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

Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches

Mrs. Mynam Sushma Devi (M.Tech)
Assistant Professor
Department of Information Technology
JBIET Autonomous TG,
Moinabad, India
sushma.m5809@gmail.com

Nagaram Akshitha Goud
Student,
Department of Information Technology
JBIET Autonomous TG,
Moinabad, India
nagaramakshitha29@gmail.com

Mr.Vijay Prasad Tiwari
Student,
Department of Information Technology,
JBIET Autonomous TG,
Moinabad, India
vt5587038@gmail.com

Chawan Rahul
Student,
Department of Information technology
JBIET Autonomous TG,
Moinabad, India
rahulchawan494@gmail.com

Sathugari Shravya Reddy
Student,
Department of Information Technology
JBIET Autonomous TG,
Moinabad, India
shravyareddy1011@gmail.com

Abstract: Online recruiting platforms have transformed the employment process; yet, they have also led to the emergence of bogus job listings, resulting in financial losses for job searchers. A deep learning algorithm is suggested to identify online recruiting fraud (ORF) with a fresh dataset derived from Fake Job Posting, Pakistan Job Posting, and US Job Posting datasets. The method utilizes Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pre-training Approach (RoBERTa) to convert task specifications into numerical vectors. To address the significant class imbalance in the dataset, the SMOTE variation, SMOBD, is utilized for efficient class balancing. The experimental framework incorporates these advanced characteristics with a two-dimensional Convolutional Neural Network (CNN2D) for job categorization. The results indicate that the integration of BERT features and SMOBD with CNN2D attains the greatest classification accuracy of 98.68%. This technique mitigates the shortcomings of obsolete datasets, offering a comprehensive solution for identifying fake job listings and substantially aiding in the avoidance of online recruiting frauds.

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

I. INTRODUCTION

The internet has significantly changed many facets of human existence in the era of modern technology, including how people look for work and how businesses find talent. Online platforms have essentially replaced traditional hiring techniques, making the employment process more productive, convenient, and efficient. The advent of internet recruiting platforms, or E-recruitment, provides employers with a practical way to advertise job vacancies and job searchers with a way to investigate career prospects. [1] Usually, these systems enable employers to post job postings with information about prerequisites, compensation ranges, perks, and other amenities. In response, job searchers peruse these sites, find employment that fit their interests and skill set, and apply. After reviewing resumes, the company shortlists individuals, conducts interviews, and completes additional procedures to complete the recruiting process [2].

Due to limitations on in-person encounters and the need for distant operations, the use of e-recruitment platforms increased dramatically during the COVID-19 epidemic. [3] The World Economic Outlook Report states that the pandemic's economic effects caused the worldwide unemployment rate to peak at 13% in 2020, up from 7.3% in 2019 and 3.9% in 2018 [4]. Many businesses switched to online recruitment tactics as a result of the extraordinary increase in unemployment, providing job searchers with an easy and convenient approach to locate possibilities despite the global crisis.

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

Companies sought to maintain recruitment efforts and serve the increasing number of job seekers impacted by mass layoffs by switching to online job postings [5].

But the increased use of e-recruitment platforms has also made it easier for online scammers to take advantage of the system. With promises of rich opportunities and alluring incentives, fraudulent job advertising have become more common, preying on unsuspecting job searchers. For those looking for real jobs, these frauds frequently result in monetary losses, identity theft, and psychological suffering [6]. The increase in these fraudulent operations emphasizes the necessity of strong systems to successfully identify and stop online recruiting fraud. In order to preserve credibility and confidence in online recruiting processes, protecting job seekers against frauds has become crucial [7].

The incorporation of cutting-edge technology and intelligent systems can assist reduce the risks related to online recruitment fraud while improving the effectiveness and dependability of e-recruitment platforms by solving these issues. In order to provide a safer and more productive online recruiting environment for both companies and job seekers, this article investigates creative ways and answers to these problems.

II. RELATED WORK

Although online recruiting platforms have become quite popular, they are increasingly being targeted by fraudulent activity, which puts job searchers at serious danger. Many academics have investigated different techniques for identifying bogus job advertisements in order to remedy this. Artificial Neural Networks (ANNs) have shown promise in accurately classifying fraudulent postings when used in online recruiting fraud detection. Nasser et al. [3] highlighted the model's capacity to adapt and understand intricate linkages within datasets, resulting in successful classification outputs, by using ANN to find patterns in recruiting fraud.

Fraud detection has also made extensive use of machine learning techniques. In a research on the categorization of authenticity in job advertising, Lokku [4] used machine learning algorithms to examine the textual and structural aspects of job ads. The study demonstrated how supervised learning techniques may identify trends that point to fraudulent activity, laying the groundwork for future research into data-driven strategies.

Habiba et al. [5] performed a comparative study of several data mining methods for identifying fraudulent job advertisements. This study assessed many algorithms and outlined the advantages and disadvantages of each method for identifying bogus job postings. The results highlighted how crucial it is to use the right models and preprocessing methods in order to increase prediction accuracy, especially when working with unbalanced datasets.

An automated detection method for online recruiting fraud was presented by Vidros et al. [7], who concentrated on the distinctive traits and behavioral patterns of fraudulent job posts. The study used machine learning algorithms to examine textual elements from a publicly available dataset, providing important insights into how fraudsters trick job searchers. In order to improve detection performance, the study emphasized the necessity of strong feature extraction methods.

Dutta and Bandyopadhyay [8] used a variety of machine learning techniques to study the identification of fraudulent job recruiting. Their study highlighted the importance of feature engineering in enhancing model performance and showed how various algorithmic techniques may be used to improve outcomes. The paper outlined the difficulties of dealing with real-world datasets, including imbalances and noise, and suggested ways to lessen these problems.

Alghamdi and Alharby [9] developed an effective model for detecting online recruiting fraud by analyzing job listings using cutting-edge machine learning techniques. Their study concentrated on finding important markers of fraudulent activity, such language trends and inconsistent job descriptions. The work opened the door for more advanced detection systems by demonstrating how integrating many characteristics and algorithms may improve detection accuracy.

According to Lal et al. [10], ensemble learning has become a potent method for detecting fraud. To increase prediction accuracy, they created ORFDetector, an ensemble-based approach that combines many classifiers. The study demonstrated the benefits of integrating several methods to lower mistakes and boost the detection system's resilience, especially when dealing with unbalanced datasets.

Another successful method for spotting bogus job advertisements is behavioral feature extraction. In order to identify frauds in online job postings, Nindyati and Nugraha [13] investigated the use of behavioral variables including reaction patterns and

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

user interactions. Their study demonstrated how behavioral data may be used to supplement conventional textual and structural characteristics, offering a more thorough comprehension of fraudulent activity.

Together, these studies show how online recruiting fraud detection has evolved, emphasizing the significance of combining sophisticated algorithms, feature extraction strategies, and data pretreatment approaches. However, issues like feature selection, class imbalance, and dataset quality continue to be crucial areas for development. New approaches to overcome these constraints and improve the efficacy of fraud detection systems can be created by expanding on the results of earlier studies.

III. MATERIALS AND METHODS

Using a unique dataset gathered from sources such as Fake Job Posting [16], Pakistan Job Posting [18], and US Job Posting [17], the proposed approach seeks to identify online recruiting fraud (ORF). To improve the accuracy of fraud detection, the system uses cutting-edge deep learning algorithms. Job information are transformed into numerical vectors using Bidirectional Encoder Representations from Transformers (BERT) [15] and Robustly Optimized BERT Pre-training Approach (RoBERTa) [12]. The first step in the process is to apply BERT and RoBERTa to the raw dataset for preliminary analysis. Effective class balancing is made possible by the integration of the SMOTE [14] variation SMOBD, which addresses the problem of class imbalance. By merging SMOBD-enhanced datasets with BERT and RoBERTa characteristics, the method is further improved. In order to ensure a thorough and efficient system for detecting fraudulent job posts, the approach culminates in a strong integration of BERT characteristics with SMOBD and a two-dimensional Convolutional Neural Network (CNN2D) for job categorization.

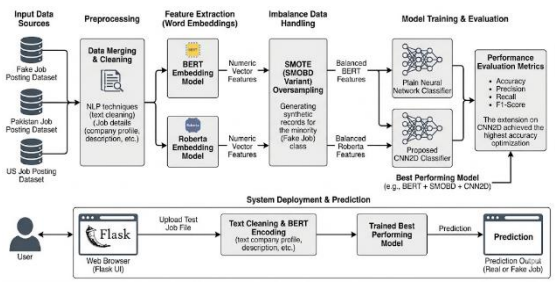
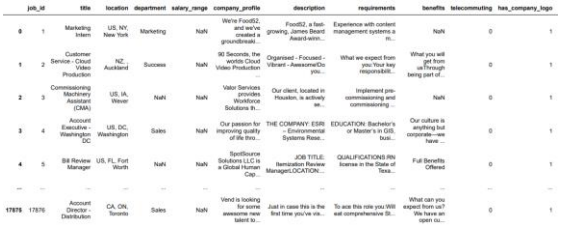


Fig.1 Proposed Architecture

The system converts work information into numerical vectors by utilizing the BERT [15] and Roberta [12] models. Several machine learning models, such as SMOTE-CNN2D, SMOTE [14], and SMORD, are trained using these vectors. Performance indicators including as accuracy, precision, recall, and F1-score are then used to test and assess trained models. To guarantee the resilience of the model, shuffling is used.

i) Dataset Collection:

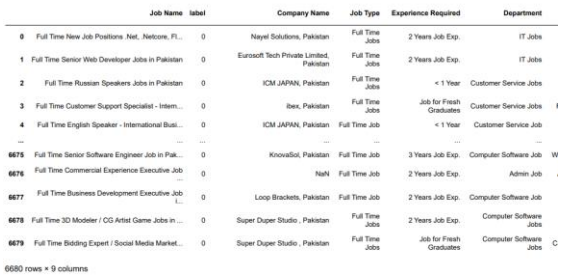
This collection is made up of job advertisements from several online sites that have been classified as fraudulent or phony [16]. It includes corporate names, comprehensive job descriptions, and more metadata. Models for detecting fraudulent job postings are trained and assessed using the dataset. For feature extraction [13] and classification tasks, the data is extracted from a CSV file and examined.



	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 groundbreak...	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 - AwesomeDo 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 eal comprehensive St...	What can you expect from us? We have an open cu...	0	1

Fig.2 Fake Job Posting Dataset

Job advertisements from Pakistan [18] that were gathered between December 2019 and March 2021 are included in the dataset. A job listing with details like the job title, firm, and description is represented by each item. To indicate that a job advertising is authentic, a label of 0 is given. The Pakistani job market's bogus job postings may be identified and categorized with the use of this dataset.



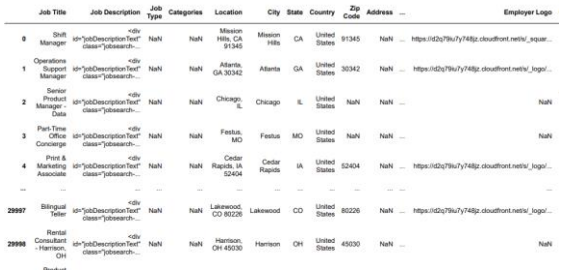
	Job Name	label	Company Name	Job Type	Experience Required	Department	
0	Full Time New Job Positions .Net, .Netcore, Fl...	0	Nayel Solutions, Pakistan	Full Time Jobs	2 Years Job Exp.	IT Jobs	
1	Full Time Senior Web Developer Jobs in Pakistan	0	Eurosoft Tech Private Limited, Pakistan	Full Time Jobs	2 Years 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	I
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	W
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 (...	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	C

6680 rows × 9 columns

Fig.3 Pakistan Job Posting Dataset

Real estate marketing-related job posts from the United States are included in this collection [17]. It contains corporate names, job titles, descriptions, and other pertinent information. To denote valid job ads, the data is tagged with 0. It is used to train algorithms to identify phony job postings in the US employment market.

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



	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 order to transform task information into numerical vectors, the SentenceTransformer library is used to import the BERT and RoBERTa models. Because they can perform Natural Language Inference (NLI) tasks and provide reliable semantic representations of text, the RoBERTa model [12] and the BERT model [15] were specially selected. After being loaded, the models may evaluate job descriptions and provide insightful embeddings, which are crucial for the fraud detection phases that follow.

iii) Pre-Processing:

BERT and RoBERTa models are imported at this pre-processing stage in order to vectorize work details into numerical vectors. Data distribution is examined using visualization tools, and then shuffling is used to check for randomness.

a) Visualization: In this stage, the distribution of fraudulent job listings is analyzed using data visualization using Matplotlib and Seaborn. In order to shed light on the relationship between various job categories and fraud, the first figure shows the number of fraudulent vs non-fraudulent posts, broken down by employment type. The distribution of fraudulent job advertisements across different experience criteria is seen in the second plot. By highlighting patterns and trends in the dataset, these visualizations aid in a deeper comprehension of the variables affecting recruiting fraud.

b) Vectorization: This stage involves reading, cleaning, and getting ready to extract features from job descriptions from three datasets [13]. BERT and RoBERTa embeddings are created by extracting and processing the labels and descriptions from the datasets. While RoBERTa encoding is used to create tensor-based representations, BERT encoding is used to transform job descriptions into numeric vectors using the BERT paradigm. These embeddings are stored for further model training. For fraud detection jobs, the procedure guarantees that the job text data is suitably vectorized.

c) Shuffling: To guarantee randomization and avoid model bias during training, the dataset is shuffled in this stage. Random indices are used to shuffle the related labels and the BERT and RoBERTa [12] embeddings. This guarantees that the data is well-mixed, which is essential for efficiently training machine learning models. The model's effectiveness and generalizability in identifying fraudulent job ads may then be more thoroughly assessed by dividing the shuffled dataset into training and testing sets.

iv) Training & Testing:

For both BERT and RoBERTa embeddings, the dataset is divided into training and testing sets. Twenty percent of the data is set aside for testing, while the remaining eighty percent is used for training. To get ready for categorization, the labels are transformed into a one-hot encoded format. Each model's training and testing sets are arranged once the BERT and RoBERTa features are loaded and processed. In order to accurately measure the model's performance, this guarantees that it is trained on a variety of datasets and assessed on previously unreported data.

v) Algorithms:

BERT + Actual Data: By extracting semantic information from the text, BERT [15] transforms job descriptions into meaningful embeddings. By using its pre-trained language model for precise text recognition and classification, this makes it possible to classify job advertisements as real or fraudulent.

RoBERTa + Actual Data: To create embeddings for job descriptions, a more reliable version of BERT called RoBERTa [12] is employed. By handling complicated and varied text structures more effectively and providing a deeper comprehension of the information for categorization, it increases the accuracy of identifying fake job posts.

BERT + SMOBD SMOTE: Class imbalance is addressed by using SMOBD SMOTE [14], which creates synthetic samples for the minority class. The model's capacity to identify fraudulent postings is then improved by using BERT embeddings on these samples to categorize job listings.

RoBERTa + SMOBD SMOTE: This method balances the dataset and enhances the job description representation by merging RoBERTa

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

embeddings with SMOBD SMOTE. The simulated data enhances RoBERTa's [12] sophisticated feature extraction capabilities, improving classification accuracy in identifying fake job listings.

BERT + SMOBD SMOTE + CNN2D: This method combines a 2D Convolutional Neural Network (CNN2D) for feature extraction, SMOBD SMOTE for data balancing, and BERT embeddings. CNN2D improves classification performance for identifying fraud in job advertisements by capturing spatial links within the embeddings.

IV. RESULTS AND DISCUSSION

Accuracy: The capacity of a test to accurately distinguish between healthy instances and patients is known as accuracy. The percentage of true positive and true negative in each analyzed instance should be calculated in order to measure a test's accuracy. This may be expressed mathematically as:

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

Precision: Precision measures the percentage of samples or incidents that are accurately identified as positives. Therefore, the following formula may be used to determine the precision:

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

Recall: In ML, recall is a statistic that assesses a model's capacity to find every pertinent instance of a given class. It provides information about how well a model captures instances of a certain class and is calculated as the ratio of properly predicted positive observations to all actual positives.

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

F1-Score: The accuracy of a model is assessed using the F1 score, a machine learning assessment metric. It integrates a model's recall and accuracy scores. The accuracy measure calculates the number of times a model correctly predicted the whole dataset.

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

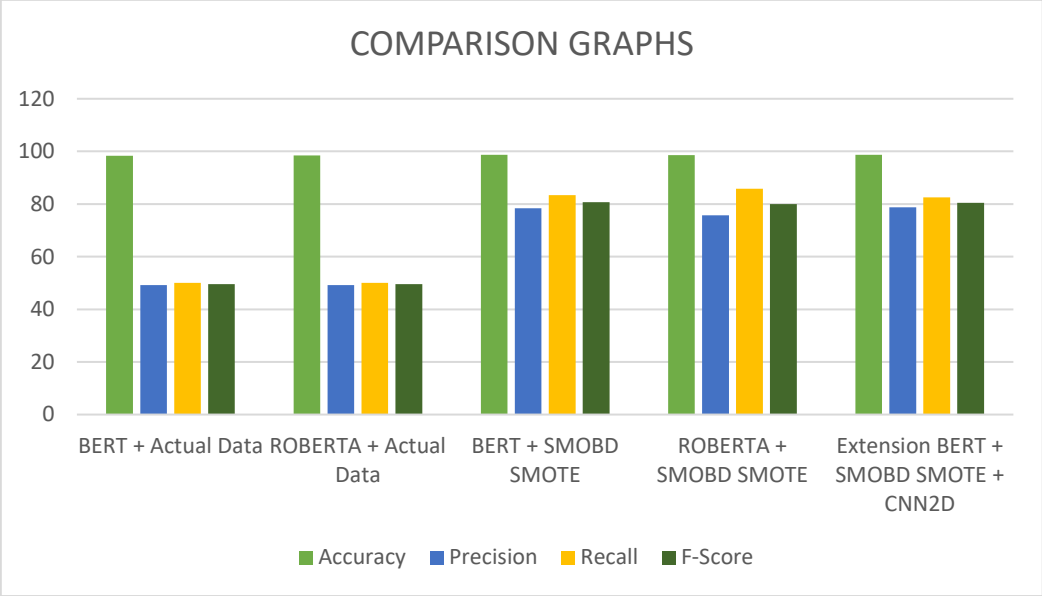
For every method in Table 1, we assess the performance parameters of accuracy, precision, recall, and F1-score. The best results are obtained by BERT + SMOBD SMOTE + CNN2D. The metrics of various algorithms are also shown in the table below for comparison.

Table.1 Performance Evaluation Metrics

Algorithm Name	Accuracy	Precision	Recall	F-Score
BERT + Actual Data	98.387245	49.193622	50.000000	49.593533
ROBERTA + Actual Data	98.497205	49.248603	50.000000	49.621457
BERT + SMOBD SMOTE	98.662146	78.339538	83.392363	80.649130
ROBERTA + SMOBD SMOTE	98.579676	75.771937	85.768812	79.918306
Extension BERT + SMOBD SMOTE + CNN2D	**98.680473**	**78.719966**	**82.563375**	**80.515399**

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 shows the F1 score in green, recall in light yellow, accuracy in light green, and precision in blue. With the highest values when compared to the other models, the BERT + SMOBD SMOTE + CNN2D algorithm performs better than the others in every metric. These facts are shown graphically in the graph above.

V. CONCLUSION

In conclusion, the growing frequency of fraudulent job advertisements on digital platforms is successfully addressed by the suggested approach for identifying online recruitment fraud (ORF). The system improves the capacity to detect fraudulent job adverts by combining many sophisticated deep learning techniques, such as Robustly Optimized BERT Pre-training Approach (RoBERTa) and Bidirectional Encoder Representations from Transformers (BERT). Class imbalance problems are greatly reduced by using a new dataset that includes postings from several sources and applying the SMOTE SMOBD approach, guaranteeing reliable model training and assessment. The findings show that the maximum accuracy of 98.68% was attained when BERT features and SMOBD were combined with a Convolutional Neural Network (CNN2D). This illustrates how well the suggested approach can differentiate between legitimate and fake job advertisements. The study offers a useful framework that helps shield job seekers against online frauds by using a multifaceted approach to ORF detection, thus leading to a more safe hiring process in the digital realm.

By investigating new machine learning approaches, such as ensemble methods and sophisticated feature extraction algorithms, the group hopes to improve the identification of online recruiting fraud in future work. The model's capacity to extract contextual information from job advertisements may be enhanced by including recurrent neural networks (RNNs) and attention processes. Additionally, performance on smaller datasets may be optimized by experimenting with transfer learning from pre-trained models. These improvements are meant to improve the precision and effectiveness of spotting bogus job postings.

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.

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

[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