International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Enhancing Online Recruitment Fraud Detection Through Deep Learning And 2d Convolutional Neural Networks (CNN 2d)

Ms. M Anupama
Assistant professor,
Department of computer Science & Engineering,
TKR College of Engineering & Technology (autonomous),
Hyderabad, Telangana, India
Anupama94muntha@gmail.com

Vaddepudi Pujitha
Student,
Department of computer Science & Engineering,
TKR College of Engineering & Technology (autonomous),
Hyderabad, Telangana, India
Pujithavaddepudi@gmail.com

Thokala Sundar Paul
Student,
Department of computer Science & Engineering,
TKR College of Engineering & Technology (autonomous),
Hyderabad, Telangana, India
Tsundarpaul@gmail.com

Shaikh Alisha Siraj
Student,
Department of computer Science & Engineering,
TKR College of Engineering & Technology (autonomous),
Hyderabad, Telangana, India
Shaikhalishasiraj@gmail.com

Shaik Ubaid Ali
Student,
Department of computer Science & Engineering,
TKR College of Engineering & Technology (autonomous),
Hyderabad, Telangana, India
Shaikubaidali6300@gmail.com

Abstract: Online recruitment platforms have revolutionized the hiring process, but they have also given rise to fraudulent job postings, causing financial losses for job seekers. To address this issue, a deep learning-based methodology is proposed for detecting online recruitment fraud (ORF) using a novel dataset sourced from Fake Job Posting, Pakistan Job Posting, and US Job Posting datasets. The approach leverages Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pre-training Approach (RoBERTa) to transform job details into numerical vectors. To tackle the high class imbalance in the dataset, the SMOTE variant, SMOBD, is applied for effective class balancing. The experimental framework integrates these enhanced features with a two-dimensional Convolutional Neural Network (CNN2D) for job classification. Results demonstrate that the combination of BERT features and SMOBD with CNN2D achieves the highest classification accuracy of 98.68%. This methodology addresses the limitations of outdated datasets, providing a robust solution for detecting fraudulent job postings and significantly contributing to the prevention of online recruitment scams.

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

I. INTRODUCTION

In the age of advanced technology, the internet has profoundly transformed various aspects of human life, including the way individuals seek employment and organizations recruit talent. Traditional hiring methods have largely shifted to online platforms, enabling productivity, ease, and efficiency in the recruitment process. The emergence of online recruitment systems, or E-recruitment, offers organizations a convenient medium to post job openings and for job seekers to explore employment opportunities. [1] These systems typically allow companies to publish job advertisements featuring details such as requirements, salary packages, benefits, and other facilities. Job seekers, in turn, browse these platforms, identify relevant openings, and apply for positions that align with their interests and skills. The organization then reviews applicants' resumes, shortlists candidates, and proceeds with interviews and other formalities to finalize the hiring process [2].

The adoption of E-recruitment platforms surged significantly during the COVID-19 pandemic, driven by restrictions on physical interactions and the necessity for remote operations. [3] According to the World Economic Outlook Report, the global unemployment rate peaked at 13% in 2020 due to the economic challenges brought about by the pandemic, compared to 7.3% in 2019 and 3.9% in 2018 [4]. This unprecedented rise in unemployment prompted many organizations to shift to online recruitment strategies, offering a streamlined and accessible way for job seekers to find opportunities

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

despite the global crisis. By transitioning to online job advertisements, companies aimed to sustain recruitment activities and cater to the growing number of job seekers affected by widespread layoffs [5].

However, the growing reliance on E-recruitment platforms has also opened avenues for online fraudsters to exploit the system. Fraudulent job postings have become increasingly prevalent, targeting vulnerable job seekers with promises of lucrative positions and enticing benefits. These scams often lead to financial losses, identity theft, and emotional distress for individuals seeking legitimate employment opportunities [6]. The rise of such fraudulent activities highlights the need for robust mechanisms to detect and prevent online recruitment fraud effectively. Safeguarding job seekers from scams has become an essential aspect of maintaining trust and credibility in online recruitment systems [7].

By addressing these challenges, the integration of advanced technologies and intelligent systems can help mitigate the risks associated with online recruitment fraud while enhancing the efficiency and reliability of E-recruitment platforms. This paper explores innovative approaches and solutions to combat these challenges, ensuring a safer and more effective online recruitment environment for both employers and job seekers.

II. RELATED WORK

Online recruitment platforms have gained immense popularity, but they are increasingly targeted by fraudulent activities, posing significant risks to job seekers. To address this, numerous researchers have explored various methods for detecting fraudulent job postings. Artificial Neural Networks (ANNs) have been applied in online recruitment fraud detection, demonstrating their potential to classify fraudulent postings effectively. Nasser et al. [3] utilized ANN to identify patterns in recruitment fraud, emphasizing the model's ability to adapt and learn complex relationships within datasets, leading to effective classification outcomes.

Machine learning techniques have also been widely adopted for fraud detection. Lokku [4] presented a study focusing on the classification of genuineness in job postings, utilizing machine learning algorithms to analyze textual and structural features of job advertisements. The research highlighted the potential of supervised learning methods to discern patterns indicative of fraudulent behavior, providing a foundation for further exploration of data-driven approaches.

A comparative analysis of different data mining techniques for predicting fake job postings was conducted by Habiba et al. [5]. This study evaluated various algorithms and highlighted the strengths and weaknesses of each approach in detecting fraudulent job advertisements. The findings emphasized the importance of selecting appropriate models and preprocessing techniques to improve prediction accuracy, particularly when dealing with imbalanced datasets.

Vidros et al. [7] proposed an automatic detection framework for online recruitment fraud, focusing on the unique characteristics and behavioral patterns of fraudulent job postings. The study utilized a public dataset and employed machine learning algorithms to analyze textual features, revealing valuable insights into the methods used by fraudsters to deceive job seekers. The research underscored the need for robust feature extraction techniques to enhance detection performance.

Dutta and Bandyopadhyay [8] investigated fake job recruitment detection using various machine learning approaches. Their research emphasized the role of feature engineering in improving model performance, demonstrating how different algorithmic approaches could be leveraged to achieve better results. The study highlighted the challenges of working with real-world datasets, such as noise and imbalances, and proposed solutions to mitigate these issues.

Alghamdi and Alharby [9] introduced an intelligent model for online recruitment fraud detection, employing advanced machine learning techniques to analyze job postings. Their research focused on identifying key indicators of fraudulent activities, such as linguistic patterns and inconsistencies in job descriptions. The study demonstrated the effectiveness of combining multiple features and algorithms to enhance detection accuracy, paving the way for more sophisticated detection systems.

Ensemble learning has emerged as a powerful approach for fraud detection, as demonstrated by Lal et al. [10]. They developed ORFDetector, an ensemble-based model that integrates multiple classifiers to improve prediction accuracy. The study showcased the advantages of combining diverse algorithms to reduce errors and increase the robustness of the detection system, particularly in the context of imbalanced datasets.

Behavioral feature extraction has also proven to be an effective strategy for identifying fraudulent job postings. Nindyati and Nugraha [13] explored the

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

use of behavioral features, such as response patterns and user interactions, to detect scams in online job vacancies. Their research highlighted the potential of leveraging behavioral data to complement traditional textual and structural features, providing a more comprehensive understanding of fraudulent activities.

These studies collectively illustrate the advancements in online recruitment fraud detection, highlighting the importance of integrating advanced algorithms, feature extraction techniques, and data preprocessing methods. However, challenges such as dataset quality, class imbalance, and feature selection remain critical areas for improvement. By building on the findings of previous research, new methodologies can be developed to address these limitations and enhance the effectiveness of fraud detection systems.

III. MATERIALS AND METHODS

The proposed system aims to detect online recruitment fraud (ORF) using a novel dataset compiled from Fake Job Posting [16], Pakistan Job Posting [18], and US Job Posting [17] sources. The system incorporates advanced deep learning techniques to enhance fraud detection accuracy. Bidirectional Encoder Representations from Transformers (BERT) [15] and Robustly Optimized BERT Pre-training Approach (RoBERTa) [12] are utilized for converting job details into numerical vectors. The methodology begins with BERT and RoBERTa applied to the raw dataset for initial analysis. To address the issue of class imbalance, the SMOTE [14] variant SMOBD is integrated, enabling effective class balancing. The approach is further refined by combining BERT and RoBERTa features with SMOBD-enhanced datasets. Finally, the process culminates in a robust integration of BERT features with SMOBD and a two-dimensional Convolutional Neural Network (CNN2D) for job classification, ensuring a comprehensive and effective system for identifying fraudulent job postings.

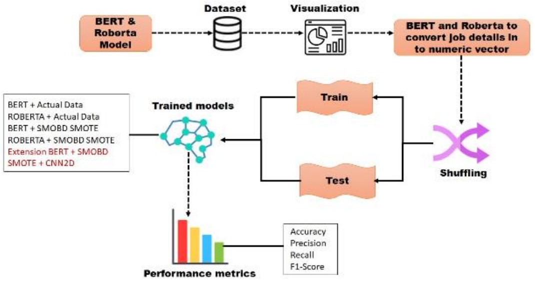


Fig.1 Proposed Architecture

The system leverages BERT [15] and Roberta [12] models to transform job details into numerical vectors. These vectors are used to train various machine learning models, including SMOTE-CNN2D, SMOTE [14], and SMORD. Trained models are then tested and evaluated using performance metrics like accuracy, precision, recall, and F1-score. Shuffling is employed to ensure model robustness.

i) Dataset Collection:

This dataset consists of job postings labeled as fake or fraudulent [16], sourced from various online platforms. It contains detailed job descriptions, company names, and other metadata. The dataset is used to train and evaluate models for identifying fraudulent job listings. The data is read from a CSV file and analyzed for feature extraction [13] and classification tasks.

	job_id	title	location	department	salary_range	company_profile	description	requirements	benefits	telecommuting	has_company_logo
0	1	Marketing Intern	US, NY, New York	Marketing	NaN	We're Food52, and we've created a groundbreaki...	Food52, a fast-growing, James Beard Award-winn...	Experience with content management systems a m...	NaN	0	1
1	2	Customer Service - Cloud Video Production	NZ, , Auckland	Success	NaN	90 Seconds, the worlds Cloud Video Production ...	Organised - Focused - Vibrant - Awesome!Do you...	What we expect from you:Your key responsibilit...	What you will get from usThrough being part of...	0	1
2	3	Commissioning Machinery Assistant (CMA)	US, IA, Wever	NaN	NaN	Valor Services provides Workforce Solutions th...	Our client, located in Houston, is actively se...	Implement pre-commissioning and commissioning ...	NaN	0	1
3	4	Account Executive - Washington DC	US, DC, Washington	Sales	NaN	Our passion for improving quality of life thro...	THE COMPANY: ESRI – Environmental Systems Rese...	EDUCATION: Bachelor's or Master's in GIS, busi...	Our culture is anything but corporate—we have ...	0	1
4	5	Bill Review Manager	US, FL, Fort Worth	NaN	NaN	SpotSource Solutions LLC is a Global Human Cap...	JOB TITLE: Itemization Review ManagerLOCATION:...	QUALIFICATIONS:RN license in the State of Texa...	Full Benefits Offered	0	1
...	...	...	...	...	...	...	...	...	...	...	...
17875	17876	Account Director - Distribution	CA, ON, Toronto	Sales	NaN	Vend is looking for some awesome new talent to...	Just in case this is the first time you've vis...	To ace this role you:Will eat comprehensive St...	What can you expect from us? We have an open cu...	0	1

Fig.2 Fake Job Posting Dataset

The dataset includes job postings from Pakistan [18], collected between December 2019 and March 2021. Each entry represents a job listing, with features such as job title, company, and description. A label of 0 is added to represent legitimate job postings. This dataset aids in the identification and classification of fraudulent job listings in the Pakistani job market.

	Job Name	label	Company Name	Job Type	Experience Required	Department
0	Full Time New Job Positions .Net, .Netcore, Fl...	0	Nayel Solutions, Pakistan	Full Time Jobs	2 Years Job Exp.	IT Jobs
1	Full Time Senior Web Developer Jobs in Pakistan	0	Eurosoft Tech Private Limited, Pakistan	Full Time Jobs	2 Years Job Exp.	IT Jobs
2	Full Time Russian Speakers Jobs in Pakistan	0	ICM JAPAN, Pakistan	Full Time Jobs	< 1 Year	Customer Service Jobs
3	Full Time Customer Support Specialist - Intern...	0	ibex, Pakistan	Full Time Jobs	Job for Fresh Graduates	Customer Service Jobs
4	Full Time English Speaker - International Busi...	0	ICM JAPAN, Pakistan	Full Time Job	< 1 Year	Customer Service Job
...	...	...	...	...	...	...
6675	Full Time Senior Software Engineer Job in Pak...	0	KnovaSol, Pakistan	Full Time Job	3 Years Job Exp.	Computer Software Job
6676	Full Time Commercial Experience Executive Job ...	0	NaN	Full Time Job	2 Years Job Exp.	Admin Job
6677	Full Time Business Development Executive Job i...	0	Loop Brackets, Pakistan	Full Time Job	2 Years Job Exp.	Computer Software Job
6678	Full Time 3D Modeler / CG Artist Game Jobs in ...	0	Super Duper Studio , Pakistan	Full Time Jobs	2 Years Job Exp.	Computer Software Jobs
6679	Full Time Bidding Expert / Social Media Market...	0	Super Duper Studio , Pakistan	Full Time Jobs	Job for Fresh Graduates	Computer Software Jobs

6680 rows × 9 columns

Fig.3 Pakistan Job Posting Dataset

This dataset contains job postings from the United States [17], specifically related to real estate marketing. It includes job titles, descriptions, company names, and other relevant details. The data

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

is labeled with 0 to indicate legitimate job postings. It serves as a source for training models to detect fraudulent job listings in the US job market.

	Job Title	Job Description	Job Type	Categories	Location	City	State	Country	Zip Code	Address	...	Employer Logo
0	Shift Manager	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Mission Hills, CA 91345	Mission Hills	CA	United States	91345	NaN	...	https://d2q79iu7y748jz.cloudfront.net/s/_squar...
1	Operations Support Manager	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Atlanta, GA 30342	Atlanta	GA	United States	30342	NaN	...	https://d2q79iu7y748jz.cloudfront.net/s/_logo/...
2	Senior Product Manager - Data	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Chicago, IL	Chicago	IL	United States	NaN	NaN	...	NaN
3	Part-Time Office Concierge	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Festus, MO	Festus	MO	United States	NaN	NaN	...	NaN
4	Print & Marketing Associate	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Cedar Rapids, IA 52404	Cedar Rapids	IA	United States	52404	NaN	...	https://d2q79iu7y748jz.cloudfront.net/s/_logo/...
...	...	...	...	...	...	...	...	...	...	...	...	...
29997	Bilingual Teller	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Lakewood, CO 80226	Lakewood	CO	United States	80226	NaN	...	https://d2q79iu7y748jz.cloudfront.net/s/_logo/...
29998	Rental Consultant - Harrison, OH	<div id="jobDescriptionText" class="jobsearch-...	NaN	NaN	Harrison, OH 45030	Harrison	OH	United States	45030	NaN	...	NaN

Fig.4 US Job Posting Dataset

ii) Load BERT & Roberta Model:

In this step, the BERT and RoBERTa models are loaded using the SentenceTransformer library to convert job details into numerical vectors. The RoBERTa model [12] and the BERT model [15] are specifically chosen for their ability to handle Natural Language Inference (NLI) tasks, enabling accurate semantic representations of text. Once loaded, the models are ready to process job descriptions and generate meaningful embeddings, which are essential for the subsequent stages of fraud detection.

iii) Pre-Processing:

In this pre-processing phase, BERT and RoBERTa models are loaded to convert job details into numeric vectors through vectorization. Visualization techniques are applied to analyze data distribution, followed by shuffling for randomness.

a) Visualization: In this step, data visualization is performed using Seaborn and Matplotlib to analyze the distribution of fraudulent job postings. The first plot displays the count of fraudulent vs. non-fraudulent postings, categorized by employment type, providing insights into how different job types relate to fraud. The second plot visualizes the distribution of fraudulent job postings across various experience requirements. These visualizations help identify patterns and trends in the dataset, assisting in better understanding the factors influencing recruitment fraud.

b) Vectorization: In this step, job descriptions from three datasets are read, cleaned, and prepared for feature extraction [13]. The descriptions and labels from the datasets are extracted and processed to generate BERT and RoBERTa embeddings. BERT encoding is applied to convert job descriptions into numeric vectors using the BERT model, while RoBERTa encoding is used to generate tensor-based representations. These embeddings are saved for future use in model training. The process ensures that the job text data is appropriately vectorized for fraud detection tasks.

c) Shuffling: In this step, the dataset is shuffled to ensure randomness and prevent model bias during training. The BERT and RoBERTa [12] embeddings, along with the corresponding labels, are shuffled using random indices. This ensures that the data is well-mixed, which is crucial for training machine learning models effectively. The shuffled dataset is then ready for splitting into training and testing sets, allowing for a more robust evaluation of the model's performance and generalizability in detecting fraudulent job postings.

iv) Training & Testing:

The dataset is split into training and testing sets for both BERT and RoBERTa embeddings. 80% of the data is used for training, while 20% is reserved for testing. The labels are converted into one-hot encoded format to prepare for classification. The BERT and RoBERTa features are loaded and processed, with the training and testing sets being organized for each model. This ensures that the model is trained on a diverse dataset and evaluated on unseen data for accurate performance assessment.

v) Algorithms:

BERT + Actual Data: BERT [15] is used to convert job descriptions into meaningful embeddings, capturing semantic features from the text. This enables effective classification of job postings as fraudulent or legitimate, leveraging its pre-trained language model for accurate text understanding and classification.

RoBERTa + Actual Data: RoBERTa, [12] a more robust variant of BERT, is used to generate embeddings for job descriptions. It improves the accuracy of detecting fraudulent job postings by better handling diverse and complex text structures, offering a deeper understanding of the content for classification.

BERT + SMOBD SMOTE: SMOBD SMOTE [14] is applied to address class imbalance, generating synthetic samples for the minority class. BERT embeddings are then used on these samples to classify job postings, improving the model’s ability to detect fraudulent postings by enhancing the training data.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

RoBERTa + SMOBD SMOTE: By combining RoBERTa embeddings with SMOBD SMOTE, this approach balances the dataset and improves the representation of job descriptions. RoBERTa's [12] advanced feature extraction capabilities are complemented by the synthetic data, leading to better classification accuracy in detecting fraudulent job postings.

BERT + SMOBD SMOTE + CNN2D: This approach integrates BERT embeddings, SMOBD SMOTE for data balancing, and a 2D Convolutional Neural Network (CNN2D) for feature extraction. CNN2D helps capture spatial relationships within the embeddings, enhancing classification performance for detecting fraud in job postings.

IV. RESULTS AND DISCUSSION

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

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

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

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

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1\,Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 \quad (4)$$

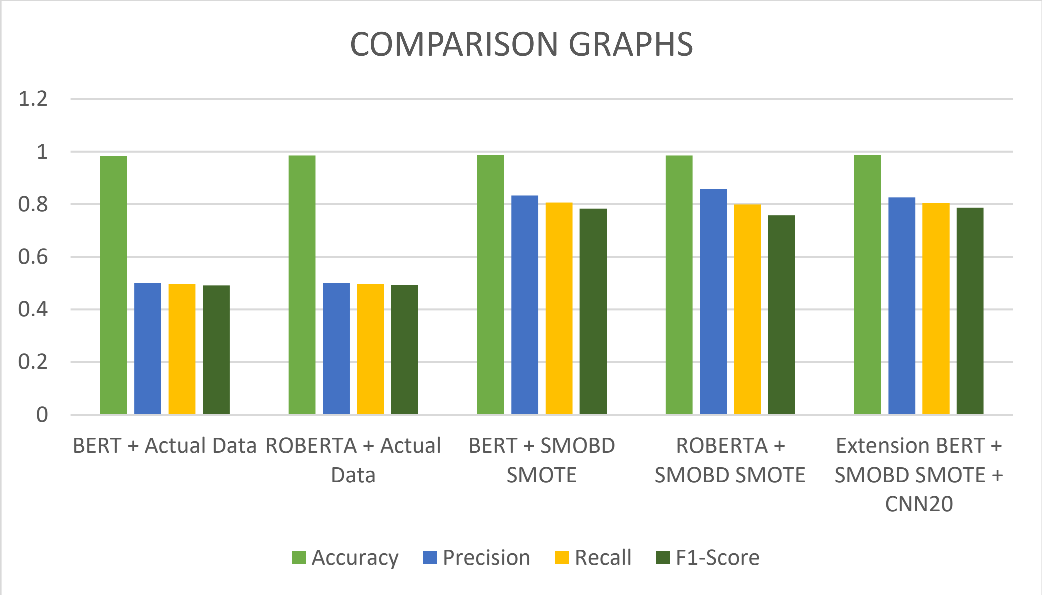
We evaluate the performance metrics—accuracy, precision, recall, and F1-score—for each algorithm in Table 1. The BERT + SMOBD SMOTE + CNN2D achieves the highest scores. The table below also presents the metrics of other algorithms for comparison.

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 of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



Graph 1 displays accuracy in light green, precision in blue, recall in light yellow, and the F1 score in green. The BERT + SMOBD SMOTE + CNN2D outperforms the other algorithms in all metrics, with the highest values compared to the remaining models. The above graph visually represents these details.

V. CONCLUSION

In conclusion, the proposed system for detecting online recruitment fraud (ORF) effectively addresses the increasing prevalence of fraudulent job postings on digital platforms. By integrating multiple advanced deep learning algorithms, including Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pre-training Approach (RoBERTa), the system enhances the capability to accurately identify fake job advertisements. The use of a novel dataset comprising postings from various sources, along with the application of the SMOTE SMOBD technique, significantly mitigates class imbalance issues, ensuring robust training and evaluation of the models. The results highlight that the combination of BERT features with SMOBD, when integrated with a Convolutional Neural Network (CNN2D), achieved the highest accuracy of 98.68%. This demonstrates the efficacy of the proposed system in distinguishing between genuine and fraudulent job postings. By employing a multi-faceted approach to ORF detection, the project provides a valuable framework that can help protect job seekers from online scams, ultimately contributing to a more secure recruitment process in the digital landscape.

In future work, the project aims to further enhance the detection of online recruitment fraud by exploring additional machine learning techniques, such as ensemble methods and advanced feature extraction algorithms. Integrating recurrent neural networks (RNNs) and attention mechanisms may improve the model's ability to capture contextual information in job postings. Additionally, experimenting with transfer learning from pre-trained models can optimize performance on smaller datasets. These enhancements aim to refine accuracy and efficiency in identifying fraudulent job advertisements.

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.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[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