2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) | 979-8-3315-3366-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICDCECE65353.2025.11035352

Comprehensive Approach to Fraudulent Job Post Detection Using Machine Learning and BERT Models

Sri Sai Suhas Sanisetty
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
ssuhaas24@gmail.com

Geetaa Namrithaa S
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
geetaanamrithaa@gmail.com

Surya Vamshi Kotamaraja
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
Surya.kotamraja@gmail.com

Bommu Nagendra Reddy
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
reddynagendra904@gmail.com

Susmitha Vekkot
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
v_susmitha@blr.amrita.edu

Bhavana V
Department of Electronics and Communication Engineering
Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
v_bhavana@blr.amrita.edu

***Abstract*—In recent times there has been a significant rise in fraudulent job postings posing risks to job seekers and recruitment platforms. The paper presents a framework combining Machine Learning (ML) and advanced Natural Language Processing (NLP) techniques to detect the fake job postings. The model uses data that has been integrated from multiple sources and uses traditional ML models and NLP models. Feature extraction methods such as fraud keyword detection and sentiment polarity analysis were done using TextBlob and VADER to enhance interpretability. Results show that the BERT model has achieved 100% accuracy by outperforming the other models and also shows the system's ability to differentiate between real and fake job postings. Comparative analysis was done to show the superiority of deep learning models over the traditional approaches. The system provides a robust solution to enhance job security, prevent cyber fraud, and assist recruitment platforms in identifying fraudulent job postings.**

***Index Terms*—Fake Job Detection, Machine Learning, NLP, BERT, Data Imbalance, Classification Metrics.**

I. Introduction

The rise of online job platforms has made job seekers and employers convenient to find or post jobs. But, due to the rise of this digital transformation, it has led to a significant increase in fake job postings and poses risks to individuals or to an organization. Scammers use these types of platforms to trick the applicants, which may lead to financial loss and identity will be compromised [1]. There are a few traditional methods that are used to detect fake job postings, like relying on manual reviews and rule-based systems. These methods are very time consuming and are not effective as the scammers periodically change their tactics. To improve the detection process, Machine learning (ML) models and Natural Language processing techniques (NLP) are used in recent studies which have shown that ML algorithms can easily identify deceptive patterns in job advertisements and it is very effective [2]. One advanced NLP technique is BERT (Bidirectional Encoder Representations from Transformers) [3]. It is a very powerful tool and is highly effective for text classification tasks as it understands the context of the words by just looking at the words before and after them and this enables a deeper understanding of linguistic nuances. This ability to capture the context from both the directions will help BERT to detect any subtle indications of fraud content in job postings. Incorporating sentiment analysis provides a further insight into the emotional tone of the job descriptions. Sentimental analysis will measure the polarity of the text and classify them as positive, negative or neutral. This type of approach has been used effectively in various fields to assess the credibility of the textual content [4]. This article aims to develop a robust and comprehensive framework to detect the fake job postings by integrating machine learning (ML) models and NLP techniques like BERT based text classification with sentiment and emotion analysis. In this study we have created a dataset for fake job postings, then proceeded with preprocessing the text data and at last extracted the relevant features. Subsequently, a BERT model is fine-tuned on the dataset to classify if the postings are fraud or legit. The performance of the model is measured by performance metrics such as accuracy, precision, recall, and F1 score. The main contributions are:

- Developed a robust dataset for job postings with labeled fraudulent and legitimate entries.
- Implemented Machine Learning models and BERT based classification model which is enhanced with sentiment

and emotion analysis.
- Provided a comparative analysis of the model's performance against four traditional ML algorithms.

The work process is divided into different sections, Section II provides an overview of recent works in fake job posts detection and their limitations. Section III gives the details of the proposed model including data collection, Section IV focuses on its implementation and results. Finally, Section V concludes the paper, offering insights into potential directions for future research.

II. Related Works

This section discusses about the previous works that have been implemented on fake job postings' detection. The authors in [5] proposed an ensemble model by combining ML models like Random Forest, Logistic Regression, and Extra Trees with feature selection techniques like PCA and Chi Square and used SMOTE for balancing the dataset. They achieved accuracy of 99% and demonstrated robust performance but limits real time applications. The authors in [6] utilized Gradient Boosting with features derived from bag of words, word embeddings which helps to classify fraud job types like identity theft and fraud schemes. The model achieved an F1 score of 0.88 and highlights the importance of feature interpretability. The authors in [7] emphasized feature selection and dataset balancing through SMOTE and also combined Logistic regression, random forest, and extra trees. The model achieved 99% accuracy but still requires significant computational resources. The authors in [8] introduced a ML based web tool that uses URL based web scraping for detection of fake job postings. It provided a user-friendly interface and real time detection but its reliance on URL validation limited its text based fraud detection. The authors in [9] used the POESM framework and tested 8 classifiers and found that random forest performs the best in terms of accuracy in precision. Lack of use of NLP techniques constrained models' ability to capture contextual nuances. The authors in [10] conducted a comparative analysis of classifiers like KNN, decision trees, SVM, and DNN. The model performed well while using DNN and achieved 98% accuracy on EMSCAD dataset. The authors in [11] proposed a hybrid deep learning model by combining Bi-LSTM and Bi-BRU with a focal loss function to tackle class imbalance. The model achieved an accuracy of 98.6% with balanced precision and recall but still required significant computational resources, limiting its scalability in real time scenarios. The authors in [12] used LSTM and GRU to capture temporal patterns in a text. The model achieved good accuracy but with class imbalance. The paper highlighted the importance of contextual understanding in tasks related to fraud detection. Research on detection of fake job postings has evolved from traditional models like random forests to advanced deep learning approaches such as LSTMs, GRUs, BERT and few attention mechanisms and improved contextual understanding and addressed the class imbalance. Hybrid models containing combination of deep learning models with ensemble methods has achieved better accuracy and efficiency than others. Few significant advancements' also have been made but challenges like generalization across the datasets and real time detection remain, which highlights the opportunities for further innovation in fraud detection systems.

III. Proposed Model

The pipeline of the proposed model is shown in the Fig.1

A. Dataset Collection and Creation

The paper utilizes three datasets and two of them are publicly available in Kaggle and one dataset was created manually. The custom dataset was generated using Python by manually giving input job posting recruitment messages that we get through messages and online job portals. Libraries such as pandas were used for data handling and structuring. Ensured to include essential features such as job title, description, requirements, company, location, and employment type to improve overall efficiency of the model. At the final stage, the three datasets were combined to enhance models' performance and ensured providing a balanced representation of fake and real job postings. Table I gives us insights of the dataset that have been combined.

TABLE I: Dataset Description for Fake Job Post Detection

Attribute	**Description**
Total no of job postings	151,729
No of job postings (Train)	121,383
No of job postings (Test)	30,346
No of fraudulent job postings (Train)	60,692
No of fraudulent job postings (Test)	15,173
No of legitimate job postings (Train)	60,691
No of legitimate job postings (Test)	15,173
Dataset source	EMSCAD, LinkedIn Postings, and Custom Created Dataset

B. Data Preprocessing and Feature Engineering

To prepare the dataset for modelling, few various data preprocessing steps have been implemented, like handling missing values and encoding categorical variables which help to ensure the machines' readability. Feature engineering was also conducted to identify the important features. Features such as fraud related keywords and sentiment polarity were based on their relevance to the job content and they are very effective in distinguishing fraud postings. TextBlob and VADER sentiment analysers have been used to determine sentiment polarity and emotion categories.

C. Machine Learning models

Few Classical ML models like Logistic Regression, Random Forest, SVC, and XGBoost, were employed.

- Logistic regression: Calculates the probability of the job posting if it's fraudulent or not using a sigmoid function.
- Random Forest: It is a combination of multiple decision trees which helps to improve the prediction accuracy through bagging. The final prediction is made by majority voting among the trees.

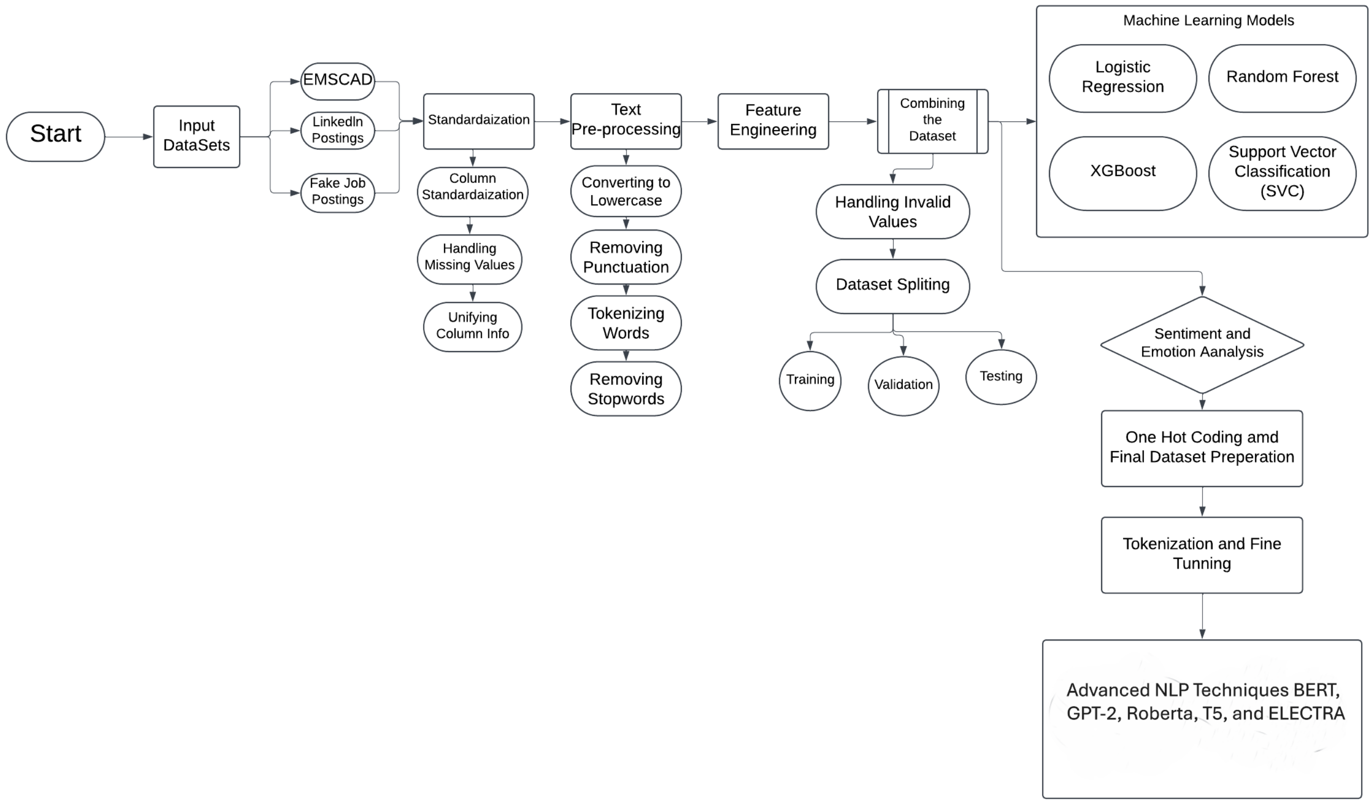

Fig. 1: Pipeline of the Proposed Model

- SVC (Support Vector Classifier): This model finds the hyperplane that maximizes the margin between the classes.
- XGBoost model: It's a gradient boosting model which optimizes the objective function.

All these models were trained on the combined dataset and also did hyper parameter tuning for better optimization.

D. Sentiment and Emotion Analysis

To achieve better results, sentiment and emotion categories (positive, negative, and neutral) had been extracted from the job descriptions using TextBlob and VADER sentiment analyzers. TextBlob is useful for handling context specific job descriptions and is very effective in capturing the overall sentiment tone. VADER is rule based and it is very effective in identifying sentiment in short job descriptions with domain specific terms. Then, proceeded with One Hot Encoding of the emotion categories.

E. Advanced NLP Techniques

Used Fine tuned transformer models, like BERT and GPT-2 and used their contextual understanding capabilities for detecting fraudulent job postings [13].

- BERT (Bidirectional Encoder Representations from Transformers): It uses a transformer architecture to understand the bidirectional context in a text. It tokenizes the input sequence, and embeddings are generated using:

$$h_0^l, h_1^l, \ldots, h_n^l = \text{Transformer}(E_0, E_1, \ldots, E_n) \quad (1)$$

Here, E_i represents token embeddings, and h_i^l denotes the output at the l-th layer. The final output embeddings are passed to a classification head

$$P(y|X) = \text{Softmax}(W \cdot h_n^L + b) \quad (2)$$

Here, W is the weight matrix, b is the bias term, and L is the total number of transformer layers.

- GPT-2 (Generative Pretrained Transformer 2): It uses a unidirectional transformer to generate context aware embeddings. The input sequence is tokenized, and all the hidden states are computed in an iterative manner.

$$h_t = \text{Transformer}(h_{t-1}, E_t) \quad (3)$$

Where E_t is the embedding of the t-th token, and h_t represents the hidden state at step t.
GPT-2 generates probabilities for the next work using:

$$P(w_t|w_{<t}) = \text{Softmax}(W_o \cdot h_t + b_o) \quad (4)$$

Here, W_o is the output weight matrix, and b_o is the bias term.

F. Model Fine tuning and Evaluation

The tokenized dataset is split into training and validation sets. The pre trained weights of BERT and GPT-2 had been utilized and then fine-tuned the model on the dataset with a classification head. Loss functions, such as Cross Entropy Loss, are minimized to optimize the models' performance. The overall evaluation is performed using metrics like accuracy, precision, recall, and F1 score.

IV. Results and Discussion

For the implementation of the proposed model for detecting fake job postings we have used Python 3, with a combination of advanced Machine Learning (ML) and Natural Language Processing (NLP) techniques. This project utilizes three datasets those are the EMSCAD dataset [14], LinkedIn postings between 2023 to 2024 dataset, and finally the Fake Postings dataset. First two datasets have been directly downloaded from Kaggle but the third dataset was created by us using python libraries. To ensure consistency, the missing values in all the critical columns have been replaced with predefined values. Few columns have been standardized by combining two or more columns into a single unified column. Text columns such as description and requirements has undergone extensive preprocessing, which includes converting text to lowercase, removing the punctuations and stop-words, tokenizing the words, and also reconstructed them into a clean string. By doing this cleaning process it reduced noise and enhanced the quality of the data for downstream tasks. Then at the last we have combined all the three datasets and made an enhanced dataset which had 151729 samples.

Feature engineering had played a very important role in enriching the dataset with few additional contexts. Identified the most important features in description column and it helped to capture meaningful patterns. Fig.2 shows the Important Feature analysis.

- Description length: It indicates the total number of characters in a job description.
- Word Count: It shows the total number of words present in the job description and its importance to provide insights into the verbosity of job postings.
- Sentiment polarity: It helps to capture the tone of job descriptions.
- Fraud Keywords: It is strongly correlated with fraudulent postings.

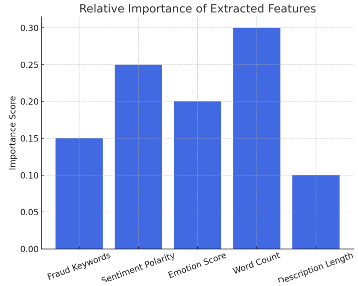

Fig. 2: Important Feature Analysis

To provide valuable insights into nature of job postings we have used sentiment and emotion analysis [15]. TextBlob has been used to calculate the polarity of a sentiment by producing scores ranging from -1 (negative sentiment) to 1(positive sentiment). Also used VADER (Valence Aware Dictionary and Sentiment Reasoner) it is rule based sentiment analysis tool in the Natural Language Toolkit (NLTK) library that is used to analyze the sentiment of a text. It basically uses a dictionary of rules and words to determine a sentiment for a piece of text. So, it categorizes the descriptions into emotional classes. Finally did OHE (One Hot Encoding) it is a machine learning technique that converts all the categorical data into numerical values so that it can be used efficiently in machine learning algorithms and at last a final combined data set was created successfully. Fig. 3 shows the sentiment and emotional analysis output.

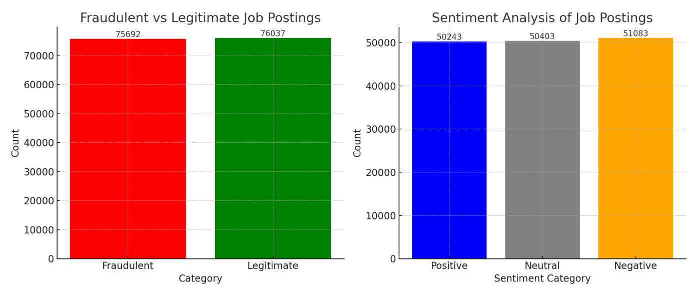

Fig. 3: Sentiment and Emotion Analysis

Multiple Machine learning models have been used for binary classification to detect fraudulent job postings. Logistic Regression, Random Forest, XGBoost, and Support vector classifier (SVC) have been used and random forest has obtained the highest accuracy (99.93%) among these models. Random Forest is an ensemble learning model and it builds multiple decision trees and aggregates its outputs and it can also handle imbalanced datasets very well and it is also robust to over fitting due to its averaging mechanism. Hyper parameter tuning was also conducted on Random Forest Model using GridSearchCV method. Parameters like number of estimators, maximum tree depth, minimum samples required to split a node, and minimum samples per leaf were optimized. The tuned Random Forest model showed little bit more improvement in precision, recall, and F1 score. However, ML models' often show lower recall compared to deep learning techniques. This is due to their reliance on manual engineered features, which may struggle to capture few linguistic patterns and contextual dependencies present in the job postings. Models like Random Forest and SVC they work perfectly with structured data but they struggle little bit across complex, unstructured data. To enhance the performance, deep learning models were used to evaluate the models' performance. First, we went with the state-of-the-art model BERT it is an NLP model which was developed by google and it learns to understand the context of unlabeled text. So, to understand the contextual relationships within a text we have used Hugging Face's (bert-based-uncased) transformer model. It is basically a bidirectional transformer model that analyzes words in text by considering contexts of both left-to-right and right-to-left. This approach has made analyzing the job descriptions more effective. The BERT tokenizer was applied to the description column with a maximum sequence length of 128, and also ensured the compatibility of the model. We went with two approaches while doing BERT one is Pre-Fine-Tuning performance and another one is Post-Fine-Tuning-Performance. Before fine tuning the BERT model, the model failed to perfectly identify the fraudulent job postings (minority class). The model has overall achieved an accuracy of 94%, but the precision and recall scores were skewed towards the majority class. After Fine tuning, by integrating sentiment and emotion analysis. The model had nearly achieved perfect performance. Utilized GPT-2 (Generative Pre-Trained Transformer 2) model, this model is also a state of the art generative model which processes the text in a sequential manner and also it uses deep learning techniques to predict the next word in a sequence. It is an auto regressive model which uses a function called softmax to estimate the probability distribution of each word in a sequence. It has a unidirectional architecture and has generative capabilities which is used for tasks that require an understanding of sequential patterns,such as detecting formulaic language associated with the fraudulent postings. Table II shows us the classification report for Pre and Post-Fine-Tuning performance of both BERT and GPT-2 which highlights the need of additional features and fine tuning to address the class imbalance. Table III shows the comparison of results between the NLP techniques.

TABLE II: Pre and Post Fine-Tuning Performance Metrics

Model	Fine-Tuning	Class	Precision %	Recall %	F1-Score %
BERT	Pre	0 (Legitimate)	94	100	97
	Pre	1 (Fraudulent)	98	21	34
	Post	0 (Legitimate)	100	100	98
	Post	1 (Fraudulent)	100	97	98
GPT-2	Pre	0 (Legitimate)	92	90	91
	Pre	1 (Fraudulent)	91	25	40
	Post	0 (Legitimate)	99	100	100
	Post	1 (Fraudulent)	94	92	93

There is a slight trade-off between accuracy, precision, and recall. ML models such as Logistic regression and Random Forest shows high accuracy and precision but exhibits lower recall, which indicates their limited capability in detecting fraud job postings and these models are more prone to false negatives. While deep learning models showcased a better balance between precision and recall after fine tuning. The BERT model achieved perfect accuracy and precision, but it required more computational resources and longer training times compared to other models. GPT-2 model was slightly less accurate than BERT but it performed well in recall, which makes it suitable for identifying fraud cases effectively.

TABLE III: Performance Metrics for all the models

Model	Accuracy %	Precision %	Recall %	F1-Score %
Logistic Regression	99.67	95.18	100	97.53
Random Forest	99.93	98.92	99.95	99.26
XGBoost	99.83	97.62	99.85	98.73
SVC	99.25	89.72	100	94.58
Fine-Tuned BERT	100	100	97	98
Fine-Tuned GPT-2	99.7	94	98	93
Fine-Tuned RoBERTa	99.95	99	98	99
Fine-Tuned T5	99.85	97	96	97
Fine-Tuned ELECTRA	99.9	98	97	98

Finally, the model is tested in two approaches. The user can give the input text or the user can provide the input image. Depending on the input the model will let us know if the given input is a fake job posting or a real job posting. For images we use OCR (Optical Character Process) [16] so it converts an image text into a machine readable text format. Once we extract the text from the image, then we can proceed with output results based on the extracted test and the code will show if the extracted text from the image is fake or real. Fig.4 shows the message used and Fig.5 shows us the extracted text from image after using OCR. Fig. 6 shows if the content in image is a fake job posting or a real one.

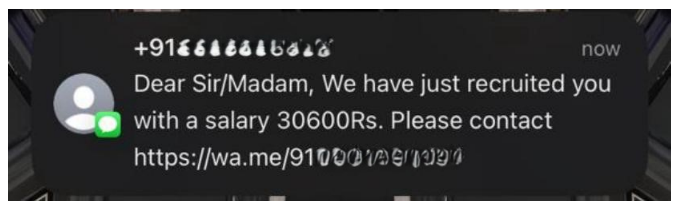

Fig. 4: Message Image used as Input

```
Processing image file...
Extracted Text: +91[illegible] now
Dear Sir/Madam, We have just recruited you
with a salary 30600Rs. Please contact
https://wa.me/91[illegible]
```

Fig. 5: Extracted text from the input message image

```
Prediction for the image content: Fake Job Posting
```

Fig. 6: Output shown after extracting the text from the input image

The minimum system requirements for implementing the proposed model include an Intel Core i5 processor, 8 GB RAM, and an NVIDIA GTX 1650 GPU with 4 GB VRAM.

The model is implemented using Python 3.8+, with essential libraries such as TensorFlow, PyTorch, and Hugging Face Transformers. The system should run on Windows 10 or Ubuntu 18.04 LTS for optimal performance. Table IV shows the comparison between the State of the Art Models (SOTA)

TABLE IV: Comparison of SOTA Models for Fake Job Post Detection

Year	Model/Technique	Accuracy/Metric	Dataset
2022	Machine Learning Algorithms Tailored for Online Platforms	Not Explicitly Mentioned	(EMSCAD) dataset
2023	Comparative Study Using SVM, Naïve Bayes for Fake Job Post Prediction	Varied Based on Algorithm	(EMSCAD) dataset
2023	Sequential Networks (LSTM, GRU) for Fake Job Prediction	Improved Reliability Metrics	(EMSCAD) dataset
2024	Bi-LSTM and Bi-GRU Architectures with Focal Loss for Class Imbalance Handling	98.6% Accuracy, 85% F1-Score	(EMSCAD) dataset
2024	BERT-Based Model Enhanced with Sentiment and Emotion Analysis	98.6% Accuracy, 85% F1-Score	Employment Scam Aegean Dataset (EMSCAD)
2024	**Proposed Model**	**99.9%Accuracy, 98% F1-Score**	**Custom Combined Dataset**

V. Conclusion and Future Scope

The detection of fake jobs uses deep learning models such as BERT, GPT-2, etc. to process job descriptions and mark them authentic or fake. The systematic feature engineering techniques such as sentiment polarity analysis, word count, and description length improved interpretability and efficiency. This better accuracy, precision, recall, and F1 score were attained with the fine-tuning of pre-trained transformer models, and they proved to work in the detection of deceptive patterns from job-postings Because of this emotion analysis and one hot encoding, the dataset was enriched and granularity was achieved in the classification process. Then again, a comparative evaluation between BERT and GPT-2 was carried out on their performance regarding efficiency the former being superior for precision and recall.However, the model's performance depends on quality and diversity of the dataset provided, which may not fully represent emerging scam tactics, and also relies heavily in text-based features, limiting models' applicability to multimedia job postings that include images or videos. The future scope of this work is to expand the dataset concerning different job domains and languages to make the model more generalized for use across different contexts. More multi-modal sources, for example, reviewer ratings given by employers or feedback given by users on jobs, would produce a better outcome. Experiments could be carried out on lightweight transformer models for improving the system in an environment that has resource limitations like mobile applications. Streaming data coupled with real-time fraud detection is another interesting option that will allow modifications and prompt feedback to users.

References

[1] T. Bhatia and J. Meena, "Detection of Fake Online Recruitment Using Machine Learning Techniques," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 300-304, doi: 10.1109/ICAC3N56670.2022.10074276.

[2] B. Pandey, T. Kala, N. Bhoj, H. Gohel, A. Kumar and P. Sivaram, "Effective Identification of Spam Jobs Postings Using Employer Defined Linguistic Feature," 2022 1st International Conference on AI in Cybersecurity (ICAIC), Victoria, TX, USA, 2022, pp. 1-6, doi: 10.1109/ICAIC53980.2022.9897059.

[3] Pourkeyvan, Alireza, Ramin Safa, and Ali Sorourkhah. "Harnessing the power of hugging face transformers for predicting mental health disorders in social networks." IEEE Access 12 (2024): 28025-28035.

[4] Pillai, Aravind Sasidharan. "Detecting Fake Job Postings Using Bidirectional LSTM." arXiv preprint arXiv:2304.02019 (2023).

[5] Afzal, Hina, et al. "Identifying fake job posting using selective features and resampling techniques." Multimedia Tools and Applications 83.6 (2024): 15591-15615.

[6] Naudé, Marcel, Kolawole John Adebayo, and Rohan Nanda. "A machine learning approach to detecting fraudulent job types." AI SOCIETY 38.2 (2023): 1013-1024.

[7] Anbarasu, V., S. Selvakani, and Mrs K. Vasumathi. "Fake Job Prediction Using Machine Learning." ubiquity 13.1 (2024): 12-20.

[8] C. Prashanth, D. Chandrasekaran, B. Pandian, K. Duraipandian, T. Chen and M. Sathiyanarayanan, "Reveal: Online Fake Job Advert Detection Application using Machine Learning," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-6, doi: 10.1109/DELCON54057.2022.9752784.

[9] B. P. Prathaban, S. Rajendran, G. Lakshmi and D. Menaka, "Verification of Job Authenticity using Prediction of Online Employment Scam Model (POESM)," 2022 1st International Conference on Computational Science and Technology (ICCST), CHENNAI, India, 2022, pp. 1-6, doi: 10.1109/ICCST55948.2022.10040305.

[10] Habiba, Sultana Umme, Md Khairul Islam, and Farzana Tasnim. "A comparative study on fake job post prediction using different data mining techniques." 2021 2nd international conference on robotics, electrical and signal processing techniques (ICREST). IEEE, 2021.

[11] Akhila, K., et al. "Improving Online Job Authenticity Detection Using Deep Learning and Focal Loss." 2024 International Conference on Data Science and Network Security (ICDSNS). IEEE, 2024.

[12] Ranparia, D., Kumari, S. and Sahani, A., 2020, November. Fake job prediction using sequential network. In 2020 IEEE 15th international conference on industrial and information systems (ICIIS) (pp. 339-343). IEEE.

[13] P. Vancha, H. Nagarajan, V. S. Inakollu, D. Gupta and S. Vekkot, "Word-Level Speech Dataset Creation for Sourashtra and Recognition System Using Kaldi," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039985.

[14] H. C. Dasari and V. Bhavana, "Enhancing Emotional Insight: The Benefits of Integrating Various Approaches for Better Emotion Recognition," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 1689-1698, doi: 10.1109/ICAAIC60222.2024.10575686.

[15] A. Swathi, S. Menon and V. Bhavana, "Impact of image size on Human Facial Expression Recognition: A Relative Study," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 687-692, doi: 10.1109/ICCSP60870.2024.10543518.

[16] R. Ramanathan, S. Ponmathavan, N. Valliappan, L. Thaneshwaran, A. S. Nair and K. P. Soman, "Optical Character Recognition for English and Tamil Using Support Vector Machines," 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, Bangalore, India, 2009, pp. 610-612, doi: 10.1109/ACT.2009.155.