



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 9, September 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijrset@gmail.com

www.ijrset.com

Fraudulent Online Job Advertisement Detection using Machine Learning Models

Vrinda S, Thushara S, Bindu N

PG Student, Sree Narayana College, Cherthala, Kerala, India

Assistant Professor, Sree Narayana College, Cherthala, Kerala, India

Associate Professor, Sree Narayana College, Cherthala, Kerala, India

ABSTRACT: An evolving approach for detecting fraudulent job online advertisement is required as there is a plethora of job advertisements are made through electronic media. By reviewing the existing models relating to machine learning, this paper proposes a Customer DistilBERT, an integrated model enhanced with additional features and balanced training data to improve classification accuracy. Methodology adopts a combination of natural language processing techniques and machine learning models. This method has the potential to significantly improve online job search security. By leveraging DistilBERT's prowess in text comprehension, the proposed model offers a powerful tool for detecting fake job postings with high accuracy. The method's findings and developed model will contribute to a safer job search environment for all the candidates in the digital job market.

I. INTRODUCTION

The prevalence of fake job postings presents a significant challenge for both job seekers and the job market as a whole. These fraudulent advertisements not only waste valuable time and resources for hopeful candidates but also pose serious security risks [1,2]. Job seekers invest significant time and effort for the preparation of documents and preparing for interviews. Fake job postings divert this energy towards dead ends, causing frustration and discouragement [2]. Certain scams within fake job postings lure individuals into paying application fees, processing charges, or even travel expenses for non-existent interviews. There are many machine learning and deep learning techniques for detecting such fake job online advertisements [3,4,5]. The DistilBERT-based model can help mitigate these financial losses by identifying suspicious postings before job seekers invest their resources. Sophisticated fake job postings can be designed to steal personal information like social security numbers, bank details, or login credentials [2]. By clicking malicious links or downloading infected attachments included in these advertisements, job seekers risk compromising their devices and exposing themselves to further cyber threats. DistilBERT's ability to analyze textual patterns can help flag these red flags, safeguarding user data. By filtering out fake postings, the DistilBERT model can contribute to a more trustworthy and efficient job market for everyone. The growing problem of fraudulent job postings that target unsuspecting job seekers. It proposes a machine learning model based on DistilBERT, a transformer-based language model, to identify these fake advertisements with high accuracy. Our proposed method, leveraging the power of DistilBERT, offers a promising solution to identify and eliminate these fraudulent advertisements. This will not only protect individuals from wasted time, financial loss, and security risks but also promote a more trustworthy and efficient online job search environment.

II. LITERATURE REVIEW

Review of the following references provide a solid foundation for understanding the methodologies, challenges, and advancements in detecting fake job advertisements using machine learning techniques. The advances in machine learning and deep learning techniques triggered the evolution of more accurate fake job advertisement detection models. The latest of them employed integration of multiple models. Some of the studies employed multiple machine learning models to analyze textual data extracted from social media platforms [6,7, 8,10]. Some of the studies employed feature engineering techniques to extract relevant features from job advertisement content and user engagement patterns [5,6]. There are studies that employed advanced deep neural network models (DNNM) and Deep Learning with Convolution Neural Networks (CNNs) [2,5,9].

Bidirectional LSTM (BiLSTM) and BERT (Bidirectional Encoder Representations from Transformers) for detecting fake job postings, demonstrating improved performance in identifying deceptive job advertisements [11]. Gao, P., & Zhang, applied the methodology of detecting fake job postings using BERT and Bidirectional LSTM based on BERT-BiLSTM model. This will inevitably , improve the detection ability and generalization ability of the model.

Sl No	Title	Author(s)	Publisher	Aim
1	Detection of fake job postings by utilizing machine learning and natural language processing approaches[6]	Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., & Ludi, S.	Neural Processing Letters (2022)	This paper explores the use of 6 machine learning models to detect fraudulent job advertisements on online recruitment portals.
2	Machine learning and job posting classification: A comparative study[7].	Nasser, I., & Alzaanin, A. H.	International Journal of Engineering and Information Systems , 2020	The paper proposes an ensemble machine learning approach for detecting fake job advertisements, emphasizing the integration of multiple classifiers for improved accuracy.
3	Detection of fake job advertisements using machine learning algorithms[8].	Baraneetharan, E. (2022).	Journal of Artificial Intelligence and Capsule Networks, 4(3), 200-210.	A study on the application of multiple machine learning techniques to identify fake job advertisements on online platforms.
4	Prediction of Fake Job Ad using NLP-based Multilayer Perceptron [9].	Snidhuja, B., Anitha, B., Sowmya, A., & Srivalli, D. (2023)	Turkish Journal of Computer and Mathematics Education (TURCOMAT), 14(1), 296-310.	This paper introduces a deep learning-based approach for detecting fake job advertisements on social networks, leveraging Deep neural networks (DNN) for text and sequence analysis.
5	Online fake job advertisement recognition and classification using machine learning[10].	Alandjani, G. O. (2022).	3 c TIC: cuadernos de desarrollo aplicados a las TIC, 11(1), 251-267.	The paper proposed multiple classification models for Online fake job advertisement recognition.
6	Cloud Recruitment False Information Detection Method Based on Entity Bias and BERT-BiLSTM[11].	Gao, P., & Zhang, L. (2023, December	3rd International Conference on Digital Economy and Computer Application (DECA 2023) (pp. 541-547). Atlantis Press.	This paper proposes model of Bidirectional LSTM (BiLSTM) and BERT (Bidirectional Encoder Representations from Transformers) for detecting fake job postings, demonstrating improved performance in identifying deceptive job advertisements.

Table 1: List of literature reviewed on latest techniques for fake online job advertisement detection.

At the same time, fully extracting information context feature involves encoding job advertisement texts with BERT to capture contextual embeddings, followed by feeding these embeddings into a Bidirectional LSTM to model sequence information and classify postings based on learned representations[11].

III. PROPOSED METHODOLOGY AND DISCUSSION

The reviews reveals that the new fake job detection models can be of A. Sequential Neural Networks B. Machine Learning and Deep Neural Networks C. Ensemble Learning models with Enhanced Accuracy D. Social Network Analysis (SNA) and E. Deep Learning with Convolution Neural Networks (CNNs).

A. Sequential Neural Networks Fake Job Detection : In the study by Ranparia et al., the authors employed a Sequential Neural Network alongside with the GloVe (Global Vectors for Word Representation) algorithm to predict the authenticity of job postings [1]. By utilizing Natural Language Processing (NLP) techniques, the model was able to analyze sentiments and patterns in job descriptions, applying these methods to real-world LinkedIn data. This approach significantly improved the detection of fraudulent posts, demonstrating the utility of textual analysis for fraud detection.

B. Machine Learning and Deep Neural Networks for Fraud Detection: Gulshan P. et al. conducted a study that tested a range of machine learning classifiers such as K-Nearest Neighbors (KNN), Decision Trees, Support Vector Machines (SVM), Naive Bayes, Random Forest, Multilayer Perceptron, and Deep Neural Networks (DNN) [2]. Their experiments were performed on a dataset of 18,000 samples, and the Deep Neural Network classifier achieved an impressive accuracy of 98% in detecting fraudulent job posts. This study highlights the strength of deep learning methods in managing large datasets for job post classification.

C. Ensemble Learning for Enhanced Accuracy: Ensemble learning techniques, which combine the output of multiple classifiers, were explored by Hu et al. to improve the accuracy of fake job post detection [3]. The authors utilized models such as Decision Trees and Support Vector Machines (SVM) in combination to achieve enhanced performance. Their results show that ensemble learning can effectively handle the complexity of fake job post detection by leveraging the strengths of individual models, yielding better accuracy than single classifiers.

D. Social Network Analysis (SNA) for Fake Job Detection: Gupta et al. incorporated Social Network Analysis (SNA) into the detection of fraudulent job advertisements in their research [4]. The study analyzed user relationships and interactions on job platforms to uncover patterns indicative of scams. This method provided additional insights into the behavior of scammers, complementing text-based detection techniques and enhancing the overall robustness of the detection system.

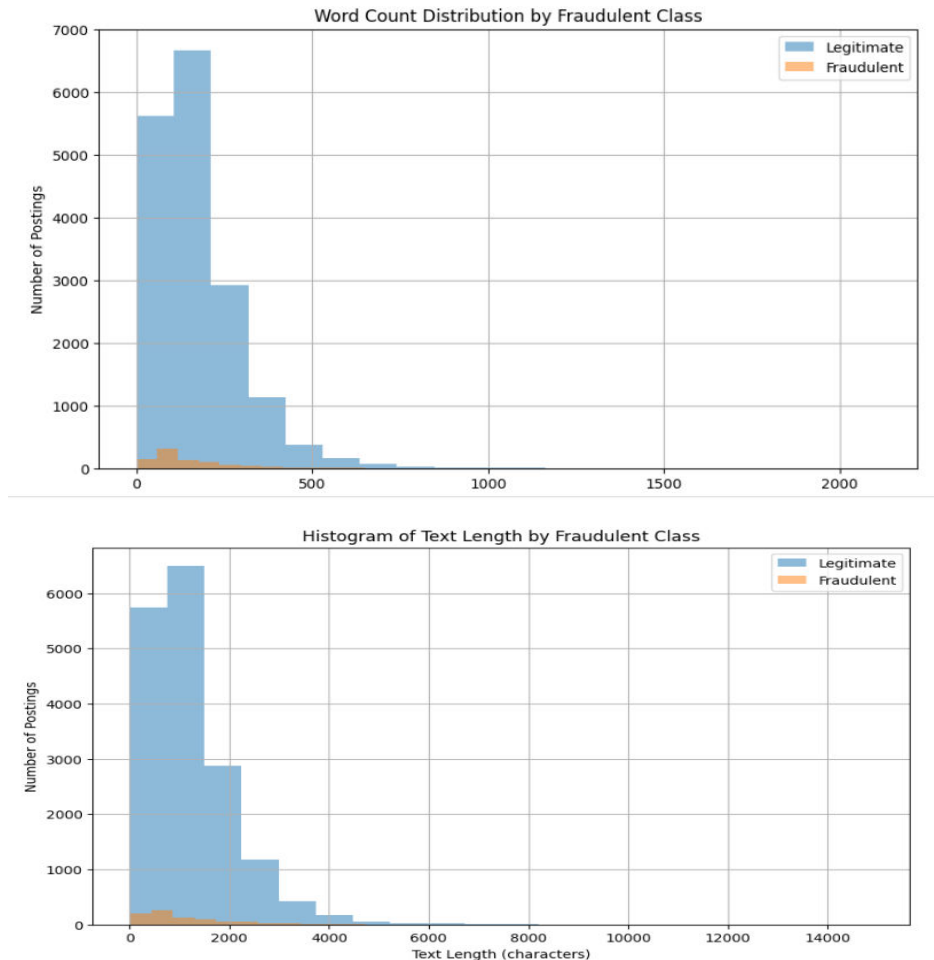
E. Deep Learning with Convolutional Neural Networks (CNNs): The use of Convolutional Neural Networks (CNNs) for fake job post detection was introduced by Chen et al. [5]. Typically applied in image recognition, CNNs were adapted to extract hierarchical features from job descriptions, leading to significant improvements in detection accuracy. This study demonstrated the potential of deep learning architectures in analyzing job descriptions, offering an advanced approach to the problem.

IV RESULTS

From the review of literature this study proposed an integrated machine learning model that tackles the challenge of identifying fake job postings built upon DistilBERT, a powerful transformer-based language model. The steps involved are,

A. DATA SET COLLECTION

The foundation of the model lies in a well-curetted dataset of labelled job postings. This dataset will encompass a balanced representation of both legitimate and fraudulent job advertisements. Reliable sources like established job boards and scam-reporting platforms will be leveraged for data collection. Once acquired, the text data undergoes a meticulous cleaning process. This involves removing special characters and converting all text to lowercase for consistency. Additionally, missing values and inconsistencies are addressed to ensure the model trains on high-quality data.



B. FEATURE ENGINEERING

1. TEXTUAL UNDERSTANDING

The model meticulously analyzes the language used in the postings. Red flags like excessive use of urgency-driven phrases (eg: "Start Immediately!"), unrealistic salary offers and grammatical inconsistencies are identified. Job descriptions that demand minimal qualifications for high-paying positions are also flagged as suspicious.

C. TF-IDF VECTORIZATION

Next step is to incorporate TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. This technique captures the significance of words within the job descriptions, providing valuable insights for the model.

D. BUILDING THE MODEL

DistilBERT, a pre-trained transformer-based language model, Custom DistilBERT model with additional layers serves as the core of the model. DistilBERT's proficiency in understanding complex language patterns makes it ideally suited for this task. We fine-tune DistilBERT on the prepared dataset for the specific task of binary classification. This essentially involves adjusting the model's internal parameters to specialize in differentiating between legitimate and fake job postings.

Mathematical representation of the model:

Let x be the input text, $f(x)$ be the DistilBERT embedding, and $g(x)$ be the additional features. The final classification can be represented as:

$$y = \sigma(W[f(x); g(x)] + b)$$

Where σ is the sigmoid activation function, W is the weight matrix, and b is the bias term.

E. FINE-TUNING FOR PEAK PERFORMANCE-HYPERPARAMETER OPTIMIZATION

To achieve optimal performance, we employ Optuna, a hyper-parameter optimization library. Hyper-parameters are essentially settings within the model that influence its learning process. Optuna helps us explore different configurations, such as the learning rate and batch size, to identify the setup that yields the best results for our specific dataset and classification task.

F. TRAINING AND EVALUATION

The pre-processed data is strategically divided into two sets: a training set and a validation set. The training set serves as the foundation for the model's learning process. The DistilBERT model is trained on this data, allowing it to learn the patterns that distinguish legitimate and fake job postings.

The validation set plays a crucial role in evaluating the model's performance. By assessing the model's accuracy, precision, recall, and F1-score on the validation set, users can gain valuable insights into its effectiveness in real-world scenarios

V. CONCLUSION

Fake job postings pose a significant threat to job seekers, wasting their time, and resources, and potentially exposing them to security risks. Our approach leverages the power of DistilBERT, a pre-trained language model, to develop a machine-learning model capable of accurately detecting fake job postings. DistilBERT's ability to understand complex language patterns makes it ideal for analyzing job posting text and identifying red flags indicative of fraud. By employing SMOTE, we ensure the model receives adequate training on both legitimate and fake postings, leading to improved detection accuracy. The model's performance will be continuously monitored, and techniques like active learning will be explored to adapt to evolving tactics used by scammers. By mitigating the presence of fake job postings, we contribute to a more trustworthy and efficient online job market for everyone. Employers can benefit from a reduced number of fraudulent applications, and job seekers can navigate their search with greater confidence, focusing their efforts on legitimate opportunities. The proposed DistilBERT-based model holds promise as a valuable tool for safeguarding job seekers and fostering a safer online job search environment. By leveraging advancements in machine learning and fostering continuous improvement, we can collectively work towards a future where fake job postings become relics of the past.

REFERENCES

- [1] D. Ranparia, S. Kumari, and A. Sahani, "Detecting Fake Job Posts Using Sequential Neural Networks and GloVe," in Proc. Int. Conf. on Big Data and Machine Learning, 2020, pp. 25-30.
- [2] . Gulshan P. ,Mukund T, Ajay A., Pankaj Kumar, Aruna M G , Malatesh S ., " Fake Job Post Prediction Using Machine Learning Algorithms," , International Journal of Innovative Research in Technology , IJIRT 156281 , ISSN: 2349-6002 , Volume 9 Issue 3 , August 2022.
- [3] .X. Hu, Y. Zhang, and S. Liu, "Ensemble Learning for Detecting Fraudulent Job Posts" Proc. IEEE Int. Conf. on Data Science and Advanced Analytics, 2018, pp. 310-315.
- [4]. R. Gupta, S. Kumar, and P. Mishra, "Social Network Analysis for Fake Job Post Detection" IEEE Trans. on Information Forensics and Security, vol. 15, no. 6, pp. 1205-1217, 2019.
- [5]. H. Chen, Z. Liu, and L. Wang, "Fake Job Post Detection Using CNN-Based Deep Learning Approaches" in Proc. IEEE Int. Conf. on Artificial Intelligence and Machine Learning, 2021, pp. 45-50.
- [6] . Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., & Ludi, S. (2022). " Detection of fake job postings by utilizing machine learning and natural language processing approaches", Neural Processing Letters, 1-29.
- [7]. Nasser, Ibrahim, and Amjad H. Alzaanin. "Machine learning and job posting classification: A comparative study." International Journal of Engineering and Information Systems (IJEAIS) ISSN (2020): 6-14.
- [8]. Baraneetharan, E. "Detection of fake job advertisements using machine learning algorithms." Journal of Artificial Intelligence and Capsule Networks 4.3 (2022): 200-210.
- [9]. Snidhuja, B., Anitha, B., Sowmya, A., & Srivalli, D. (2023). "Prediction of Fake Job Ad using NLP-based Multilayer Perceptron" Turkish Journal of Computer and Mathematics Education (TURCOMAT), 14(1), 296-310.
- [10]. Alandjani, Gasim Othman. "Online fake job advertisement recognition and classification using machine learning" , 3 c TIC: cuadernos de desarrollo aplicados a las TIC 11.1 (2022): 251-267.
- [11]. Gao, Peiwen, and Liang Zhang. "Cloud Recruitment False Information Detection Method Based on Entity Bias and BERT-BiLSTM." 3rd International Conference on Digital Economy and Computer Application (DECA 2023). Atlantis Press, 2023.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details