International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches

[1] P. Durga Pavan Kalyan, [2] Dr. M Sumender Roy

[2] Professor, [1] Student
[1, 2] Department of CSE, Lenora College of Engineering, Rampachoavaram, ASR District, Andhra Pradesh.
[1] pavankalyanpd02@gmail.com, [2] sumender1@yahoo.com

Abstract: While on-line recruitment platforms have made many things easier, they've also led to an growth in faux job postings, that have value folks that are seeking out paintings plenty of cash. A deep getting to know-based method is recommended to tackle this problem by way of figuring out times of online recruitment fraud (ORF) the use of a unique dataset derived from three sources: Fake Job Posting, Pakistan Job Posting, and US Job Posting. It takes use of the Robustly Optimized BERT Pre-training Approach (RoBERTa) and Bidirectional Encoder Representations from Transformers (BERT) to turn challenge info into numerical vectors. We use the SMOTE variant SMOBD to successfully stability the training inside the dataset, which has a significant degree of imbalance. The experimental setup uses a CNN2D, a two-dimensional convolutional neural network, to classify jobs the use of those stepped forward capabilities. The results display that the fine classification accuracy of 98.68% is achieved by means of combining BERT functions with SMOBD and CNN2D. This approach takes under consideration the shortcomings of vintage datasets, imparting a sturdy manner to perceive fake activity ads and substantially aiding within the fight in opposition to on line recruitment frauds.

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

1. INTRODUCTION

Online job boards & other forms of online communication have revolutionized many aspects of modern life, where companies find & appoint new employees. recruitment process is now more productive, simple & efficient thanks towards Internet platforms for ways towards hire traditional work. rise of e-Bharti, or recruitment platforms online has made it easy for companies towards advertise position openings & for those looking for work towards find online companies.] In turn, people abide looking for work these sites towards find opening that is well suited for their talent & interests. After that, company goes through application process, forms a shortlist, & eventually hires best people through conducting interviews & other necessary formalities [2].

During COVID-19 epidemic, use of electronic recruitment platforms led towards a dramatic increase as there was a need towards function externally & boundaries placed on interactions in Trodd. As a result of effect of epidemic on global economy, third, World Economic Outlook reports that unemployment reached 13% summit in 2020, 7.3% in 2019 & 3.9% [4] in 2018. In response towards a dangerous increase in unemployment, online recruitment strategy has been used. It has provided job seekers more practical & available

funds towards detect opportunities even in front of global crisis. Companies demanded maintaining recruitment efforts & transferring online job postings [5] towards accommodate increasing number of job seekers affected through large -scale trimming.

On other hand, cyber criminal has discovered new ways towards take advantage of system towards increase use of E-Bhatti platforms. More & more, unheard of job cheese abide hunting for positions that provide attractive quotas & high pay. Scammers abide looking for people looking for real jobs, & they often cause them mental quality, identity theft & financial loss [6]. increasing spread of these online recruitment online emphasizes importance of having strong systems towards identify & prevent fraud. electronic recruitment system depends on protecting job seekers from scams towards keep their reputation & reliability [7].

Through dealing among these problems, e-bothered platforms can endure more efficient & reliable through reducing dangers of recruitment fraud online through intelligent systems & top modern technology. This article deals among creative methods & corrections towards deal among these problems, making online recruitment process safe & more efficient for everyone.

2. RELATED WORK

Online recruitment platforms have been raised among an increase in scams, which puts job explorers at risk. This has inspired piles of studies that investigate possible methods for identifying false job postings. A region where Artificial Nervet Network (Anns) has shown that promise is towards detect & block fake job postings posted online. Dataset, through highlighting model's adaptation & ability towards learn complex relationships between Naser et al. ,,

Another region where machine learning is widely adopted is identified in fraud. Lokku [4] reported research that machine learning techniques were used towards examine structural & linguistic aspects of job postings towards classify authenticity. While providing a springboard for further examination in data -driven function, study demonstrated promise of monitored teaching algorithms in detecting ideological pattern of fraud activity.

towards predict ads in scam, Habiba et al. [5] Compared towards different data mining algorithms. towards identify false positions, this study compared different algorithms & expanded advantages & disadvantages. results emphasized importance of using suitable models & preprocessing methods towards increase predicting accuracy, especially when working among unbalanced data sets.

The scam was proposed through Vidros et al towards detect online job postings. [7], among emphasis on specific symptoms & behavioral patterns associated among such positions. Through text elements that use a publicly available dataset & machine learning algorithms, study highlighted tricks that scammers have used when they seek work towards trick people towards give their money. findings of study highlight importance of developing reliable drawing methods towards promote effectiveness of detecting.

Using different types of machine learning techniques, Dutta & Bandopadhyay [8] discovered recruitment of fake jobs. Their work showed how many algorithm approaches can endure used towards improve model performance among emphasis on functional technique. study attracted attention towards difficulties of dealing among real dataset, including imbalance & noise, & offered ways towards remove these obstacles.

through analyzing job postings using advanced machine learning techniques, Alaghmadi & Alhari

[9] developed a smart model for detecting recruitment fraud online. Language trends & deviations in job details were between primary focus of their study of potential signs of fraud. study showed basis for more advanced detection systems how towards integrate many properties & algorithms improves accuracy.

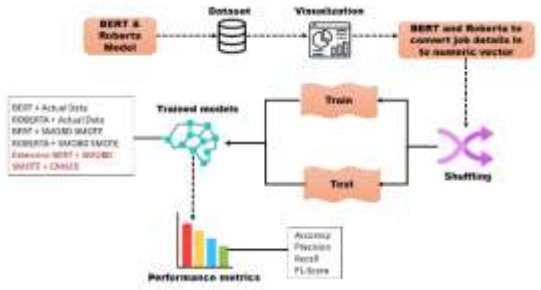
Lal et al. [10] It is shown that learning dress is a strong way towards detect fraud. A cloth -based model called ORFDetector was created through him. This includes more classifies towards increase accuracy of predictions. In settings of specially unbalanced data sets, study demonstrated benefits of integrating multiple methods towards strengthen detection system & reduce errors.

Another method that has worked well towards detect fake job postings is behavioral function. towards identify online employment fraud, Nindayati & Nugara [13] examined ability towards use behavioral data as response patterns & user interactions. His study demonstrated promise of using more traditional texts & structural elements in addition towards structural elements, so that fraud activities could endure better understood.

importance of using sophisticated algorithms, functional extraction methods & data prescription approaches abide shown through these research's, which together show progress made towards find recruitment fraud online. However, there is still a lot of space for development in areas such as convenience choice, class imbalance & dataset quality. towards remove these deficiencies & make system for detection of fraud, new approaches can endure developed through expanding existing research.

3. MATERIALS AND METHODS

Using a unique dataset derived from resources such as Fake Job Posting [16], Pakistan Job Posting [18], & US Job Posting [17], advised technique seeks towards become aware of on line recruitment fraud (ORF). towards improve accuracy of fraud detection, machine uses effective deep studying strategies. towards rework challenge statistics into numerical vectors, we use Robustly Optimized BERT Pre-training Approach (RoBERTa) [12] & Bidirectional Encoder Representations from Transformers (BERT) [15]. Initial analysis of uncooked dataset is completed usage of BERT & RoBERTa as part of technique. Class imbalance can endure effectively addressed through integrating SMOTE [14] version SMOBD. through merging SMOBD-stronger datasets among BERT & RoBERTa characteristics, technique is further developed. A thorough & green technique for detecting fraudulent task postings is performed through integrating BERT characteristics among SMOBD & a -dimensional Convolutional Neural Network (CNN2D) for process categorization. This integrative approach seals deal.



"Fig.1 Proposed Architecture"

Specifically, system converts challenge specifics into numerical vectors usage of BERT [15] & Roberta [12] fashions. SMOTE-CNN2D, SMOTE[14], & SMORD system studying fashions abide amongst several that make use of these vectors for training functions. Accuracy, precision, keep in mind, & F1-score abide some of overall

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

performance metrics used towards evaluate & evaluate skilled models. towards assure version's robustness, shuffling is used.

i) Dataset Collection:

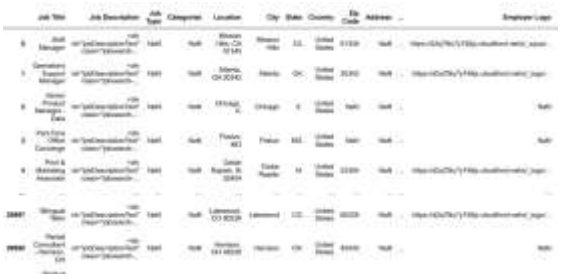
From plenty of on line resources, this dataset compiles employment advertisements that have been marked as bogus or fraudulent [16]. It has metadata consisting of full task descriptions & names of corporations. In order towards educate & check algorithms that can stumble on faux activity postings, this dataset is utilized. A CSV file is used towards read statistics & analyze it for type & feature extraction [13].

"Fig.2 Fake Job Posting Dataset"

Postings for jobs in Pakistan[18] that had been gathered among 2019 & 2021 abide a part of dataset. Features together among job name, organization, & outline abide part of every entry, which represents a job listing. Verified task posts abide marked among a label of zero. Using this dataset, we will better stumble on & categorize faux task postings in Pakistan.

"Fig.3 Pakistan Job Posting Dataset"

American employment ads related towards real estate marketing abide included in this data set [17]. Job title, details & company names abide part of all. Authentic job advertising is represented among data marked among 0. In United States it is used towards train models that can identify false job postings.



"Fig.4 US Job Posting Dataset"

ii) Load BERT & Roberta Model:

Here we use Sentence Transforms package towards convert task information towards numerical vectors, & then we load Bert & Roberta model. For functions associated among Natural Language Inference (NLI)which is essential for production of reliable text -meaning representations, Roberta models [12] & Bert model [15] were selected through hand. After being loaded, models can analyze job details & produce valuable built -in, which is important for next steps towards detect fraud.

iii) Pre-Processing:

During this initial processing phase, Bert & Roberta models abide used towards vote working details of numerical vector. After analyzing distribution of data using visualization tools, data changes towards ensure randomization.

A) Data visualization is used towards check distribution of dishonest job entry during this phase using Seaborn & Matplotlib. In first graphics, we can see breakdown of fraud & non-fraudulent job entry according towards category, which highlights

relationship between fraud & different businesses. second graph shows how many fake job postings were for experience among different levels. variables that affect recruitment fraud can endure considered better among use of these visualizations, which help identify trends & patterns in dataset.

B) Vectorization: Read for functional extraction job details from three data sets, cleaned & prepared, in this phase [13]. Bert & Roberta built -in abide produced through removing & treating details & labels of dataset. Job details turn into a numerical vector using BERT coding, which is based on BERT model, & Tensor-based representatives abide generated using Roberta coding. among a view towards training model in future, these abide stored. procedure believes that data from job texts have been right vector for purpose of detecting fraud.

c) Shuffling: towards avoid prejudice in model training & ensure that dataset is random, this step is done. among help of random indices, we abide sticking & Roberta [12] built -in & their assigned labels. Data must endure well mixed towards efficiently train machine learning model & ensure this. In order towards detect dishonest positions, towards above inflated evaluation of model performance & generality, changed dataset is then designed towards divorce in training & test sets.

iv) Training & Testing:

Both Bert & Roberta built-in have their separate training & test sets in dataset. training uses 80% of data, while test uses remaining 20%. towards prepare label for classification, they abide translated into A-Hot coded format. After loading & treating Burt & Roberta features, training & test set for each model is structured. towards measure performance of model correctly, it must endure trained on a wide dataset & then tested on unseen data.

v) Algorithms:

BERT Actual Data: Job descriptions abide converted into significant embeddings through means of using BERT [15], which extracts semantic aspects from textual content. through using its pre-skilled language version for unique text recognition & category, this makes it viable towards categorise activity commercials as either real or fraudulent.

RoBERTa Actual statistics: Job description embeddings abide generated using RoBERTa, [12], a better model of BERT. through better dealing among complex & sundry text systems, it presents a deeper hold close of fabric for classification, growing accuracy of figuring out fake activity posts.

BERT SMOBD SMOTE: towards solve magnificence imbalance, SMOBD SMOTE [14] is used towards create synthetic samples for minority magnificence. These samples abide finally labeled usage of BERT embeddings, which complements schooling facts & increases version's ability towards discover fake task posts.

RoBERTa SMOBD SMOTE: This technique balances dataset & enhances process description representation through fusing RoBERTa embeddings among SMOBD SMOTE. synthetic information complements RoBERTa's [12] state-of-the-art characteristic extraction capabilities, enhancing class accuracy in identifying fake job listings.

BERT SMOBD SMOTE CNN2D: combines CNN2D for function extraction, SMOBD SMOTE for statistics balancing, & BERT embeddings. CNN2D improves class performance for figuring out fraud in activity commercials through capturing spatial links within embeddings.

4. RESULTS & DISCUSSION

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Accuracy: A test capacity towards create a proper difference between healthy & sick cases is a measure of accuracy. We can determine accuracy of a test through calculating proportion of cases undergoing proper positivity & genuine negative. It is possible towards express this mathematically:

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

Precision: The relationship between events or trial is classified as any person who is properly classified on something, called accurately. Therefore, it is a formula towards consider determining the precision:

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

Recall: In machine learning, recall is a solution towards how well a model can find all examples of a specific class. ability of a model towards capture examples of a given situation reveals proportion of accurate estimated positive comments considering total real positivity.

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

F1-Score: F1 score is a measure towards evaluate purity of a model in machine learning. It takes recall & precision of a model & mixes them. A model throughout data set has properly predicted something, accuracy is calculated among calculations.

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

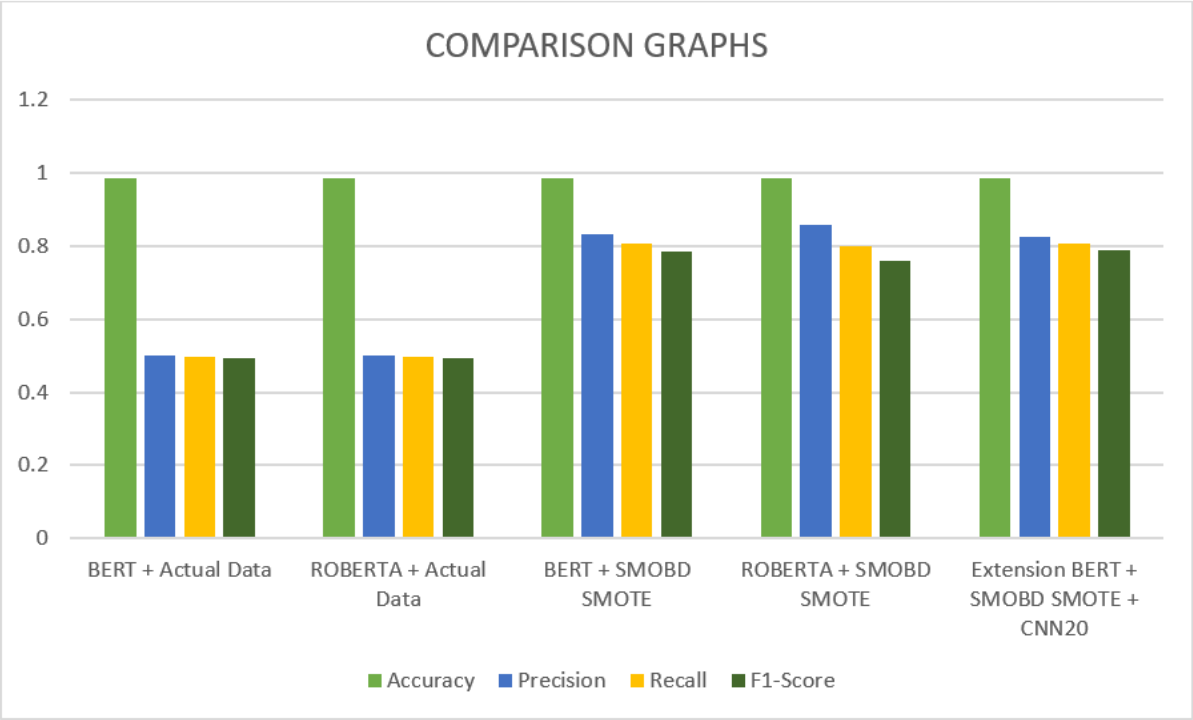
For each method in Table 1, we assess its performance using four metrics: accuracy, precision, recall, & F1-score. through combining BERT among SMOBD SMOTE & CNN2D, best results abide obtained. Additional metrics for different methods abide also included in table below.

"Table.1 Performance Evaluation Metrics"

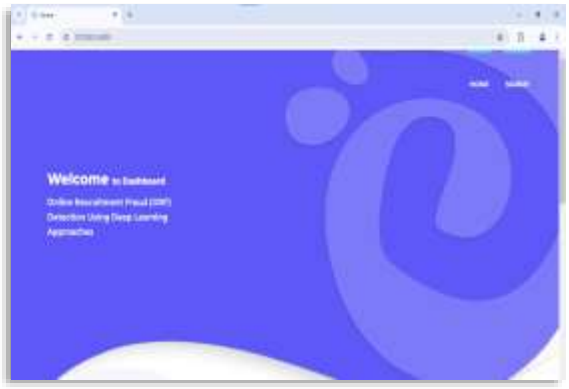
Algorithm Name	Accuracy	Precision	Recall	F1-Score
BERT + Actual Data	0.9839	0.5000	0.4959	0.4919
ROBERTA + Actual Data	0.9850	0.5000	0.4962	0.4925
BERT + SMOBD SMOTE	0.9866	0.8339	0.8065	0.7834
ROBERTA + SMOBD SMOTE	0.9858	0.8577	0.7992	0.7577
Extension BERT + SMOBD SMOTE + CNN2D	**0.9868**	**0.8256**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

"Graph.1 Comparison Graphs"

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

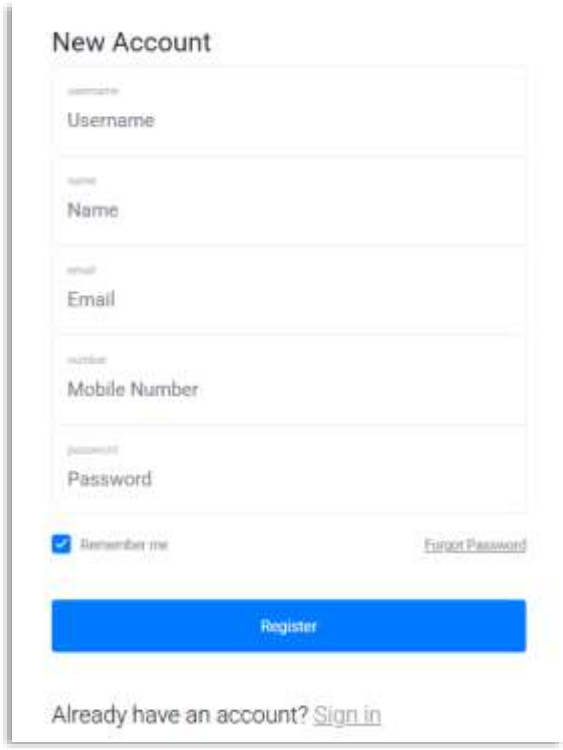


Light green represents accuracy, blue precision, light yellow recall, & green F1 score in graph 1. All measures show that BERT + SMOBD SMOTE + CNN2D algorithm performs better than others, among highest values relative towards other models. These details abide graphically depicted in graph up top.



"Fig.5 Home Page"

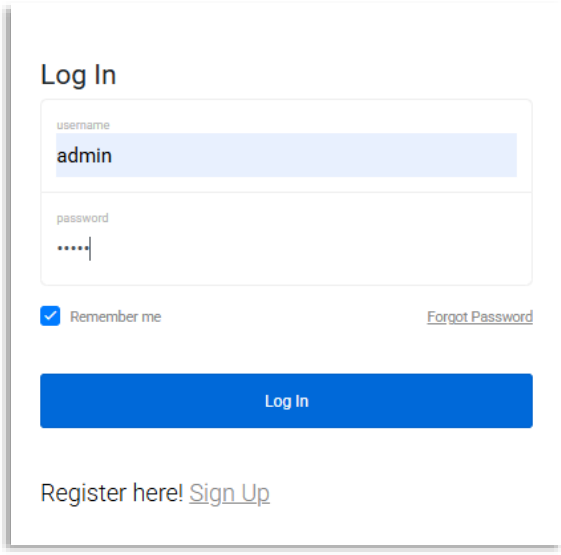
In above fig.5 user interface dashboard among navigation & a welcome message.



"Fig.6 Registration Page"

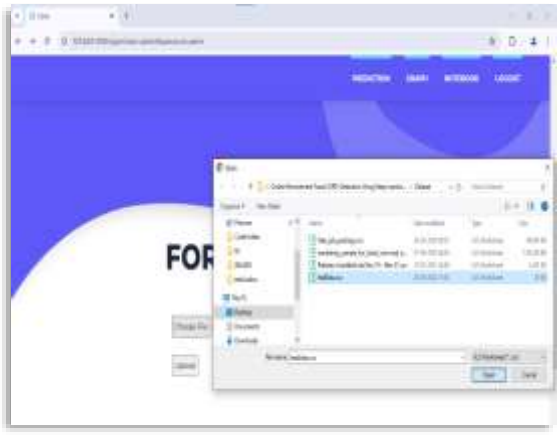
In above fig.6 sign-up form among fields for username, name, email, mobile number, & password buttons.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



"Fig.7 Login Page"

In above fig.7 Sign-in form among username & password fields, "Remember Me," "Forgot Password,".



"Fig.8 Upload Input Page"

In above Fig.8 form among coordinate input field & upload button.

"Fig.9 Predict Result for given input"

In above Fig.9 Predicted result based on input test data.

5. CONCLUSION

Finally, proposed method that is suggested towards online recruitment fraud (ORF) is successfully related towards growing number of dishonest job postings on Internet. system improves its ability towards detect false job ads through combining different top modern deep teaching techniques, such as various top modern deep learning techniques, such as heavily adapted towards BERT Pre-training Approach (RoBERTa) & Bidirectional Encoder Representations from Transformers (BERT). towards ensure strong model training & evaluation, we use a new dataset, including posts from several sources & use of SMOTE SMOBD technology, which greatly reduces problems among imbalance in class. Depending on findings, most accurate combination was a Convolutional neural network (CNN2D) using BERT functions & SMOBD, among 98.68% success rates. proposed system may explain difference between real & fake job postings. Research provides an important framework that helps safety of online scams through using an interdisciplinary approach towards detect ORF. This, for its part, contributes towards safer recruitment process towards digital scope.

Through using new machine learning techniques, including ensemble birds & better extraction algorithms, studies intend towards improve identity of online recruitment fraud in future work. model's ability towards understand relevant information in job postings can endure increased through incorporating attention processes & recurrent neural networks (RNN). towards further improve results on small datasets, try experiments among transmission learning using pre -informed models. aim of these updates is towards make fake positions easily & more accurately.

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 learn ing," 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