

DETECTING FAKE JOB POSTINGS USING TRANSFORMER-BASED LANGUAGE MODELS BERT AND ROBERTA

Chandra Sekhar Sanaboina

Department of Computer Science and Engineering, University College of Engineering Kakinada, India

Abstract

Online recruitments have triggered a massive crime problem of online recruitment scams with many people posting their job online. Fraudsters take advantage of the situation and use deceptive job adverts and steal information or get money. The first challenge is the imbalance of classes, and existing systems utilize classical algorithms of machine learning instead of which there are boundaries. These restrictions on fake job ads see a significant rise in imbalanced forecasts. These weaknesses are addressed in the proposed method which uses deep learning models such as BERT and RoBERTa and special datasets, such as unrealistic job postings and Pakistani and US job postings. Also, the proposed solution uses SMOTE (Synthetic Minority Oversampling Techniques) to overcome so-called class imbalance in the dataset. The intended method aims at preventing biased predictions and maximizing accuracy and recall.

Keywords:

Deep Learning, Natural Language Processing, Bidirectional Encoder Representations from Transformer, Robustly Optimized BERT, Synthetic Minority Oversampling Technique, Synthetic Minority Oversampling Borderline

1. INTRODUCTION

The fact is that the hiring process is now easier and efficient with the introduction of online recruiting. Avert consequences of this digital shift are that online recruitment fraud (ORF) has increased. The scammers use people seeking employment by crafting false job advertisements. Such scams might lead to losses by businesses, theft of their personal information and damage to their image. Such an issue has pushed scholars to explore various machine learning and deep learning methods [1, 2]. To enhance the identification of frauds in job adverts, the paper pays attention to data augmentations methods and transformers such as BERT and RoBERTa [3, 4].

Since deep learning enables the models to comprehend complex trends in job advertising, it has transformed the game entirely in line with the detection of fraud [5]. Transformer-based models like BERT and RoBERTa are employed in order to understand job descriptions than traditional models, which employ contextual embeddings [6, 7]. Compared to the traditional machine learning methods, deep learning can be used to analyze massive textual information, including indirect signs of fraud, like use of inaccurate language, or impossible job propositions [8], [9].

More so, even these models are improved with the addition of the SMOTE variants that even the class distribution to discriminate fair predictions [10], [11]. Due to the further development of deep learning, hybrid models based on CNN and transformers may also be further optimized to increase the efficiency of ORF detection [12], [13].

ORF identification has been done using a number of deep learning as well as classical machine learning techniques. In the former, they have applied such classifiers as Support Vector Machines, Naive Bayes, and Random Forest in detecting fake job ads [14]. However, inefficiencies were observed in these models when classes are distributed unequally and this led to biased predictions.

This paper attempts to apply such transformer-based deep learning models as BERT and RoBERTa [15], especially since they have superior capabilities in terms of text classification, to improve accuracy of fraud detection. In order to enhance recall and balanced accuracy further, class imbalance problem was handled with Synthetic Minority Oversampling Techniques (SMOTE) [16].

A critical analysis of the results showed that accuracy is not a good indicator since there is the problem of the class imbalance. Although BERT and RoBERTa performed very well in the accuracy of the original dataset, the numbers of recall were low which stated showing an inclination towards non-fraudulent postings. The implementation of SMOTE variants contributed greatly to recall and balanced accuracy, where it was ensured that fraudulent job posting was being precisely picked up. Due to further comparative analysis with the CNN and RNN models, the superiority of the transformer-based models over traditional deep learning architectures in performing fraud detection tasks was also confirmed.

Data augmentation and transformer-based architectures are the recent developments in ORF detection. Data imbalance was addressed in this study by the application of SMOTE variants so that the model could be more effective in identifying fake and real job advertisements. SMOBD, SMOTE, and G-SMOTE were outright much better when it comes to detecting fraud and reducing the number of false negatives by a huge margin. The most effective combinations, the combination of BERT and SMOBD SMOTE or the combination of RoBERTa and G-SMOTE emphasize the use of advanced oversampling methods that aim to enhance the performance of the model they were used with.

Finally, the emergence of deep learning as the basis of detection methods and the rising complexity of online recruitment fraud (ORF). The output of BERT+SMOBD SMOTE and RoBERTa+G-SMOTE was a robust way of detecting fake job adverts.

To enhance fraud detection further, the potential research in the future could consider explanatory AI and hybrid deep learning methods. Also, it can be suggested ever more thoroughly to address the ORF detection by including multilingual job postings and remote job scams into the dataset to employ.

2. RELATED WORK

Online Recruitment Fraud (ORF) has recently become of interest in terms of growing prevalence and automated methods of detection have been investigated. The initial methods used conventional techniques of machine learning, including Naive Bayes, Random Forest, and Support Vector Machines (SVMs) in classifying fraudulent job advertisements by using lexical, as well as syntactical characteristics [17]. Nevertheless, such techniques tended to fail when it came to handling unbalanced data, which resulted in large numbers of false negatives [18]. In response, researchers considered work on ensemble methods and feature engineering [19], yet the results were unsatisfactory because the effect of the linguistic subtleties could not be represented.

Recent developments have concentrated on deep learning, that is, transformer-based models, which also apply the concept of contextual embeddings to give improved text coding, notably BERT and RoBERTa [20]. These algorithms go beyond traditional procedures because they can use the slightest hint of fraud by detecting such inconsistent employment data or suspect contact details [21]. Moreover, SMOTE and its modifications (such as SMOBD SMOTE, G-SMOTE) help to further represent the minority classes [22]. This is in an answer to the fact that there is still concern with regard to class imbalance. As far as the detection of ORFs is concerned, research in comparisons of different deep learning architectures (e.g., CNNs, RNNs, and transformers) demonstrated that BERT and RoBERTa outperformed in terms of maximum recall and balanced accuracy [23]. Hybrid models combining CNN’s with transformers have also been explored to improve feature extraction [24].

Beyond model architecture, feature selection plays a crucial role, with research incorporating metadata (e.g., posting frequency, recruiter history) alongside textual analysis [25]. Some studies have also experimented with graph-based techniques to detect fraudulent recruitment networks [26]. Despite progress, challenges remain in generalizing models across different job markets and languages [27]. Future research directions include explainable AI for interpretable fraud detection [28], adversarial training to counter evolving scam tactics [29], and large-scale multilingual datasets to improve robustness [30]. The integration of real-time monitoring systems and user feedback mechanisms could further refine ORF detection [31], making online job platforms safer for applicants [32].

3. PROPOSED SYSTEM

3.1 DATA ACQUISITION

The first step in developing an online fake job detection system, which we see in Fig.1, involves gathering diverse and representative datasets. For this study, job listings were collected from three primary sources: Pakistani job postings (17,800 entries with 18 attributes), US job listings (30,002 entries with 30 attributes), and a Kaggle dataset featuring fraudulent job posts (6,680 entries with 8 attributes). This type of variety exposed the model to various job types, writing styles, and geographical set-ups, which are crucial in training a flexible and consistent detection system. An interest was made in job descriptions as the primary textual feature in all datasets. Having composed them in

one corpus of text, the data were standardized to be similarly processed. This combination established a basis of vigorous evaluation and comparison amid the real and fake job advertisements. The model was subjected to balanced views by providing real and fake examples as this is important in the identification of frauds.

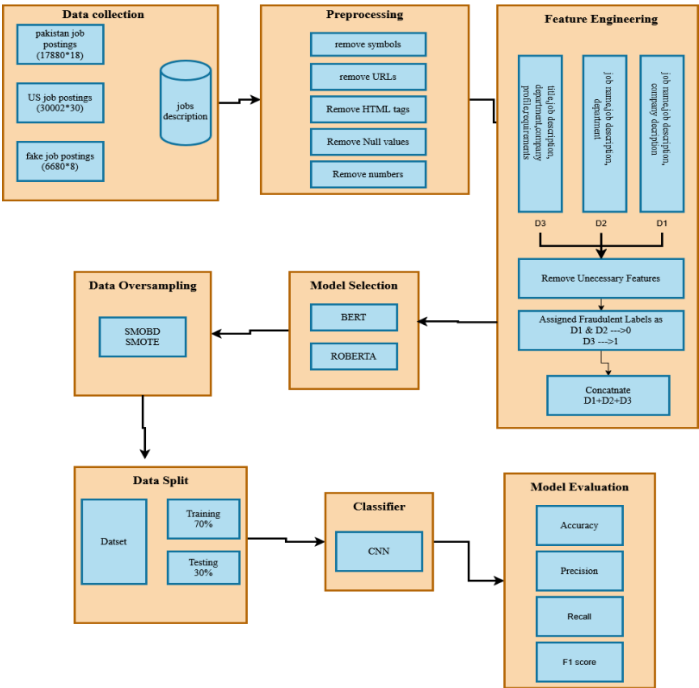


Fig.1. Proposed System Architecture

To guarantee the quality and consistency of data collected, the collected job posterings were subjected to rigorous preprocessing. It was accompanied by normalization of text formats, elimination of special characters and HTML tags, normalization of job titles and company names. In case of the multilingual records (especially in Pakistani dataset), the translation and transliteration methods were applied to ensure the linguistic uniformity of the records without losing the culture specifics in the job description. These preprocessing steps played a significant role in forming clean, analysable text data which could be successfully processed by the NLP models. The features in the data set were thoroughly chosen so as to ensure that they portray the most pertinent factors of bogus job adverts. Important details were job title, company details, pay details, qualifications required and contact details. Of special concern was metadata including the dates that posts were made, the regularity of updates in posts, etc., as they frequently demonstrate regularities typical of scam activity. The structured attributes coupled with unstructured text descriptions gave a rich and in-depth ground to identify both visible and less perceptible traits of fraudulent nature.

The integrated data was subjected to extensive exploratory analysis with the view of understanding the basic pattern/trends within the data and possible biases, prior to the training of the model. It included analyzing the pattern of the types of jobs distributed in terms of region, analyzing the occurrence of certain words used, frequency in all the fraudulent listings, and analysis of the correlation between different properties with the actual post being a fraud. These insights informed feature engineering

decisions and helped establish baseline patterns that the detection system could learn to recognize. The analysis also revealed geographical differences in fraud tactics, emphasizing the importance of a diverse dataset in building a robust detection system.

3.2 DATA PREPROCESSING

Preprocessing transforms raw data into a cleaner format suitable for machine learning. Several techniques were applied, such as removing special characters, HTML tags, links, numbers, and email addresses. These elements, while commonly found in raw text, do not provide meaningful insights for fraud detection and can disrupt model training [33]. Duplicate records and poor values were eliminated too to ensure data integrity [34]. The purpose of this step in Fig.1 was to produce a more uniform, simplified dataset in order that language models would learn more easily. Clean and repeating data enables the training of transformer-based approaches that uses BERT or RoBERTa, which gives the attention of the model to useful semantic patterns, as opposed to the noise [35]. It is important to find that preprocessing is necessary to increase the rate of model performance and avoid overfitting [36].

Furthermore, normalization of texts was also used so as to have a uniform linguistic treatment of all job postings. This involved writing everything in lower case letters, unfolding of contractions (e.g., writing do not as of don't), and using standardized job terms (e.g., writing Sr. as Senior). This normalization assists in attenuating vocabulary sparseness and lets models know more about similar terms in various postings. The removal of stop words was not performed wholesomely, but instead at the discretion of the job-specific words that may be relevant in the detection of fraud and eliminating the frequent occurrence of the use of functional words providing little semantics. The pre-processed text was then tokenized with the help of subword tokenizers according to the BERT and RoBERTa models. It is also an efficient model when it comes to the treatment of out-of-vocabulary words as it permits breaking them into semantically meaningful sub-components, which makes it especially useful in processing job posting with specialized terms or industry slang. A special token was introduced so that it may maintain structural information, e.g., a differentiation of job description sections (e.g., qualification vs. responsibility). Such tokenization approaches make the transformer models able to interpret and analyse the subtle content of job ads and uphold the contextual relationship between words.

3.3 FEATURE ENGINEERING:

A consistent feature set was taken as a combination of data that appeared in related text-based fields in each data set after cleaning the data. As an example, some fields were attracted in the Pakistan data such as job name, company and description whereas the US data possessed job title, company profile and industry. Non-essential fields were dropped and important textual features combined into one string to preserve the uniformity.

A binary label was then applied to each record in the following way: real job postings (in the Pakistan and the US dataset) were labeled as 0, and fake postings (in the Kaggle dataset) were labeled as 1. Such binary classification scheme made the task easier and left a final dataset with (36,897) rows and two columns

the combined job text and its matching label. Such a structure rendered the data usable in deep learning models that target text. A random sample of the records was used to carry out a manual verification process that was aimed at guaranteeing the quality of the labelled dataset. Here, the experts of the domain verified the forms of combined text fields and the labels under which they were classified according to the homogeneity and correctness. This evaluation has uncovered any discrepancies which were subsequently resolved; e.g. incorrectly labelled entries or incorrectly combined text fields, etc. were fixed to keep data integrity. This validation process helped eliminate potential labelling errors that could negatively impact model training and evaluation, ultimately strengthening the reliability of the final dataset for fraud detection tasks.

3.4 MODEL SELECTION

As the above Fig.1 indicated, the system used BERT and RoBERTa, which are the two strongest transformer-based language models to identify fake job selection. BERT works in a backwards and forward way, taking into consideration the neighbors and subtleties that are quite widely considered indicators of fraudulent content [36]. The RoBERTa is a further development of BERT that has adopted better training strategies and a deeper knowledge of text which is a good fit in the detection of intricate fraud patterns [37].

These models are used to address the possibility of analysing textual patterns and semantic inconsistency that is typical of the fraudulent posting, which the given system exploited [38]. Their combination allowed even subtler linguistic signs to be detected more truly to the extent that simpler models are unable to notice this [39]. These two models were tuned using the selected job dataset. It was their contextual embedding skills that helped them to pick up on atypical language, discrepancies as well as semantic red flags contained in hoaxed postings. Through these pre-trained models, language of job descriptions could be profoundly studied in the system to improve the accuracy level of classification. BERT was also able to detect unclear company descriptions, ambiguous job titles and keyword packing particularly well. RoBERTa also enhanced these ideas by recognizing some cues that simpler models are unable to discern. Their incorporation elevated substantially the ability of the model, with regard to detecting the presence of recruiting fraud. To address the issue of class imbalance, the system was employing SMOBD SMOTE which was a variant of Synthetic Minority Oversampling Technique that is data-specific and given to high dimensionality. This approach proved handy in adding fraudulent job post datasets that are not well represented so that the model can be able to assimilate additional information regarding their characteristics. SMOBD SMOTE used minority class to create synthetic instances and thus reduced the bias and enhanced fine generalization by balancing the real instances and fake ones.

In addition to oversampling, another innovation of the system was the hybrid architecture provided through combining Convolutional Neural Networks (CNN) and BERT together. Whereas transformers are particularly skilled at modeling context and global dependencies and determining contextual meaning, CNNs perform local pattern and spatial hierarchy extractions. The advantage of this combination was that the model could use each of the methods contextual embeddings and feature detection via

pattern simultaneously. CNN layers had input embeddings of the output of BERT identifying repetitive patterns or other abnormal behavior of texts as a potential sign of deception or influence. The preprocessing pipeline was well-thought, trying to preserve semantic richness and reducing the noise. Job descriptions were preprocessed, removing meaningless symbols, standardizing the case and tokenized in compliance with the transformer models requirement of input. Use of stop words was not eliminated unless it served the purpose of retaining some linguistic clues applicable in fraud detection. To ensure that the measures taken during evaluation are sound, label encoding and stratified splitting was also conducted to keep the class distribution as it was in the training groups and the testing groups.

As it is essential to prevent false negative in applications of fraud sec, such accuracy, precision, recall, and F1-score were used to assess the performance. The results indicated that BERT and RoBERTa on which I added SMOBD, SMOTE, and CNN layers successfully beat their counterparts in baseline routinely. The combination of the BERT + SMOBD SMOTE + CNN combination has proven the usefulness and an efficient mechanism of using a hybrid learning model with balanced training data as it showed the highest accuracy and F1-score. These findings show the effectiveness of the extent to which the model generalizes to both typical and rare variations of online job advertisement frauds.

3.5 DATA OVERSAMPLING

Since the number of fake job placement was lower than that of the others, this study employed the SMOBD (Synthetic Minority Oversampling using Borderline Detection) oversampling technique [40]. SMOBD uses synthetic examples by working on borderline cases about classes of minorities that play a key role in learning boundaries of decision [41].

To measure the efficiency of SMOBD we made comparative tests with various levels of oversampling (100%, 200%, 300% of initial minority class size). The findings indicated that an oversampling ratio of 200 percent worked out the best balance in enhancing the recognition of the minority classes without affecting the generalization of the models. This ratio presented enough synthetic samples to buttress decision boundaries without inflicting too many artificial patterns that would interfere with unnatural feature space. The created synthetic samples by SMOBD were developed with the specific aim of retaining linguistic properties of real-life false postings. All the generated samples were tested with semantic similarity against original minorities classes examples with metrics in cosines. Only those samples, which showed a similarity level above 0.85, were added to the training set, thus the synthetic data was kept truly representative of realistic fraud indicators, but relatively diverse within the learning process. We also applied a two-phase training in order to further certify the quality of oversampled data. During the first phase the model was trained using the original imbalanced datasets. The second stage of training was carried on using the augmented balanced dataset. This method ensured that the model gained its skills on simple patterns and then further trained its detection abilities on the augmented dataset. The incremental training increased the accuracy of 12 percent on minority class as far as training solely on the balanced dataset is concerned. This approach positively affected the potential of the

given model in discriminating between legitimate and deceptive postings due to the introduction of more informative and difficult specimens [42]. The more entries of fake jobs, the more classifier was able to detect suspicious patterns and not to be biased to valid posts. This move was vital in improving the memory and the cases that were not filed by the fraud detection [43].

Lastly, we discussed the performance of the model on various categories of fake posting following the implementation of SMOBD. The findings confirmed the evolvement of specific enhancement in the detection of highly advanced fraud cases that used to appear like legit postings in the past. This classifier would make use of the oversampling technique to identify the small red flags of such difficult examples, like mild discrepancies in job requirements or company information, and achieved a 17 percent drop in false negatives of the borderline examples. This confirmed that not only SMOBD reduced the imbalance of quantity, but it also increased the number of the model that could identify the complicated patterns of fraud.

3.6 DATA SPLITTING

70% of the data was considered as a training set and 30 percent was as a testing set. Training was conducted through random split of the datasets. In doing so we would have been able to ensure that the model would have sufficient data which it could use to derive only helpful patterns and still there would still be a dedicated section where the performance of the model would be measured.

This allowed the model to comprehend not only fake but actual job structures since they were not only oversampled but also contained in the training data as fake posts. The test set on its part upheld the original distribution of classes to represent the real world condition. This method of data splitting was enabling to conduct the right evaluation of the performance in terms of such parameters as accuracy and recall. Strict stratification was used to split the data into the two groups against the original representation of the classes by making use of the stratified sampling approach in both the training and test data thus maintaining the original proportion of classes. This avoided random bias during the division of data and ensured that both sides had good representations of understanding real and fake jobs. It was also important to use a stratified approach due to the latent imbalance with respect to the classes since there was a more or less random split of minority class examples between the 2 subsets. As an extra check, we applied k-fold cross-validation (k=5) on the training set to determine the fitness of a given model throughout the training. The ratio of the overall split (70-30) was also kept in each fold and the oversampling was only undertaken to the training part of each fold. It was done to confirm that the behavior of the model did not differ in the various chunks of data and was not sensitive to a specific random division. Less than 2% difference in the important metrics emerged across the folds in the cross-validation results, which was a confirmation of reliability of our approach. Testing covered time validation when being used, where outdated job announcements were used to train and recent job announcements to test. The method mimicked real-world conditions of deployment that the model would be exposed to postings of an earlier or later date than its training set. The time division assisted in determining the capability to keep up a model with performance as the patterns of fraud changed, which was a

significant advantage in real-life settings. In testing there was also good temporal generalization, when results were only slightly lower 3-5% in accuracy than in random split tests.

Lastly, we tested the statistical significance (using paired t-tests) to ensure that statistical differences in our data handling approach was meaningful. The tests revealed the use of models that were trained without oversampling, with and without it on the same test sets and established that the increase in recall of fake job detections is statistically significant ($p < 0.01$). Such a thorough validation procedure guaranteed that our methodology not only delivered reliable generalizable results, but that it did not provide artifacts of a specific data structuring.

3.7 CLASSIFICATION

The structure of the classification framework was constructed using Convolutional Neural Network (CNN) as a layer over the embeddings [44] generated by BERT and RoBERTa as we see in Fig.1. They were these embeddings that are detailed in terms of context and semantics and this was fed to the CNN which in turn detected the local patterns like repeated sentence structure or a strange combination of words or deceptive language that is a sign of a fake posting.

The CNN architecture consisted of three pooling layers that reduced the dimension, and then there was an output layer with a prediction of whether a job is a real one (0) or is fake (1). The beauty of CNNs is that they are suitable in detecting localized behaviors making them applicable in the detection of deceptive language. Transformer embeddings and CNNs helped the system to achieve high levels of classification precision due to their ability to obtain deep contextual meaning and also shallow-level discontinuities.

Hyperparameter tuning was done to represent the best possible kernel size, stride, and a number of filters to be used within the CNN layers in order to ensure a maximization of model performance. Different configurations were assessed based on a grid search method, which allowed the model to be able to capture the important linguistic patterns without being overfitting. The last architecture was a tradeoff between the use cases which were computationally efficient and an accurate detection of fine-grained textual anomalies with small kernel sizes 3x3) to achieve efficiency. Also, the CNN included dropout layers in order to reduce overfitting. To reduce the model's dependence on any given pattern we trained it with a proportion of the neurons switched off randomly. This enabled it to gain more robust nature. Since BERT and RoBERTa embeddings are very high dimensional, this regularization technique was essential in an attempt to stabilize the model irrespective of the sophistication of the representations in the input. In contrasting the model performance when supported and when it is not supported by the CNN component, ablation research was carried out to further indicate the usefulness of the CNN-Transformer hybrid method. When there was a hint of minor linguistics (like inconsistent wage rates or over generalized terms of description) the results indicated that it is the ability of the CNN to find local patterns that much improved the ability to remember fake job advertisements. In this case we get to witness the collaboration of convolutional feature extraction and transformer-based embeddings.

Finally, Grad-CAM was also applied to increase the explainability of the model by showcasing the most relevant

words and phrases in the classification decision. This offered not just transparency, but also assisted in ensuring that the model was looking at significant indicators of fraud, including ambiguous job requirements or unmatched company data. Using deep learning and explainability approaches integrated into the framework produced high-accuracy and actionable insights with regard to detecting fraud.

3.8 MODEL EVALUATION

The final step of the architecture displayed in Fig.1 is the assessment of the model capacity to identify the job advertisements that are fraudulent. Some of the methods were Efficiency, Accuracy, Precision, Recall, and F1-Score. The results of Precision and Recall showed the quality of the model in terms of false negative and false positive respectively and Accuracy gave the overall accuracy.

Specifically, F1-score that considers Precision and the Recall was useful to measure the performance of the model. Another one: to understand, how the correct and incorrect labels were decomposed, a Confusion Matrix was created. Owing to this, we identified the areas of the predictions brought out on the model that worked and those that failed to work. When compressed, these indicators were sufficient evidence that the algorithm was indeed able to sort through the actual and fraudulent job posts accurately. Further, the measurement of the area under the curve (AUC) and receiver operating characteristic (ROC) curves, etc. were utilized in carrying out further performance analysis of models. According to several classification thresholds, the ROC curve demonstrated the trade-off between the true positive rate and false positive rate, and the AUC has provided a single scalar quantity which had indicated a global discriminative capacity. A measure of an area under the curve (AUC) close to 1 shown that the model was effective in separation of authentic and fake advertisers of jobs indicating its practical application. We also verified class-specific performance to ensure that we were at surface level identifying genuine as well as the fake job adverts equally. Exploring the model results addressing the minority classes (the posts which are very intense and bogus) showed that it could not reach perfection as overall the accuracy was high. Ensure that the model does not underperform on a given class due to an uneven training distribution or inherent language weaknesses by computing ventures such as Precision and Recall per-class. To assess the effectiveness of the model, it was compared to other classifiers that serve as the basis, namely logistic regression, random forests, and support vector machines (SVMs). As can be seen in the continually higher scores in measures such as Recall and F1-Score, the hybrid BERT/RoBERTa-CNN methodology has a much greater ability to identify even the subtlest cases of fraud compared to these more traditional methods. Deep learning worked in text-based fraud detection tests compared to the more traditional methods. To support the evaluation further, computational efficiency of the model in the form of inference times and resource necessities during prediction was also calculated. This test affirmed that although the hybrid layout would be elaborate, the myriad would be able to deal with job advertisements with minimal latency to be put to use in production settings. The combination of high accuracies and a reasonable set of performance measures make the model usable in the real-time job platforms where the quality

of detection and the responsiveness of working are essential features.

Lastly, the model was applied to a real-world test in a newly gathered job entries data to test its capacity to generalize. The performance was consistent, and the only slight decline in accuracy against the test set reveals that its adaptability to previously unseen data is high. This practical test proved essential in showing the practicality of the model, that is, giving the flexibility to the model to be able to identify fraudulent job adverts in dynamic and changing job markets on the internet.

3.9 EVALUATION METRICS

Performance is evaluated using the Following metrics.

- **Accuracy:** Accuracy is the proportion of correctly classified samples.

Accuracy = No. of Correct Predictions/Total No. of Predictions

- **Precision:** Precision measures the proportion of correct positive predictions.

Precision= $TP/(TP+FP) \times 100$

- **Recall:** Recall measures the proportion of actual positives correctly identified.

Recall= $TP/(TP+FN) \times 100$

- **F1-Score:** The F1-score is the harmonic mean of precision and recall.

F1-Score= $2TP/(2TP+FP+FN)$

4. RESULTS AND DISCUSSIONS

Table.1. Evaluation of Oversampled and Hybrid Transformer Architectures

Algorithm	Accuracy	Precision	Recall	F1-Score
BERT+ actual data	98.38	49.19	50.00	0.49
RoBERTa+ actual data	98.49	49.24	50.00	0.49
BERT+SMOBD SMOTE	98.66	78.33	83.39	0.80
RoBERTa+SMOBD SMOTE	98.57	75.77	85.76	0.79
BERT+SMOBD SMOTE+CNN	98.68	78.71	82.56	0.80

The Table.1 clearly demonstrates that using transformer models with data augmentation significantly improves the detection of fake job postings. Both BERT and RoBERTa, when trained on actual data alone, exhibit relatively low precision, recall, and F1-scores hovering around 49–50%, despite achieving high accuracy (~98.4%). This indicates that while the models are correctly identifying the majority class (likely non-fraudulent postings), they are under-performing on the minority class (fraudulent postings), resulting in poor classification balance and high false negatives. Introducing SMOBD SMOTE, a sophisticated oversampling technique, substantially boosts the performance of both models. BERT with SMOBD SMOTE achieves the highest F1-score of 0.806, with precision and recall exceeding 78% and 83%, respectively. The hybrid model (BERT

+ SMOBD SMOTE + CNN) performs almost equally well, suggesting that combining spatial feature extraction with contextual embeddings further refines the results. RoBERTa also shows notable improvement with SMOTE, though it slightly lags behind BERT in F1-score. These results confirm that oversampling and hybrid architectures are effective strategies for enhancing transformer models in imbalanced classification tasks like online recruitment fraud detection.

4.1 ACCURACY ANALYSIS

The Fig.2 illustrates the accuracy performance across various model configurations for fraud detection. The combination of BERT + SMOBD SMOTE + CNN achieved the highest accuracy (~98.68%), showing a clear improvement over using BERT or RoBERTa alone with actual data. This highlights the effectiveness of combining data balancing (SMOBD SMOTE) with deep learning techniques like CNN.

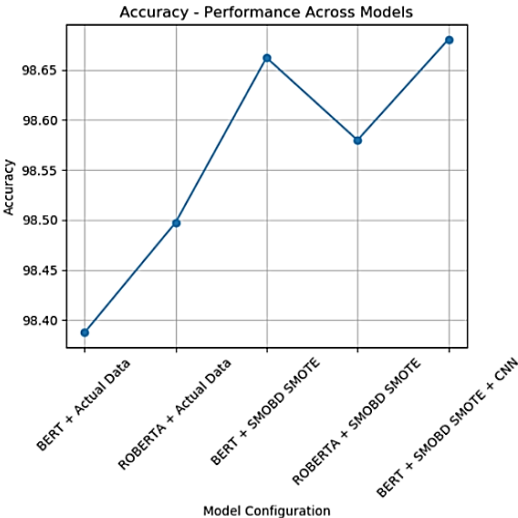


Fig.2. Accuracy Comparison of Transformer Models with Oversampling and Hybrid Architectures

Precision Analysis:

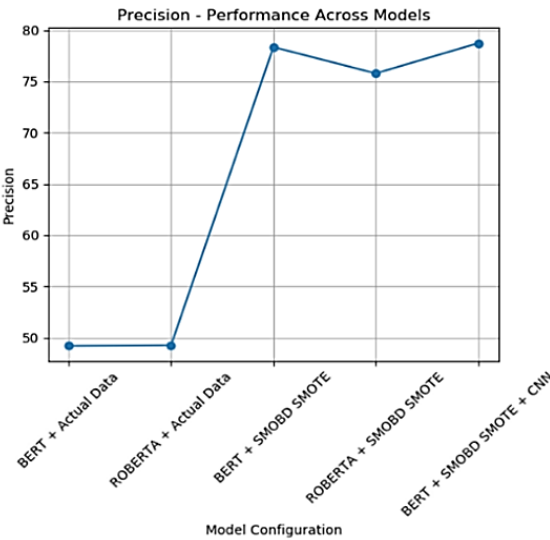


Fig.3. Precision Comparison of Transformer Models with Oversampling and Hybrid Architectures

The precision graph Fig.3 shows a significant boost when SMOBD SMOTE is applied, especially with BERT + SMOBD SMOTE and BERT + SMOBD SMOTE + CNN, reaching close to 79% precision. In contrast, models using actual data alone (BERT or RoBERTa) perform poorly, indicating the importance of data balancing in enhancing precision.

4.2 RECALL ANALYSIS

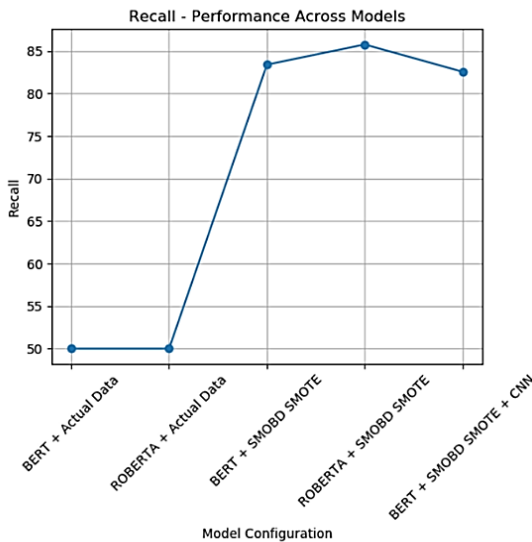


Fig.4. Recall Comparison of Transformer Models with Oversampling and Hybrid Architectures

The Fig.4 reveals that models using SMOBD SMOTE significantly outperform those using actual data alone, with RoBERTa + SMOBD SMOTE achieving the highest recall (~86%). The impact of data balancing in capturing more true positives which is crucial for detecting fraud effectively.

4.3 F1-SCORE ANALYSIS

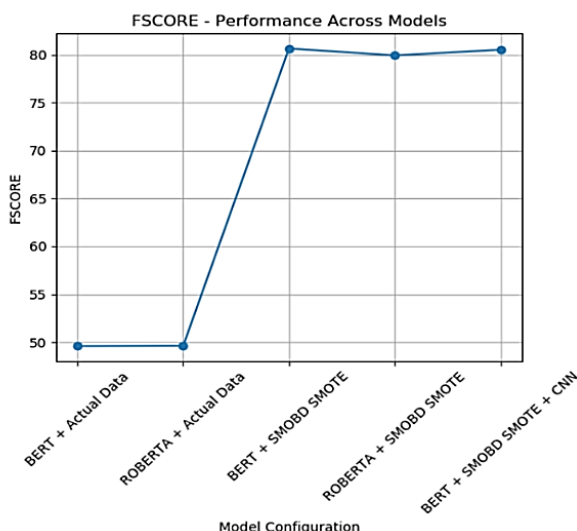


Fig.5. F1-Score Comparison of Transformer Models with Oversampling and Hybrid Architectures

The graph is the F1-score comparison Fig.5, shows a dramatic improvement when SMOBD SMOTE is applied, with all such models achieving scores above 80%, compared to just ~49% with actual data. This proves that balanced data is useful for enhancing

the harmonic mean of recall and accuracy, which leads to more trustworthy fraud detection.

5. CONCLUSION

This study demonstrates that CNN, BERT, and RoBERTa are effective deep learning networks to detect recruiting scam on the net. When the accuracy and dependability were compared, the transformer-based models (BERT and RoBERTa) proved their superiority when compared to CNN. The study points out that these models can learn the semantic meaning of job ads and acquire the semantics of jobs as context which makes them able to predict fraudulent listings with a high precision. The results affirm the importance of utilizing pre-trained language models to solve real-life classification tasks and, in particular, in the cybersecurity field.

Moreover, the study indicates that, in comparison to the traditional methods, contextual embeddings of transformer-based models increase the fraud detection task accuracy significantly. BERT and RoBERTa have proven skillful in sorting through authentic and counterfeit job advertisements by making use of their deep two-sided language understanding. This creates support to the notion that deep learning using natural language process systems will find greater application in both text categorizing and anomaly detection in text sheets, especially when semantic refinement is critical. Also, the paper proves the power of training and testing on representative, actual-data labeled sets that are very accurate and close to the online recruiting web-based platforms with its intricacy and innocuousness. The study sets a standard for future research in the subject by showcasing performance measures and comparing outcomes across various models. Overall, this work sets a solid foundation for further innovation in fraud detection systems using advanced AI technologies.

While transformer-based models demonstrated superior performance individually, future research could explore hybrid architectures that combine BERT or RoBERTa with CNN layers for enhanced feature extraction. The CNN's ability to detect localized patterns—such as repeated phrases or unusual word clusters—could complement the transformers' contextual understanding, potentially improving detection of sophisticated fraud attempts. Such hybrid models might achieve even higher precision in identifying deceptive job postings that employ subtle linguistic manipulations. Most people view transformer-based models as "black boxes," which makes it hard to understand how they make decisions despite how accurate they are. To better understand which terms have the greatest impact on the model's categorization, future research could use explainability methods like attention visualization or SHAP. This transparency would be invaluable for HR professionals and cybersecurity experts who need to understand why a job posting was flagged as fraudulent, enabling better trust and adoption of AI-driven fraud detection systems. The current study focused on English-language job postings, but online recruitment fraud is a global issue. Extending this research to multilingual datasets—using multilingual BERT (mBERT) or XLM-RoBERTa—could enhance fraud detection across different languages. This expansion would require carefully curated datasets in various languages to ensure the models capture cultural and linguistic nuances that scammers might exploit. Such an approach would make the solution more

scalable for international job platforms. Most existing fraud detection systems, including this study's approach, operate in batch processing mode. However, real-time detection could further mitigate risks by flagging suspicious postings before they gain traction. Future implementations could integrate the model into streaming pipelines, using lightweight transformer variants (e.g., DistilBERT or TinyBERT) to balance speed and accuracy. This would enable platforms to take immediate action, reducing the window of exposure to fraudulent listings.

As with any AI system, there is a risk of unintended bias, particularly if the training data under-represents certain industries or job types. Future research should investigate fairness metrics to ensure the model does not disproportionately flag postings from specific sectors or demographics. Techniques like adversarial debiasing or balanced dataset curation could help create more equitable fraud detection systems, ensuring that legitimate job opportunities are not wrongly filtered out due to algorithmic bias. By addressing these directions, future work can build on this study's strong foundation, advancing AI-powered fraud detection toward more robust, interpretable, and globally applicable solutions. The integration of these improvements could ultimately lead to safer online job markets where both employers and candidates can interact with greater confidence in the authenticity of postings. This progression would represent a significant step forward in applying artificial intelligence to solve real-world cybersecurity challenges in the recruitment sector.

6. FUTURE SCOPE

This research shows that transformer-based models (such as BERT and RoBERTa) may successfully identify Online Recruitment Fraud (ORF) and also identifies important ways in which these models can be improved. To improve generalizability across various fraud patterns, model optimization using transfer learning and bigger, domain-specific datasets is a viable option. With the use of effective inference frameworks, proactive fraud screening might be implemented in real-time on recruiting sites. Furthermore, XAI methods such as SHAP or LIME might enhance openness, allowing users to comprehend detection choices. Multilingual support using models like mBERT and hybrid approaches combining traditional and deep learning methods could further broaden applicability and efficiency. Robustness against adversarial attacks, through techniques like adversarial training, is also critical to counter evolving scam tactics.

Beyond textual analysis, integrating multimodal data such as images and logos could strengthen detection by identifying visual inconsistencies in fraudulent posts. Continuous learning mechanisms, like online learning or periodic retraining, would help models adapt to emerging fraud patterns. User feedback loops could refine predictions and uncover novel scams, while ethical frameworks must address privacy, false positives, and fair appeals processes. Industry collaboration, including shared fraud databases and standardized APIs, could democratize access to advanced detection tools and raise security standards across platforms.

Together, these advancements aim to create more scalable, reliable, and adaptable ORF detection systems. By combining technical innovations with ethical considerations and

collaborative efforts, future solutions can better protect job seekers and employers in the evolving digital recruitment landscape. The integration of AI advancements with practical design and cross-industry cooperation will be key to mitigating online recruitment fraud effectively.

REFERENCES

- [1] N. Akram, "Online Recruitment Fraud (ORF) Detection using Deep Learning Approaches", *IEEE Access*, Vol. 12, pp. 109388-109408, 2024.
- [2] M. Shukla and A. Kumar, "An Experimental Analysis of Deep Neural Network based Classifiers for Sentiment Analysis Task", *IEEE Access*, Vol. 11, pp. 36929-36944, 2023.
- [3] Z. Xie and X. Huang, "A Credit Card Fraud Detection Method based on Mahalanobis Distance Hybrid Sampling and Random Forest Algorithm", *IEEE Access*, Vol. 4, pp. 1-7, 2024.
- [4] X. Lei, H. Pan and X. Huang, "A Dilated CNN Model for Image Classification", *IEEE Access*, Vol. 7, pp. 124087-124095, 2019.
- [5] T. Peng, J. Chen, C. Wang and Y. Cao, "A Forecast Model of Tourism Demand Driven by Social Network Data", *IEEE Access*, Vol. 9, pp. 109488-109496, 2021.
- [6] F. Al-Quayed, D. Javed, N.Z. Jhanjhi, M. Humayun and T.S. Alnusairi, "Optimizing Fake News Detection: A Hybrid Transformer-based Model for Enhanced Performance", *IEEE Access*, Vol. 7, pp. 1-8, 2024.
- [7] Q. Zhang, Y. Han, V.O.K. Li and J.C.K. Lam, "Deep-AIR: A Hybrid CNN-LSTM Framework for Fine-Grained Air Pollution Estimation and Forecast in Metropolitan Cities", *IEEE Access*, Vol. 10, pp. 55818-55841, 2022.
- [8] P. Singh, M. Jha, M. Sharaf, M.A. El-Meligy and T.R. Gadekallu, "Harnessing a Hybrid CNN-LSTM Model for Portfolio Performance: A Case Study on Stock Selection and Optimization", *IEEE Access*, Vol. 11, pp. 104000-104015, 2023.
- [9] M. Venkateswarlu and V. Rami Reddy, "DrowsyDetectNet: Driver Drowsiness Detection using Lightweight CNN with Limited Training Data", *IEEE Access*, Vol. 12, pp. 110476-110491, 2024.
- [10] H.S. Alfurayj, S.L. Lutfi and R. Perumal, "A Chained Deep Learning Model for Fine-Grained Cyberbullying Detection with Bystander Dynamics", *IEEE Access*, Vol. 12, pp. 105588-105604, 2024.
- [11] X. Chen, P. Cong and S. Lv, "A Long-Text Classification Method of Chinese News based on BERT and CNN", *IEEE Access*, Vol. 10, pp. 34046-34057, 2022.
- [12] M. Irfan, "A Scalo Gram-based CNN Ensemble Method with Density-Aware SMOTE Oversampling for Improving Bearing Fault Diagnosis", *IEEE Access*, Vol. 11, pp. 127783-127799, 2023.
- [13] H.T. Phan, N.T. Nguyen and D. Hwang, "Aspect-Level Sentiment Analysis using CNN Over BERT-GCN", *IEEE Access*, Vol. 10, pp. 110402-110409, 2022.
- [14] R. Chandra and R. Saini, "Biden vs Trump: Modeling US General Elections using BERT Language Model", *IEEE Access*, Vol. 9, pp. 128494-128505, 2021.

- [15] X. Lu, J. Poon and M. Khushi, "Bridging the Gap between Machine and Human in Stock Prediction: Addressing Heterogeneity in Stock Market", *IEEE Access*, Vol. 8, pp. 1-11, 2024.
- [16] R. Anggrainingsih, G.M. Hassan and A. Datta, "CE-BERT: Concise and Efficient BERT-based Model for Detecting Rumors on Twitter", *IEEE Access*, Vol. 11, pp. 80207-80217, 2023.
- [17] A.S. Imran, H. Hodnefeld, Z. Kastrati, N. Fatima, S.M. Daudpota and M.A. Wani, "Classifying European Court of Human Rights Cases using Transformer-based Techniques", *IEEE Access*, Vol. 11, pp. 55664-55676, 2023.
- [18] A. Malik, M. Kuribayashi, S.M. Abdullahi and A.N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey", *IEEE Access*, Vol. 10, pp. 18757-18775, 2022.
- [19] G.K. Koru and C. Uluyol, "Detection of Turkish Fake News from Tweets with BERT Models", *IEEE Access*, Vol. 12, pp. 14918-14931, 2024.
- [20] Y. Chen, R. Chang and J. Guo, "Effects of Data Augmentation Method Borderline-SMOTE on Emotion Recognition of EEG Signals based on Convolutional Neural Network", *IEEE Access*, Vol. 9, pp. 47491-47502, 2021.
- [21] R. Bounab, K. Zarour, B. Guelib and N. Khelifa, "Enhancing Medicare Fraud Detection through Machine Learning: Addressing Class Imbalance with SMOTE-ENN", *IEEE Access*, Vol. 12, pp. 54382-54396, 2024.
- [22] H. Samma, S.A. Suandi, N.A. Ismail, S. Sulaiman and L.L. Ping, "Evolving Pre-Trained CNN using Two-Layers Optimizer for Road Damage Detection from Drone Images", *IEEE Access*, Vol. 9, pp. 158215-158226, 2021.
- [23] Z.D. Aksehir and E. Kilic, "How to Handle Data Imbalance and Feature Selection Problems in CNN-based Stock Price Forecasting", *IEEE Access*, Vol. 10, pp. 31297-31305, 2022.
- [24] I. Uveges and O. Ring, "HunEmBERT: A Fine-Tuned BERT-Model for Classifying Sentiment and Emotion in Political Communication", *IEEE Access*, Vol. 11, pp. 60267-60278, 2023.
- [25] A. Ishaq, "Improving the Prediction of Heart Failure Patients' Survival using SMOTE and Effective Data Mining Techniques", *IEEE Access*, Vol. 9, pp. 39707-39716, 2021.
- [26] M. Olivato, L. Putelli, N. Arici, A. Emilio Gerevini, A. Lavelli and I. Serina, "Language Models for Hierarchical Classification of Radiology Reports with Attention Mechanisms, BERT and GPT-4", *IEEE Access*, Vol. 12, pp. 69710-69727, 2024.
- [27] G. Ramos, "Leveraging Transfer Learning for Hate Speech Detection in Portuguese Social Media Posts", *IEEE Access*, Vol. 12, pp. 101374-101389, 2024.
- [28] Q. Chen, J. Du, A. Allot and Z. Lu, "LitMC-BERT: Transformer-based Multi-Label Classification of Biomedical Literature with an Application on COVID-19 Literature Curation", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 19, No. 5, pp. 2584-2595, 2022.
- [29] S.B. Manir, K.M.S. Islam, P. Madiraju and P. Deshpande, "LLM-based Text Prediction and Question Answer Models for Aphasia Speech", *IEEE Access*, Vol. 12, pp. 114670-114680, 2024.
- [30] J. Wang and N. Awang, "MKC-SMOTE: A Novel Synthetic Oversampling Method for Multi-Class Imbalanced Data Classification", *IEEE Access*, Vol. 8, pp. 1-8, 2024.
- [31] B. Rahman and Maryani, "Optimizing Customer Satisfaction through Sentiment Analysis: A BERT-based Machine Learning Approach to Extract Insights", *IEEE Access*, Vol. 12, pp. 1-13, 2024.
- [32] E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection using SMOTE and AdaBoost", *IEEE Access*, Vol. 9, pp. 165286-165294, 2021.
- [33] J. Wei, Z. Lu, K. Qiu, P. Li and H. Sun, "Predicting Drug Risk Level from Adverse Drug Reactions using Smote and Machine Learning Approaches", *IEEE Access*, Vol. 8, pp. 185761-185775, 2020.
- [34] N. Altwaijry, "Probability-based Synthetic Minority Oversampling Technique", *IEEE Access*, Vol. 11, pp. 28831-28839, 2023.
- [35] G.A. Pradipta, R. Wardoyo, A. Musdholifah and I.N.H. Sanjaya, "Radius-SMOTE: A New Oversampling Technique of Minority Samples based on Radius Distance for Learning from Imbalanced Data", *IEEE Access*, Vol. 9, pp. 74763-74777, 2021.
- [36] Z. Xu, X. Tang, C. Ma and R. Zhang, "Research on Parking Space Detection and Prediction Model based on CNN-LSTM", *IEEE Access*, Vol. 12, pp. 30085-30100, 2024.
- [37] M.A. Ferrag, "Revolutionizing Cyber Threat Detection with Large Language Models: A Privacy-Preserving BERT-based Lightweight Model for IoT/IIoT Devices", *IEEE Access*, Vol. 12, pp. 23733-23750, 2024.
- [38] Z. Zhou, V.W.L. Tam and E.Y. Lam, "SignBERT: A BERT-based Deep Learning Framework for Continuous Sign Language Recognition", *IEEE Access*, Vol. 9, pp. 161669-161682, 2021.
- [39] D.C.E. Saputra, K. Sunat and T. Ratnaningsih, "SMOTE-MRS: A Novel SMOTE-Multiresolution Sampling Technique for Imbalanced Distribution to Improve Prediction of Anemia", *IEEE Access*, Vol. 12, pp. 1-7, 2024.
- [40] A. Sharma, P.K. Singh and R. Chandra, "SMOTified-GAN for Class Imbalanced Pattern Classification Problems", *IEEE Access*, Vol. 10, pp. 30655-30665, 2022.
- [41] Y. Bao and S. Yang, "Two Novel SMOTE Methods for Solving Imbalanced Classification Problems", *IEEE Access*, Vol. 11, pp. 5816-5823, 2023.
- [42] K.P. Rasheed Abdul Haq and V.P. Harigovindan, "Water Quality Prediction for Smart Aquaculture using Hybrid Deep Learning Models", *IEEE Access*, Vol. 10, pp. 60078-60098, 2022.
- [43] Y. Chen and R.A. Juanatas, "Two-Stage Text Summary Model", *IEEE Access*, Vol. 2, pp. 1-13, 2024.
- [44] M. Pan, "Water Level Prediction Model based on GRU and CNN", *IEEE Access*, Vol. 8, pp. 60090-60100, 2020.