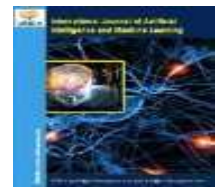




International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches

Pravin Gaupale¹, Professor Anita Yadav², Bramhadeo Wadibhasme³

¹ Email: pravin.gaupale@gmail.com

² Email: hod.cse@tgpcet.com

³ Email: bramhadeo12@gmail.com

Abstract: The new employment process in the web has revolutionized, but it also has its drawbacks like creating fake job offers and causing financial loss to job seekers. To overcome this challenge, a DL approach is proposed for ORF detection using a new dataset that was gathered from Fake Job Posting, Pakistan Job Posting and US Job Posting datasets. The technique uses BERT and RoBERTa to convert the task details into numeric vectors. To deal with the large class imbalance in the dataset, the SMOTE variation SMOBD is used to balance the classes well. This updated framework integrates these advances with a two-dimensional CNN2D for job classification. The results reveal that the model of BERT model features and SMOBD with CNN2D has the highest classification accuracy of 98.68%. This approach addresses some shortcomings of legacy datasets and gives them a solid response to the challenge of detecting a fake job posting, thereby making a significant contribution towards preventing online recruiting fraud.

“Index Terms - Class imbalance, data augmentation, deep learning, employment scam, fraud detection, machine learning, online recruitment, SMOTE, transformer-based models”.

1. Introduction

In today's technologically advanced world, the internet has revolutionised the way people search for jobs and organisations search for talent in virtually every aspect. This shift in the hiring process has largely been pushed towards a digital environment that aligns well with the productivity, convenience, and efficiency of conventional hiring practices. An online recruiting system is also referred to as E-recruitment, which allows organisations to post job vacancies and individuals seek career opportunities. [1] These systems often enable organisations to post adverts for jobs, including specifics like requirements, salary packages, benefits and other amenities. Job seekers, on the other hand, utilize these platforms to look for job openings that are relevant to their interests and talents and apply for the jobs. The firm then evaluates the resumes of the applications, shortlists the candidates and goes through interviews and other formalities to complete the recruitment process [2].

Due to physical distancing and remote working, there was a significant rise in the usage of e-recruitment platforms. There was a huge surge in the use of e-recruitment platforms as a consequence of the restrictions on physical interaction and the requirement for remote operation. The global unemployment rate rose to 13% in 2020, from 7.3% in 2019 and 3.9% in 2018, as a result of the economic challenges brought about by the pandemic, according to the World Economic Outlook Report [4]. As there was an extraordinary rise in unemployment, many organisations chose to find a way to recruit online, making it a simple and easy way for individuals to seek out opportunities during the global crisis. The job posting activity of companies shifted to online job ads to maintain recruitment activities and to meet the increasing number of job seekers who were impacted by the widespread layoffs [5].

But with the rise in use of E-recruitment platforms, there are opportunities for online fraudsters to exploit the system too. Nowadays, the fraudulent job advertising are more common. They are aimed at the vulnerable job-seekers with attractive benefits and paid employment. Such frauds can lead to financial loss, identity theft and psychological distress to people seeking legitimate job opportunities [6]. The rise of these types of fraudulent operations highlights the necessity of having broad tools that can be used to detect and prevent online recruitment fraud. Avoiding job searchers from falling prey to frauds has become a critical part for ensuring confidence and authenticity in online recruitment platforms [7].

Therefore, it will help to decrease the risk of the online recruiting fraud by introducing new technology and intelligent system into the E-recruitment platform, which, in turn, will boost the efficiency and reliability of the E-recruitment platform to address these challenges. This paper discusses novel methods and solutions

to tackle these difficulties to provide a more secure and efficient online recruitment environment for both the businesses and job seekers.

2. Related Work

Sure, online recruiting services are very popular but they're also the subject of a number of frauds that can be very dangerous for the job searcher. To reduce the negative effects of this, various researchers have explored various strategies for determining when a job is a scam. ANNs have been employed for online recruiting fraud identification and proved the capability of ANNs to effectively classify fraud online posts. The authors of Nasser et al. [3] employed ANN for pattern recognition in recruiting fraud and they found that the model was able to learn and adapt the complex relationships present in the data and achieve good classification performance.

Fraud detection is another typical use of ML techniques. Lokku [4] suggested to study the classification of the legitimacy of a job post by analysing textual and structural features of job posts by applying ML algorithms. The study also showed the possibility of supervised learning techniques to identify trends that indicate fraudulent activity and laid the foundation for further research of data-driven approaches.

In the comparative study proposed by Habiba et al. [5] they have studied various algorithms of data mining that are used to foresee bogus job postings. In this study, we experimented with various algorithms and discussed their advantages and disadvantages with regards to detecting fake job postings. The results pointed out the necessity of model and pre-processing choices to boost the prediction accuracy especially in the presence of imbalanced data sets.

The peculiarities and behaviour of fake job ads were taken into account by Vidros et al. [7] who presented an automatic detection system for on-line recruiting fraud. The paper provided valuable insights into the linguistic characteristics of fraudsters who deceive job seekers from a public data set and ML approaches. The study showed the need of powerful feature extraction strategies to improve the effectiveness of detection.

The detection of bogus job recruitment was investigated by Dutta and Bandyopadhyay [8] by various ML methods. They emphasized the importance of feature engineering as a strategy to enhance model performance and demonstrated the potential of different algorithmic techniques to achieve better results. The study identified the challenges of using real data sets, including imbalance and noise, and proposed methods to address these issues. The study pointed out the difficulties in the application of real data sets (imbalance and noise) and presented some possible solutions to these difficulties.

Alghamdi and Alharby [9] introduced an intelligent way of online recruiting fraud detection through the use of advanced ML techniques for analysing the job posts. They were looking for something that would get their attention, some linguistic pattern, some job description anomaly, something significant to the fraudulent operations. The study showed how the combination of multiple features and algorithms can improve the accuracy of detection and result in more advanced detection systems.

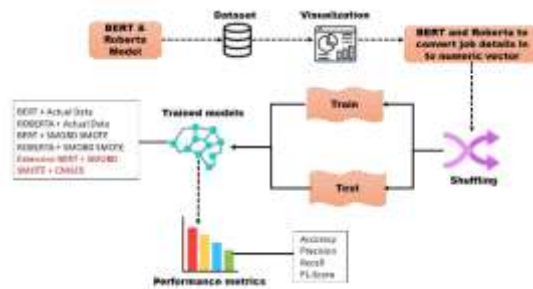
Lal et al. [10] also demonstrated that Ensemble learning is a well suited tool in the field of fraud detection. They created an ensemble model called ORFDetector, which they used a group of classifiers to boost prediction accuracy. The study showed the effectiveness of using multiple algorithms together to reduce the errors and error rates of the detection system, particularly if the data sets were unbalanced.

Behavioural feature extraction has also been proven to be a useful approach to identifying fake job postings. Nindyati and Nugraha [13] investigated the behaviour of the users through the reaction patterns and user engagement to identify the scams in online job ads. Their work has demonstrated the potential of the use of behavioural data to complement textual and structural data, and led to a greater understanding of fraudulent behaviours.

The progress in online recruitment fraud detection and the significance of sophisticated algorithms, feature extraction methods and data preparation methods are demonstrated in these studies. But there are still some important opportunities for improving the quality of the dataset, mitigating class imbalance, and choosing features. Past studies can be the basis for developing new techniques to address these limitations and improve the effectiveness of fraud detection systems.

3. Materials And Methods

The proposed solution is a new dataset gathered from the source of Fake Job Posting [16], Pakistan Job Posting [18] and US Job Posting [17] databases, which uses ORF detection. The technology is built on cutting edge DL algorithms, which help to increase the quality of fraud detection. We use BERT [15] and RoBERTa [12] for vectorisation of the job description. First, BERT and RoBERTa are used on the raw data to analyze it. For this reason, to address the class imbalance issue, we have introduced the variation of SMOTE [14] called SMOBD which is effective for class balancing. The technique is further enhanced by incorporating BERT and RoBERTa features with SMOBD enriched dataset. The method finally ends with a strong integration of BERT characteristics with SMOBD and CNN2D for job categorisation assuring a comprehensive and effective solution for recognising fraudulent job posts.



“Fig.1 Proposed Architecture”

It utilizes BERT models [15] and Roberta models [12] for numerical representation of work information. Different ML algorithms such as SMOTE-CNN2D, SMOTE [14] and SMORD are trained using these vectors. Performance indicators including accuracy, precision, recall, and F1-score are used to assess and evaluate trained models. The shuffling process is performed in order to make the model robust.

i) Dataset Collection:

The collection includes job postings marked as phony or fraudulent [16] and gathered from various sources on the Internet. It includes all job descriptions, company names and more. The data set is used for model training and validation to identify fake job postings. The data are read from a .csv format and analysed for feature extraction [13] and classification purposes.

Job ID	Job Title	Company Name	Location	Description	Experience Required	Salary	Is Fake
1	Software Engineer	ABC Inc	New York	Design, develop, and maintain software applications.	3-5 years	\$80,000	0
2	Marketing Manager	XYZ Corp	Los Angeles	Develop and execute marketing strategies to promote products and services.	5-7 years	\$95,000	0
3	Product Designer	DEF Ltd	San Francisco	Design user interfaces and product features for web and mobile applications.	2-4 years	\$75,000	0
4	Business Development	GHI LLC	Chicago	Identify and develop new business opportunities and partnerships.	3-6 years	\$85,000	0
5	Customer Support	JKL Inc	Remote	Provide timely and effective support to customers via email, chat, or phone.	1-2 years	\$45,000	0
6	Project Manager	MNO Corp	Seattle	Plan, execute, and close projects, ensuring they are completed on time and within budget.	4-6 years	\$90,000	0
7	Quality Assurance	PQR Ltd	Phoenix	Test software applications for bugs and errors, ensuring high quality.	2-3 years	\$60,000	0
8	Operations Manager	STU Inc	Denver	Oversee daily business operations, manage staff, and optimize processes.	5-8 years	\$100,000	0
9	UX Researcher	VWX Corp	Portland	Conduct user research to understand user needs and inform product design.	3-5 years	\$82,000	0
10	Systems Administrator	YZA LLC	San Diego	Manage and maintain IT infrastructure, including servers and networks.	4-7 years	\$78,000	0

“Fig.2 Fake Job Posting Dataset”

The collection is comprised of job listings from Pakistan [18] that were collected from December 2019 to March 2021. Every entry is a job posting with an assortment of characteristics like job title, company, and description. A label of "0" is added to authentic job advertisements. This dataset aims to facilitate the detection and categorisation of fake job postings in the Pakistani job market.

Job ID	Job Title	Company Name	Location	Description	Experience Required	Salary	Is Fake
1	Software Engineer	ABC Inc	New York	Design, develop, and maintain software applications.	3-5 years	\$80,000	0
2	Marketing Manager	XYZ Corp	Los Angeles	Develop and execute marketing strategies to promote products and services.	5-7 years	\$95,000	0
3	Product Designer	DEF Ltd	San Francisco	Design user interfaces and product features for web and mobile applications.	2-4 years	\$75,000	0
4	Business Development	GHI LLC	Chicago	Identify and develop new business opportunities and partnerships.	3-6 years	\$85,000	0
5	Customer Support	JKL Inc	Remote	Provide timely and effective support to customers via email, chat, or phone.	1-2 years	\$45,000	0
6	Project Manager	MNO Corp	Seattle	Plan, execute, and close projects, ensuring they are completed on time and within budget.	4-6 years	\$90,000	0
7	Quality Assurance	PQR Ltd	Phoenix	Test software applications for bugs and errors, ensuring high quality.	2-3 years	\$60,000	0
8	Operations Manager	STU Inc	Denver	Oversee daily business operations, manage staff, and optimize processes.	5-8 years	\$100,000	0
9	UX Researcher	VWX Corp	Portland	Conduct user research to understand user needs and inform product design.	3-5 years	\$82,000	0
10	Systems Administrator	YZA LLC	San Diego	Manage and maintain IT infrastructure, including servers and networks.	4-7 years	\$78,000	0

“Fig.3 Pakistan Job Posting Dataset”

This data set is related to real estate marketing and has job postings data from the USA [17]. It has job titles, descriptions, pertinent information about business names and more. The data is labelled with 0 for authentic job postings. It is employed to train algorithms to identify fake job ads in the U.S. job market.

Job ID	Job Title	Company Name	Location	Description	Experience Required	Salary	Is Fake
1	Software Engineer	ABC Inc	New York	Design, develop, and maintain software applications.	3-5 years	\$80,000	0
2	Marketing Manager	XYZ Corp	Los Angeles	Develop and execute marketing strategies to promote products and services.	5-7 years	\$95,000	0
3	Product Designer	DEF Ltd	San Francisco	Design user interfaces and product features for web and mobile applications.	2-4 years	\$75,000	0
4	Business Development	GHI LLC	Chicago	Identify and develop new business opportunities and partnerships.	3-6 years	\$85,000	0
5	Customer Support	JKL Inc	Remote	Provide timely and effective support to customers via email, chat, or phone.	1-2 years	\$45,000	0
6	Project Manager	MNO Corp	Seattle	Plan, execute, and close projects, ensuring they are completed on time and within budget.	4-6 years	\$90,000	0
7	Quality Assurance	PQR Ltd	Phoenix	Test software applications for bugs and errors, ensuring high quality.	2-3 years	\$60,000	0
8	Operations Manager	STU Inc	Denver	Oversee daily business operations, manage staff, and optimize processes.	5-8 years	\$100,000	0
9	UX Researcher	VWX Corp	Portland	Conduct user research to understand user needs and inform product design.	3-5 years	\$82,000	0
10	Systems Administrator	YZA LLC	San Diego	Manage and maintain IT infrastructure, including servers and networks.	4-7 years	\$78,000	0

“Fig.4 US Job Posting Dataset”

ii) Load BERT & Roberta Model:

Here, the BERT and RoBERTa models are loaded from SentenceTransformer library to convert task details to numerical vectors. We choose RoBERTa [12] and BERT [15] models for being able to perform NLI and accurately capture the meaning of text. Once loaded, the models are ready to analyse job descriptions and obtain valuable embeddings that are critical for the next phases of fraud detection.

iii) Pre-Processing:

The BERT and RoBERTa models are used as a pre-processing step to convert the work details into numeric vectors. Randomise the data, shuffle it and look at the distribution of data using visualisation techniques.

a) Visualization: Seaborn and Matplotlib are used to visualize the data and understand the distribution of fake job advertisements. The first plot illustrates the number of fraudulent and non-fraudful postings by employment type to gain an understanding of the association between different employment categories and fraud. The second graph shows the percentage of fake job posts by experience level. These visualisations can help you recognise patterns and trends in the dataset, and understand the factors that impact on recruiting fraud.

b) Vectorization: In this step, a description of jobs from three datasets are read, cleaned and prepared for feature extraction [13]. The descriptions and labels are taken and processed from the datasets to create BERT and RoBERTa embeddings. BERT encoding to transform job descriptions to numeric vectors with the help of BERT model. RoBERTa for encoding to get tensor based representations. These embeddings are stored until they can be used for the training of the models. This method guarantees that the job text data is vectorised suitably for the purpose of fraud detection.

c) Shuffling: This randomizes the data set and helps to prevent any skewing of the model when training it. The BERT and RoBERTa [12] embeddings and their accompanying labels are jumbled with random indices. This results in well-mixed data which is important for efficiently training models using ML. Then, the shuffled data set is ready to be separated into training and testing sets. This allows for a more robust evaluation of the model's performance and generalisability in detecting fake job listings.

iv) Training & Testing:

The dataset is split into training and test sets, with both BERT and RoBERTa embeddings. 80% of the data is utilised for training and 20% of the data is used for testing. The labels have been one hot encoded for classification. Features are loaded and processed for both BERT and RoBERTa and train and test sets are built for both models. This guarantees that the model gets trained on a range of data and tested on data not seen during training, which helps them perform well.

v) Algorithms:

BERT + Actual Data: Integrating the job description into meaningful vector to obtain semantic features from the text by using BERT [15]. Its pre-trained language model facilitates the precise understanding and classification of text helping to classify job posts efficiently as bogus and legitimate.

RoBERTa + Actual Data: Using a more robust BERT, RoBERTa [12] to generate embeddings of job descriptions. It can improve the accuracy of detecting the fraudulent job posts by better processing complex and various text structures to give full understanding of text for classification.

BERT + SMOBD SMOTE: SMOBD SMOTE [14] is used to solve class imbalance by synthesising samples of the minority class. These samples are then utilised to obtain BERT embeddings which are used to classify job listings. This in turn helps the model to better detect fake postings by enriching the training data.

RoBERTa + SMOBD SMOTE: The use of RoBERTa embeddings in combination with SMOBD SMOTE, not only balances the dataset, but also enhances the representation of the job descriptions. This mixture of synthetic data and robust feature extraction ability of RoBERTa [12] enhances the classification accuracy of bogus job postings detection.

BERT + SMOBD SMOTE + CNN2D: This method uses BERT embedding and SMOBD SMOTE data balancing method and CNN2D feature extraction. CNN2D helps to learn the spatial relationship in embeddings that enhances the classification accuracy for fraud detection in job advertisements.

4. Results & Discussion

Accuracy: The power of a test to correctly identify patients as opposed to healthy cases. The accuracy of the test should be estimated by dividing the number of true positive by the number of true negative in all test situations. This is mathematically expressed as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

Precision: Precision is the proportion of correctly classified occurrences or samples to the total number of occurrences or samples classified as positive. So, the formula for precision is:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

Recall: It is a metric used for ML to measure how well a model can identify all the relevant examples for a specific class. This is the ratio of correctly predicted positive observations to the number of positive observations, the ratio, or the rate of identification of all positive observations.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

F1-Score: F1 score is one of the statistics in ML that measures the accuracy of a model. It is a combination of precision and recall score of a . The accuracy is the total number of times a model was correct for the whole data set.

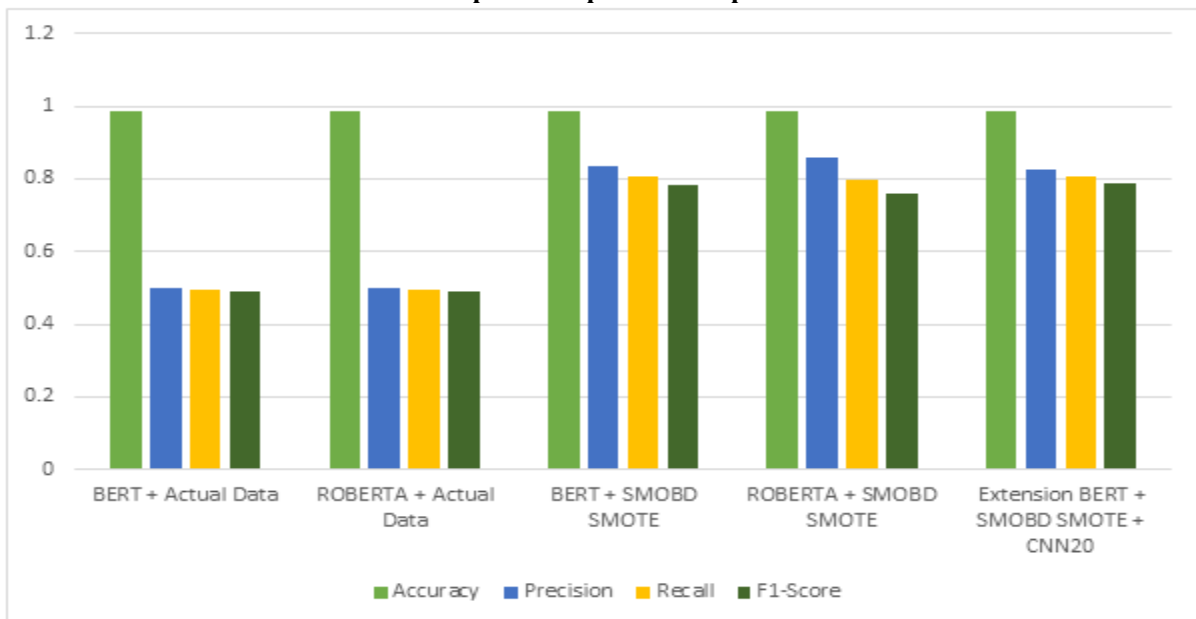
$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100 (1)$$

The performance indicators, namely accuracy, precision, recall and F1-score for each method are evaluated and shown in Table 1. The best scores are obtained by BERT + SMOBD SMOTE + CNN2D. The same metrics are included in the table for comparison with other algorithms.

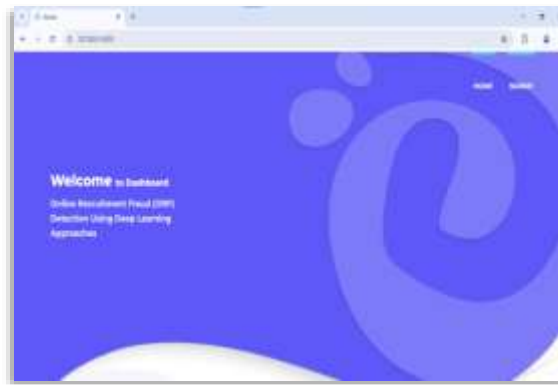
“Table.1 Performance Evaluation Metrics”

Algorithm Name	Accuracy	Precision	Recall	F1-Score
BERT + Actual Data	0.9839	0.5000	0.4959	0.4919
ROBERTA + Actual Data	0.9850	0.5000	0.4962	0.4925
BERT + SMOBD SMOTE	0.9866	0.8339	0.8065	0.7834
ROBERTA + SMOBD SMOTE	0.9858	0.8577	0.7992	0.7577
Extension BERT + SMOBD SMOTE + CNN2D	0.9868	0.8256	0.8052	0.7872

“Graph.1 Comparison Graphs”



The accuracy is represented in light green, the precision in blue, the recall in light yellow and the F1 score in green (see graph 1). In all the metrics, the BERT + SMOBD SMOTE + CNN2D algorithm is better than the other algorithms, with the highest metrics compared to the other models. These details are shown in the above graph.



“Fig.5 Home Page”

Welcome message with user interface navigation and dashboard above fig.5

A screenshot of a 'New Account' registration form. The form is titled 'New Account' and contains five input fields: 'Username', 'Name', 'Email', 'Mobile Number', and 'Password'. Below the fields, there is a 'Remember me' checkbox which is checked, and a 'Forgot Password' link. A blue 'Register' button is at the bottom. At the very bottom, there is a link that says 'Already have an account? Sign in'.

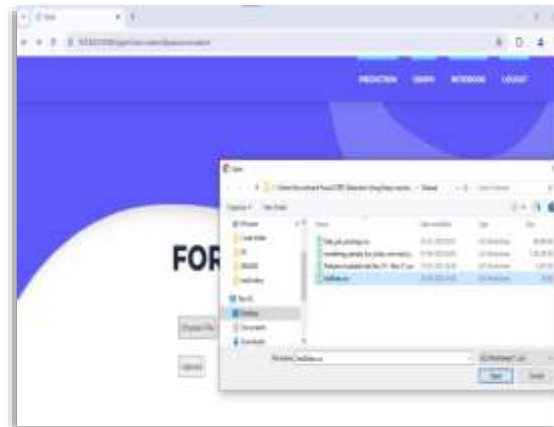
“Fig.6 Registration Page”

In above fig.6 Sign up form consists of fields like username, name, email, mobile number and buttons (password).

A screenshot of a 'Log In' form. The form is titled 'Log In' and contains two input fields: 'username' with the value 'admin' and 'password' with masked characters '.....'. Below the fields, there is a 'Remember me' checkbox which is checked, and a 'Forgot Password' link. A blue 'Log In' button is at the bottom. At the very bottom, there is a link that says 'Register here! Sign Up'.

“Fig.7 Login Page”

fig.7 Sign-in form with username and password boxes, “Remember Me”, “Forgot Password” in top.



“Fig.8 Upload Input Page”

The upper form of figure 8 has a coordinate input area and an upload button.



“Fig.9 Predict Result for given input”

As shown in above Fig.9 Predicted result based on input test data.

5. Conclusion

Last but not least, the recommended solution to the problem of ORF is able to address the growing activity of fake job ads on online platforms. The system increases its ability to correctly detect bogus job ads by incorporating a number of sophisticated algorithms for DL, including BERT and RoBERTa. The use of a fresh dataset collected from different sources and the usage of SMOTE SMOBD approach largely alleviates class imbalance problems, assuring the sound training and evaluation of the models. The results indicate that the best accuracy of 98.68% was obtained by the model which is composed of BERT combined with SMOBD and merged with a CNN2D. The results show the proposed technique is effective to differentiate real and fake job listings. This comprehensive method of detecting ORF is an important basis that may help safeguard job hunters from online scams and eventually pave the way for a safer recruitment process in the digital era.

The future work on this project will focus on further enhancement of detection of online recruitment fraud by investigating other ML approaches such as ensemble methods and more sophisticated feature extraction algorithms. A potential improvement for the model is the incorporation of RNNs and attention mechanisms that can capture the context from the job advertisements. You may also play with transfer learning from pre-trained models to optimise the performance on smaller datasets. The enhancements aim to improve the detection and efficiency of fake job posting identification.

References

- [1] G. Othman Alandjani, “Online fake job advertisement recognition and classification using machine learning,” 3C TIC, Cuadernos de Desarrollo Aplicados a las TIC, vol. 11, no. 1, pp. 251–267, Jun. 2022.
- [2] A.Adhikari, A.Ram, R.Tang, and J.Lin, “DocBERT: BERT for document classification,” 2019, arXiv:1904.08398.

- [3] I. M. Nasser, A. H. Alzaanin, and A. Y. Maghari, "Online recruitment fraud detection using ANN," in Proc. Palestinian Int. Conf. Inf. Commun. Technol. (PICICT), Sep. 2021, pp. 13–17.
- [4] C. Lokku, "Classification of genuinity in job posting using machine learning," Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, no. 12, pp. 1569–1575, Dec. 2021.
- [5] S. U. Habiba, Md. K. Islam, and F. Tasnim, "A comparative study on fake job post prediction using different data mining techniques," in Proc. 2nd Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST), Dhaka, Bangladesh, Jan. 2021, pp. 543–546.
- [6] Report Cyber. Accessed: Jun. 25, 2022. [Online]. Available: <https://www.actionfraud.police.uk/>
- [7] S. Vidros, C. Kolias, G. Kambourakis, and L. Akoglu, "Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset," Future Internet, vol. 9, no. 1, p. 6, Mar. 2017.
- [8] S. Dutta and S. K. Bandyopadhyay, "Fake job recruitment detection using machine learning approach," Int. J. Eng. Trends Technol., vol. 68, no. 4, pp. 48–53, Apr. 2020.
- [9] B. Alghamdi and F. Alharby, "An intelligent model for online recruitment fraud detection," J. Inf. Secur., vol. 10, no. 3, pp. 155–176, 2019.
- [10] S. Lal, R. Jiaswal, N. Sardana, A. Verma, A. Kaur, and R. Mourya, "ORFDetector: Ensemble learning based online recruitment fraud detection," in Proc. 12th Int. Conf. Contemp. Comput. (IC3), Noida, India, Aug. 2019, pp. 1–5.
- [11] Kanika, J. Singla, A. Kashif Bashir, Y. Nam, N. UI Hasan, and U. Tariq, "Handling class imbalance in online transaction fraud detection," Comput. Mater. Continua, vol. 70, no. 2, pp. 2861–2877, 2022.
- [12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, arXiv:1907.11692.
- [13] O. Nindyati and I. G. Bagus Baskara Nugraha, "Detecting scam in online job vacancy using behavioral features extraction," in Proc. Int. Conf. ICT Smart Soc. (ICISS), vol. 7, Bandung, Indonesia, Nov. 2019, pp. 1–4.
- [14] N.V.Chawla, K.W.Bowyer, L.O.Hall, and W.P.Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.
- [15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, arXiv:1810.04805.
- [16] S. Bansal. (2020). [Real or Fake] Fake Job posting Prediction. [Online]. Available: <https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction>
- [17] Indeed Job Posting Dataset. Accessed: Feb. 16, 2023. [Online]. Available: <https://www.kaggle.com/datasets/promptcloud/indeed-job-posting-dataset>
- [18] Pakistan's Job Market. Accessed: Feb. 16, 2023. [Online]. Available: <https://www.kaggle.com/datasets/zusmani/pakistans-job-market>