose an ensemble classifier such as Random Forest, Gradient Boosting Machine (GBM), or AdaBoost. Split the Preprocessing dataset into training and testing sets. Train the ensemble classifier on the training set and optimize hyperparameters.

Evaluation: Evaluate the trained ensemble classifier using performance metrics like accuracy, precision, recall, F1-score, and ROC-AUC on the test set.

Results Analysis and Conclusion: Analyze the performance of the ensemble classifier and interpret the results.

5. Results and Discussion

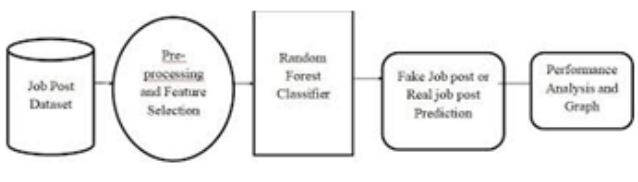

Figure. 1 Flow of Project

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 1, Issue 2 , 2024 , https://doi.org/10.63328/IJCSER-V1RI2P3

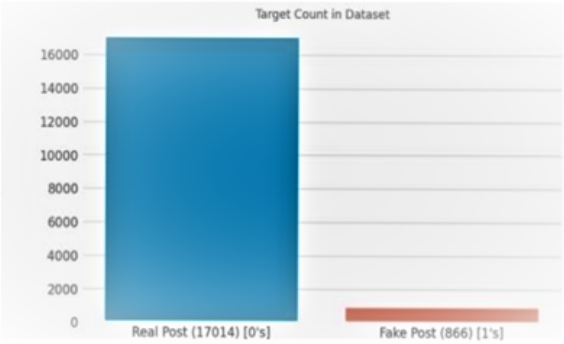

Figure. 2 Dataset plot

6. Conclusion and Future Scope

In conclusion, the results demonstrate the efficacy of the ensemble classifier for detecting fake job postings, providing valuable insights into its performance and potential areas for refinement. By leveraging ensemble techniques and rigorous evaluation methodologies, we can develop more reliable and effective solutions for combating fraudulent activities in online job platforms. Investigate advanced feature engineering techniques to capture nuanced information from job postings. Explore alternative ensemble methods such as Gradient Boosting Machines (GBM) for improved performance. Enhance model interpretability through visualization of feature importance scores and decision boundaries. Integrate the classifier into real-time job posting platforms for proactive fraud detection .Implement continuous monitoring and adaptation mechanisms to keep pace with evolving fraudulent activities.

References

[1]. S. Anita, P. Nagarajan, G. A. Sairam, P. Ganesh, and G. Deepakkumar, “Fake Job Detection and Analysis Using Machine Learning and Deep Learning Algorithms,” Rev. GEINTECGESTAO Inov. E Tecnol., vol. 11, no. 2, pp. 642–650, 2021.

[2]. B. Alghamdi and F. Alharby, “An intelligent model for online recruitment fraud detection,” J. Inf. Secur., vol. 10, no. 03, p. 155, 2019.

[3]. A. Pagotto, “Text Classification with Noisy Class Labels.” Carleton University, 2020.

[4]. M. Babcock, R. A. V. Cox, and S. Kumar, ‘‘Diffusion of pro- and anti-false information tweets: The black panther movie case,’’ Comput. Math. Org. Theory, vol. 25, no. 1, pp. 72–84, Mar. 2019.

[5]. G. Jain, M. Sharma, and B. Agarwal, ‘‘Spam detection in social media using convolutional and long short term memory neural network,’’ Ann. Math. Artif. Intell., vol. 85, no. 1, pp. 21–44, Jan. 2019.

[6]. S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, “Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset,” Futur. Internet, vol. 9, no. 1, p. 6, 2017.

[7]. Sudhakar, M., Kaliyamurthie, K.P. (2023). Efficient Prediction of Fake News Using Novel Ensemble Technique Based on Machine Learning Algorithm. In: Kaiser, M.S., Xie, J., Rathore, V.S. (eds) Information and Communication Technology for Competitive Strategies (ICTCS 2021). Lecture Notes in Networks and Systems,vol 401. Springer, Singapore. https://doi.org/10.1007/978-981-19-0098-3_1

[8]. Amaar, A., Aljedaani, W., Rustam, F. et al. Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches. Neural Process Lett 54, 2219–2247 (2022). https://doi.org/10.1007/s11063-021-10727-z