

A BERT-Based Multilingual Framework for Detecting Online Recruitment Scams in the Bangladesh Job Market

Tarannum Binte Tariq^{1*}, Saiful Islam¹

Abstract

Online recruitment scams pose a growing threat in digital job markets, particularly in emerging economies such as Bangladesh, where socioeconomic conditions create a highly vulnerable population. With a youth unemployment rate of 7.2%, a gender skewed victimization pattern where 59% of the victims are female and widespread adoption of mobile financial services such as bKash and Nagad, Bangladesh presents a uniquely challenging environment for fraud detection. Existing machine learning models for scam detection have been predominantly trained on Western datasets and therefore fail to capture fraud signals specific to emerging markets. This paper proposes a localized detection framework based on an Enhanced Multilingual BERT architecture with custom regularization, trained on a dataset of 20,000 annotated job postings collected from major Bangladeshi recruitment platforms like BDJobs and LinkedIn as well as informal channels such as Telegram and WhatsApp groups over a 21-month period. The dataset incorporates Bangladesh-specific fraud signals, including references to mobile financial services, informal communication patterns and Bangla-English code-mixing. The proposed model achieves a test accuracy of 95.6% with precision of 94.0% and 89.0% for legitimate and fraudulent postings respectively and F1-scores of 95.0% and 87.0% for the same classes. The framework outperforms existing global models on Bangladesh-specific evaluation data by 3.3 to 8.1 percentage points.

Keywords: Online Recruitment Fraud, Fraud Detection, Job Postings, BERT-Multilingual, Bangladesh Job Market, Class Imbalance

1. Introduction

Online recruitment fraud has emerged as a critical cybersecurity and socioeconomic concern, particularly in digitally expanding labor markets where job seekers increasingly rely on online platforms for employment opportunities.

*Corresponding Author email: twinkletariq286@gmail.com

¹Department of Information and Communication Technology, Bangladesh University of Professionals, Dhaka, Bangladesh

DOI: <https://doi.org/10.64494/JFST/v4i1/tt/2026/04/67-92>

This section establishes the motivation for the study by examining the global rise of recruitment scams, the specific vulnerabilities of the Bangladeshi job market, and the limitations of existing fraud detection approaches. It also defines the research contributions, research questions, and overall organization of the paper.

1.1 Background, Motivation and The Bangladesh Context

The digitalization of labor markets has expanded employment opportunities while simultaneously creating new avenues for fraud. Online recruitment fraud has become a pervasive global concern, with malicious actors employing sophisticated tactics such as website impersonation, identity theft, advance-fee demands, and personal data harvesting (Gupta & Sharma, 2018; Patel & Kumar, 2021a). The problem is particularly acute in Southeast Asia, where 69% of Vietnamese job seekers reported receiving at least one fraudulent job offer (The Economic Times, 2025). Bangladesh is similarly vulnerable because youth unemployment, rapid digital adoption, and uneven digital literacy increase job seekers' exposure to deceptive employment offers. The country has approximately 1.94 million unemployed youths and a youth unemployment rate of 7.2%, while recruitment scam reports increased by 281% in 2022, primarily affecting young adults and female victims (The Daily Star, 2023). Financial fraud accounts for approximately 17% of reported cybercrimes, with mobile financial services such as bKash and Nagad frequently exploited as payment channels (Ahmed et al., 2021; The Business Standard, 2023). Although Bangladesh had approximately 124 million mobile internet users in 2024, recruitment activities conducted through Facebook, WhatsApp, and Telegram remain weakly regulated, further increasing the risk of fraudulent postings (Bangladesh Telecommunication Regulatory Commission [BTRC], 2024; Hassan et al., 2023). These global and national conditions demonstrate the urgent need for an automated, scalable, and regionally adapted recruitment fraud detection system.

1.2 Limitations of Existing Solutions

Existing fraud detection approaches are inadequate for the Bangladeshi context. Manual content moderation does not scale to the volume of postings on informal platforms and rule-based systems are easily circumvented by minor textual variations (Zhao et al., 2020). Most machine learning models trained on Western datasets fail to identify fraud signals specific to Bangladesh, such as Bangla English code-mixing, references to local payment systems like bKash, Nagad and the informal register common to Telegram and WhatsApp postings (Liu et al., 2021; Patel & Kumar, 2021b). Widely used benchmark datasets such as EMSCAD also lack South Asian representation (Vidros et al., 2017), making it

impossible to evaluate existing models in this context without domain-specific data collection.

1.3 Research Contributions

To address the limitations of existing solutions discussed in Section 1.2, this study makes the following three contributions:

- a. **A Bangladesh-Specific Dataset:** A curated collection of 20,000 annotated job postings sourced from formal Bangladeshi job portals and informal digital channels. The dataset captures region-specific fraud patterns, local linguistic features including code-mixing, and behavioral signals unique to the Bangladeshi job market.
- b. **A Context-Aware Detection Framework:** An enhanced Multilingual BERT model adapted to the Bangladeshi context, incorporating weighted loss functions to address class imbalance and stratified sampling to ensure representative evaluation across fraudulent and legitimate job posting classes.
- c. **Rigorous Real-World Testing:** Comprehensive experiments yielded a 95.6% test accuracy, with systematic comparisons against five baseline models and an analysis of deployment considerations in resource-constrained environments.

1.4 Research Questions

This study addresses three research questions:

- a. To what extent can transformer-based NLP models augmented with Bangladesh-specific features improve recruitment fraud detection compared to general-purpose models?
- b. What classification performance can be achieved by fine-tuning BERT-Multilingual on Bangladeshi job postings in the presence of class imbalance and multilingual code-mixing?
- c. How does the proposed localized framework compare with existing global baseline models when evaluated on Bangladesh-specific data?

2. Literature Review

The literature on online recruitment scam detection spans rule-based systems, traditional machine learning models, deep learning approaches, and recent transformer-based architectures. This section critically reviews these research directions and highlights their relevance to recruitment fraud detection in multilingual and region-specific settings. Particular attention is given to the

limitations of existing global datasets and models when applied to the Bangladeshi context.

2.1 Online Recruitment Fraud

Online recruitment fraud has evolved from simple fake job advertisements into sophisticated online scams conducted through job portals, social media, email, and messaging platforms such as WhatsApp and Telegram. Fraudsters often impersonate employers, advertise unrealistic salaries or work-from-home opportunities, request advance payments, and collect victims' personal or financial information (Akram et al., 2024; Taneja et al., 2025). These online and psychologically manipulative tactics demonstrate the need for automated, context-aware recruitment fraud detection systems.

2.2 Traditional Machine Learning Approaches

Patel and Kumar (2021a) developed a framework that combined textual and structural features with Random Forest and support vector machine classifiers, reporting an accuracy of 89.2% on job postings from Western platforms. Zhao et al. (2020) added contextual and historical trends (posting frequency, account age) and the accuracy was 87.5% on LinkedIn, but platform-specific metadata is essential. Proposing multi-modal combinations of both text and social media cues, Liu et al. (2021) achieved 91.0% accuracy through multimodal fusion of text and social network signals, but their dataset remains proprietary and unavailable for external replication.

A systematic review by Akhter et al. (2025) further emphasized the importance of data preprocessing, feature extraction, feature selection, contextual analysis, and appropriate model selection in machine-learning and deep-learning-based deceptive-text detection. Although their review focused on fake-news detection, these methodological considerations are also relevant to online recruitment-fraud classification.

2.3 Deep Learning and Transformer Architectures

Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks have demonstrated effectiveness in capturing semantic structures and sequential dependencies in text (Kim, 2014; Hochreiter & Schmidhuber, 1997). Palash et al. (2022) proposed a CNN-based model using the EMSCAD dataset to detect fraudulent job postings and demonstrated that deep learning could outperform several conventional machine-learning algorithms in this classification task.

The advent of transfer learning through large-scale pre-training, as exemplified by BERT, produced significant gains over prior methods. Devlin et al. (2019) reported improvements of up to 5.1 percentage points on multiple NLP benchmarks. Subsequent BERT-based fraud detection systems have further confirmed the value of contextual embeddings for identifying deceptive content (Chen & Wang, 2022; Kumar et al., 2023). However, most existing BERT-based fraud detectors rely on English-language models. BERT-Multilingual, pre-trained on 104 languages including Bangla, remains underexplored in fraud detection contexts, despite its demonstrated effectiveness in cross-lingual classification tasks in South Asia (Pires et al., 2019).

2.4 Regional Studies

Research on fraud in developing economies remains limited. In Bangladesh, Ahmed et al. (2021) identified the misuse of mobile financial services such as bKash and Nagad, while Das and Islam (2022) highlighted regulatory gaps, limited technical intervention, a 281% rise in recruitment fraud, and financial fraud accounting for 17% of reported cybercrimes. Regional studies also show that fraud patterns vary by context, including work from home scams in Vietnam and India and advance fee fraud in Nigeria (Nguyen et al., 2023; Sharma & Patel, 2022; Okonkwo & Azubuike, 2021). These findings indicate the need for a context specific detection model. Key research contributions are summarized in Table 1.

Table 1: Summary of Key Research on Recruitment Scam Detection

Study	Region	Methodology	Key Findings	Limitations
Gupta & Sharma, 2018	Global	Survey analysis	Identified common scam strategies	No automated model
Patel & Kumar, 2021a	Global	RF/SVM with TF-IDF	89.2% accuracy on formal platforms	Lacks informal context
Zhao et al., 2020	China	Contextual features	Behavioral patterns improve detection	Single platform evaluation
Devlin et al., 2019	Global	BERT transformer	Contextual embeddings improve NLP	Not regionally adapted
Ahmed et al., 2021	Bangladesh	Qualitative MFS analysis	Identified bKash exploitation	No ML framework

Study	Region	Methodology	Key Findings	Limitations
Das & Islam, 2022	Bangladesh	Cybercrime trends	17% financial fraud rate	No detection model

2.5 Research Gaps

Three important gaps stand out from the existing work:

- a. Regional Adaptation: Most models are constructed using Western datasets, implying that they do not capture fraud indicators specific to Bangladesh, such as references to bKash or Nagad, Bangla English code-mixing, and posts shared through unofficial platforms such as Telegram or WhatsApp.
- b. Class Imbalance: Previous research has not effectively addressed the problem of extreme class disparity. Many studies have omitted minority class evaluation altogether, despite the availability of established methods for mitigating this class imbalance.
- c. Localized Datasets: Publicly available datasets such as EMSCAD lack South Asian representation and few studies account for the deployment constraints relevant to low-resource settings.

3. Dataset Development and Engineering

A robust and contextually representative dataset is essential for developing an effective recruitment scam detection system. This section presents the construction of a Bangladesh-specific dataset consisting of job postings collected from both formal recruitment platforms and informal digital channels. It details the data collection strategy, dataset schema, annotation process, distributional characteristics, linguistic composition, and preprocessing pipeline used to prepare the data for transformer-based classification.

3.1 Data Collection Strategy

The dataset was developed through a systematic data collection process over 21 months, from February 2024 to November 2025. A total of 20,000 job postings were collected from formal platforms such as BDJobs, LinkedIn, and public Facebook groups, along with informal public channels such as Telegram and WhatsApp, representing Bangladesh’s recruitment landscape. Public platform data were collected using Python 3.9, BeautifulSoup 4.9, Selenium 4.1, and Tweepy, while Telegram data were retrieved through the Telethon library. Only publicly accessible groups were used, and no private or end to end encrypted

conversations were accessed. The pipeline used SHA 256 hashing for deduplication, automated rate limiting to comply with platform terms of service, and PostgreSQL for metadata storage. Ethical approval was obtained according to Bangladesh University of Professionals guidelines, and no personally identifiable information about job seekers was retained. The overall data source distribution and platform selection process are illustrated in Figure 1.

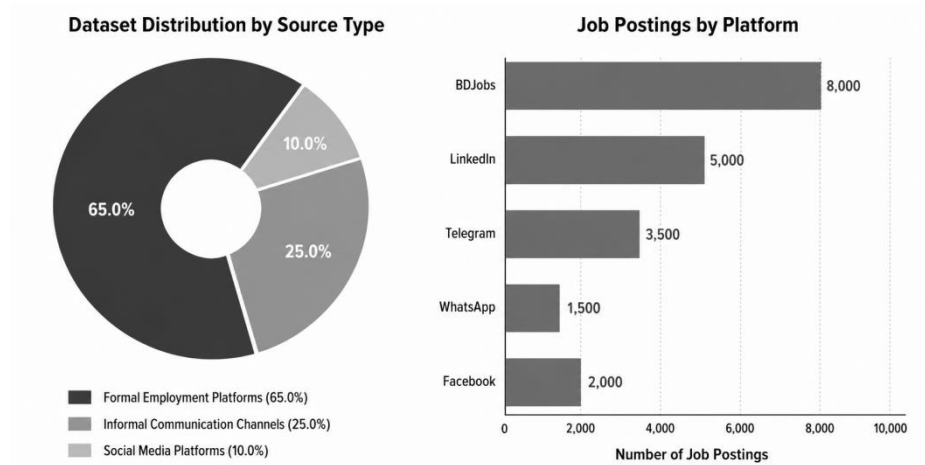


Figure 1: Data Sources and Platform Selection

3.1.1 Data Sources

- Formal Platforms (65%):** The majority of postings were sourced from BDJobs, Bangladesh's largest job portal, and LinkedIn, which has more than 3 million registered users in Bangladesh. Both platforms provide structured data and verified employer information.
- Informal Channels (25%):** Data were collected from 15 Telegram groups, contributing 3,500 postings, and several WhatsApp groups, contributing 1,500 postings. These channels have minimal moderation, making them common venues for fraudulent job advertisements.
- Social Media (10%):** The remaining 2,000 postings were collected from Facebook job groups with a combined membership exceeding 500,000 users.

3.2 Data Schema

Table 2 breaks down the full 18-field relational schema, which captures both the textual content and structural metadata of each posting.

Table 2: Dataset Schema and Field Specifications

Field Name	Data Type	Description
post_id	UUID	Unique identifier
source	String	Platform origin
date_posted	Date	Posting date
company_name	String	Employer identification
job_title	String	Position designation
job_description	Text	Primary content
salary	String	Compensation details
contact_emails	String []	Contact emails
contact_whatsapp	Boolean	WhatsApp indicator
text_for_model	Text	Consolidated text for NLP
label	String	Binary classification

3.3 Annotation Process

Labels were assigned through a rigorous three-stage annotation process.

Stage 1: Predefined rules were initially applied to label approximately 40% of the dataset by leveraging clear fraud indicators, including payment requests, unrealistic salary offers, incorrect contact information, and businesses already present on a blacklist.

Stage 2: All job postings were individually and manually reviewed by three professionals with expertise in cybersecurity and human resources. The annotators were specifically trained to identify cues unique to the Bangladeshi context, including company verification, salary validation, and job-cloning patterns.

Stage 3: In instances where annotators disagreed, which occurred approximately 20 percent of the time, a senior panel intervened and made the final call.

Cohen’s Kappa of 0.82 indicated strong inter annotator agreement. However, the 20% disagreement rate, resolved by a senior panel, shows that some fraud signals were difficult even for experts to classify, especially postings that closely resembled legitimate jobs. This ambiguity may limit model performance and supports the use of soft labels or confidence weighted training in future work.

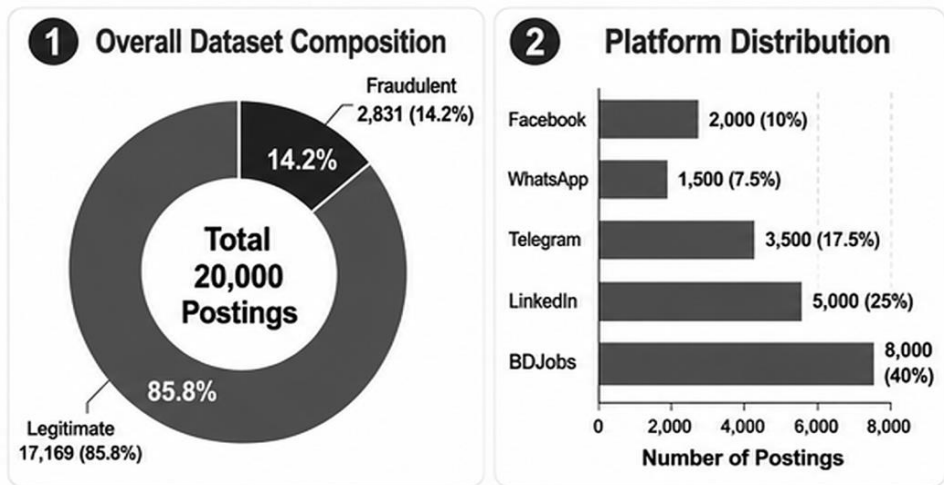
3.4 Dataset Characteristics

Figure 2 presents the main characteristics of the final dataset. The dataset contains 20,000 job postings collected from five platforms, including 17,169 legitimate postings and 2,831 fraudulent postings, as shown in Figure 2(a).

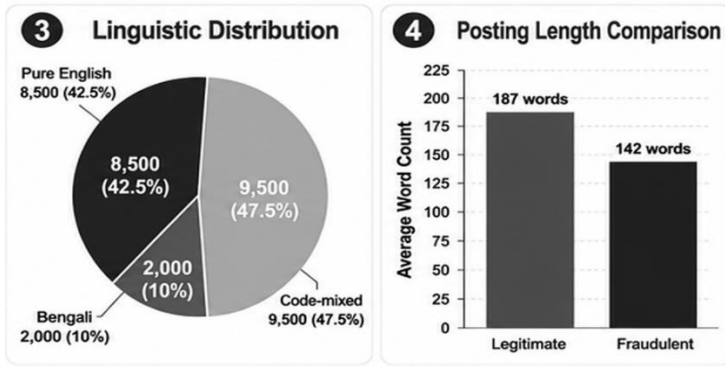
BDJobs contributes the largest portion of the dataset, followed by LinkedIn, Telegram, Facebook, and WhatsApp.

Figure 2(b) shows the linguistic distribution and posting length comparison. Code-mixed Bangla-English postings form the largest group, followed by pure English and Bangla postings. Legitimate postings are longer on average than fraudulent postings, indicating that scam posts are generally shorter and less descriptive.

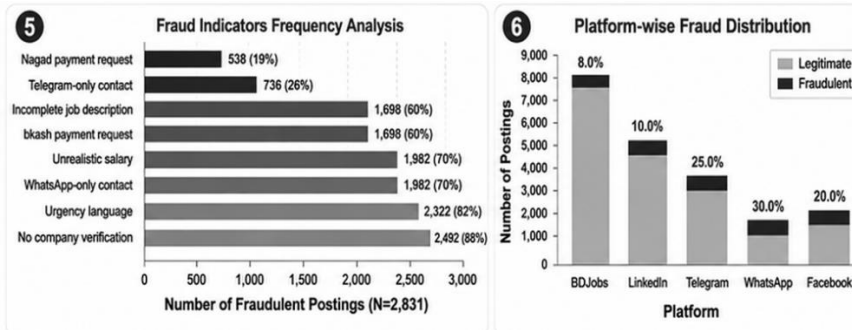
Figure 2(c) summarizes the major fraud indicators and platform-wise fraud distribution. The most frequent indicators include no company verification, urgency language, WhatsApp-only contact, unrealistic salary, bKash payment request, and incomplete job description. The platform-wise comparison shows that WhatsApp and Telegram contain higher fraud proportions than formal job platforms such as BDJobs and LinkedIn. Finally, Figure 2(d) presents the payment methods used in fraudulent postings, where bKash is the most common method, followed by Nagad and other or unspecified methods. These dataset characteristics highlight the class imbalance, linguistic diversity, platform variation, and fraud-specific behavioral patterns used for model training and analysis.



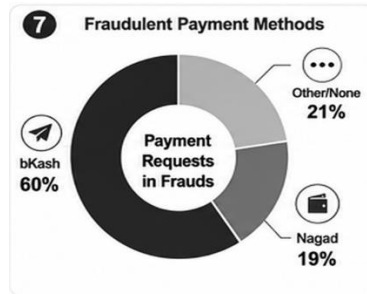
(a)



(b)



(c)



(d)

Figure 2: Dataset characteristics and distribution: (a) overall class composition and platform distribution; (b) linguistic distribution and posting length comparison; (c) fraud indicator frequency and platform-wise fraud distribution; and (d) fraudulent payment method distribution.

3.5 Preprocessing Pipeline

Figure 3 shows how raw job postings are turned into model-ready inputs through a five-step pipeline

Step 1: All relevant fields (job title, company name, description, salary, contact details) were concatenated into a single `text_for_model` field.

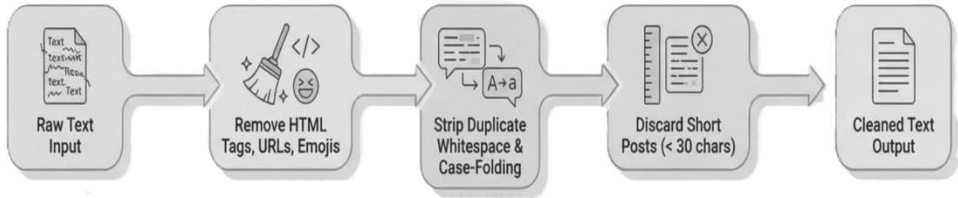


Figure 3: Data Preprocessing Pipeline

Step 2: The unified text was cleaned by removing excess whitespace, decoding HTML entities, stripping HTML tags, removing embedded URLs and masking contact information. Postings with fewer than 20 characters after cleaning were discarded (approximately 450 entries).

Step 3: Bangla English code-mixed content was deliberately preserved; Bangla text in Latin script was transliterated for representational uniformity.

Step 4: Text was tokenized using the BERT-Multilingual-cased tokenizer (vocabulary: 119,547 tokens), truncated or padded to 384 tokens with [CLS] and [SEP] tokens

Step 5: The dataset was partitioned into training (16,000 samples: 13,735 legitimate, 2,265 fraudulent), validation (2,000 samples), and test sets (2,000 samples) using stratified sampling, preserving the 85.8:14.2 class ratio across all splits.

4. Methodology

In this study, the transfer learning method is supervised and based on a pre-trained transformer, as demonstrated in Figure 4. It follows a structured methodology centered on five main steps, namely dataset preparation, preprocessing, through stratified splitting, BERT-Multilingual model building with regularization, optimized fine-tuning and class balancing and performance on held-out test data. The design placed higher concern over deployability, interpretability based on attention analysis and robustness based on regularization, although it remained modular enough to keep up with changing patterns of fraud. A planned extension of the architecture will integrate structural metadata features, including account age, posting frequency, and external link

safety scores, into the final dense layers alongside the [CLS] embedding, enabling a holistic risk score that combines textual signals with behavioral patterns.

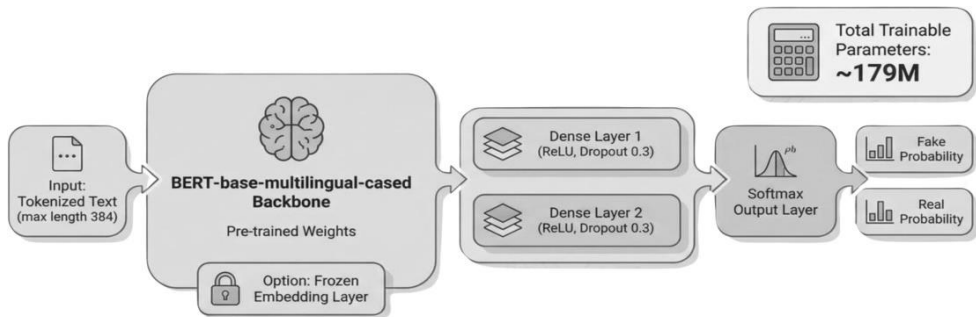


Figure 4: Enhanced BERT-Multilingual Architecture

4.1 Model Architecture: Enhanced BERT-Multilingual

The proposed framework is built upon the BERT-Multilingual architecture, a pre-trained bidirectional transformer language model based on the original BERT architecture with additional components incorporated and replacing original model to fit the specificities of the present model.

4.1.1 Base Architecture Selection

BERT Multilingual cased was selected as the base model because it supports 104 languages, including English and Bangla, and can process the code-mixed text common in Bangladeshi job postings (Devlin et al., 2019). Its cased tokenizer preserves capitalization patterns, which are useful because fraudulent postings often contain more all-caps words. It also supports effective transfer learning for cross lingual classification with fewer training samples than training a model from scratch (Pires et al., 2019). Table 3 presents the encoder configuration.

Table 3: Transformer Encoder Specifications

Component	Specification
Layers	12-layer bidirectional transformer
Hidden dimension	768
Attention heads	12 per layer
Feed-forward dimension	3,072
Trainable parameters	110 million

Component	Specification
Positional encoding	Learned embeddings (max length 512)
Activation function	GELU

4.1.2 Classification Head Design

The task-specific classification head is designed to map the [CLS] token representation to binary output probabilities, as detailed in Table 4.

Table 4: Classification Head Architecture

Layer	Configuration
Input	[CLS] token (768-dim)
Dropout	0.3
Fully connected	768 $\rightarrow$ 2
Activation	Softmax

4.2 Regularization Techniques

The regularization methods mentioned in Table 5 are employed to make the model more generalized and prevent overfitting.

Table 5: Regularization Strategy

Technique	Setting
Dropout	0.3 on hidden and attention layers
Weight decay	$\lambda = 0.01$ (AdamW)
Gradient clipping	Max norm = 1.0
Early stopping	Patience = 3 epochs

4.3 Class Imbalance Handling

Class imbalance is a critical challenge in this dataset, as legitimate postings constitute 85.8% of all samples while fraudulent postings account for only 14.2%, yielding an imbalance ratio of approximately 6:1. Without corrective measures, a classifier trained on this distribution would be biased toward the majority class, achieving high overall accuracy while performing poorly on the minority (fraudulent) class, precisely the class of greatest practical importance. To address this, three complementary strategies were employed, as illustrated in Figure 5.

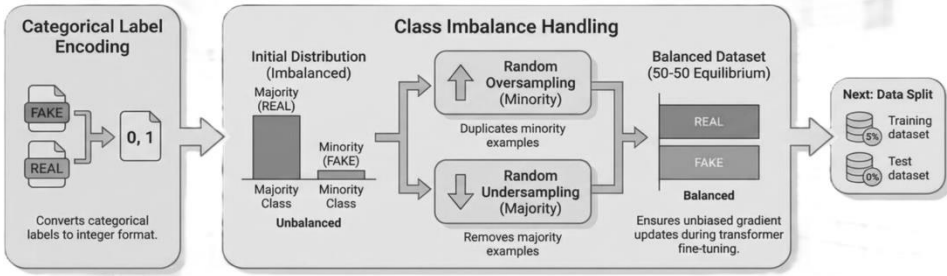


Figure 5: Class Imbalance Handling Strategy

4.3.1 Weighted Loss Function

The weights of the classes are assigned such that those occurring very rarely have a greater impact on the training:

$$w_{\text{real}} = \frac{n_{\text{total}}}{2 \times n_{\text{real}}} = \frac{20,000}{2 \times 17,169} = 0.583 \quad (1)$$

$$w_{\text{fake}} = \frac{n_{\text{total}}}{2 \times n_{\text{fake}}} = \frac{20,000}{2 \times 2,831} = 3.534 \quad (2)$$

Weighted cross-entropy is used to deal with class imbalance: mistakes on fraudulent postings are penalized with more weight:

$$L = -(w_{\text{real}} \sum y_{\text{real}} \log(\hat{y}_{\text{real}}) + w_{\text{fake}} \sum y_{\text{fake}} \log(\hat{y}_{\text{fake}})) \quad (3)$$

This drives the model to intercept more cases of fraud and increase the recall in the areas where it is needed the most.

This drives the model to intercept more cases of fraud and increase the recall in the areas where it is needed the most.

4.3.2 Stratified Sampling

Stratified random sampling is used to divide the data into training, validation and test sets to maintain the same 85.8:14.2 balance of classes in each split. This facilitates even-handed and consistent assessment and closer representation of actual performance.

4.3.3 Threshold Calibration

The default decision threshold of 0.5 can be calibrated after training. Lowering the threshold increases recall at the cost of precision; whereas raising the threshold improves precision but reduces recall. This enables cost-sensitive threshold tuning to match the error tolerance of different deployment contexts.

4.4 Training Protocol and Optimization

The training protocol was designed to ensure stable fine-tuning, reproducibility, and efficient convergence of the proposed BERT-Multilingual framework. Since transformer-based models are highly sensitive to learning rate selection, batch size, class imbalance, and overfitting, the optimization strategy was carefully configured using AdamW, learning-rate scheduling, gradient accumulation, mixed-precision training, and early stopping. This section describes the optimizer settings, training configuration, hardware environment, and comparative baseline models used to evaluate the effectiveness of the proposed approach.

4.4.1 Optimizer and Learning Rate

The AdamW optimizer is used with a learning rate of 2×10^{-5} , $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, and weight decay of 0.01. A linear warm-up over the first 10% of training steps is followed by a linear decay to zero.

4.4.2 Training Configuration

The training takes up to 5 epochs when using a batch size of 16 and gradient accumulation steps of 2, effectively achieving a batch size of 32. FP16 mixed-precision training accelerates the pace whilst a definite random seed of 42 ensures reproducibility as illustrated in Figure 6.

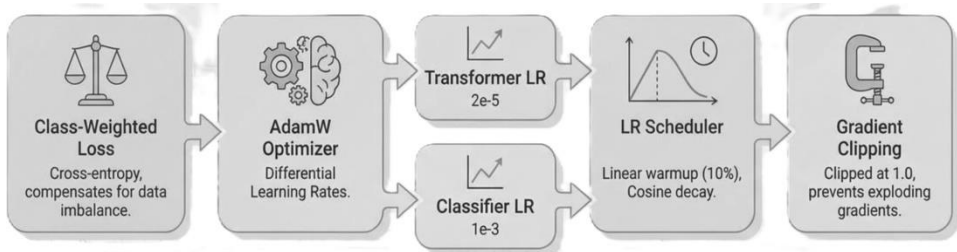


Figure 6: Training Regimen and Optimization Strategy

4.4.3 Hardware Infrastructure

All experiments were conducted using an NVIDIA Tesla V100 with 32 GB VRAM, an Intel Xeon Gold 6248 processor, and 128 GB RAM. Each epoch took about 18 minutes, with total training time of approximately 90 minutes. Although suitable for research, this hardware requirement may limit deployment on small job portals or mobile systems. Future work will use knowledge distillation and INT8 quantization to reduce model size by about 40% and improve inference speed by about 60%, with less than 2 percentage points of performance loss.

4.5 Baseline Models for Comparative Analysis

To evaluate the performance of the proposed framework, it is compared with five baseline models representing the evolution of fraud detection methods:

- a. Rule-Based System: A simple keyword-matching method that marks posts which have frequent fraud signs such as terms concerning money, time pressure and unrealistic expectations.
- b. Traditional ML (Random Forest): A Random Forest TF-IDF feature classifier with ($n_{estimators} = 100$) within a feature space of 5000 dimensional.
- c. LSTM-Based Deep Learning: A bi-directional LSTM equipped with an attention system, which is based on GloVe 300-dimensional embeddings, 2 layers, 128 hidden units
- d. Standard BERT-Base (English): BERT-base uncased fine-tuned on the same dataset with no multilingual support.
- e. BERT-Multilingual (Standard Fine-tuning): BERT-Multilingual-cased with normal fine-tuning (default dropout 0.1, no weighted classes),

5. Results and Discussion

This section evaluates the proposed framework for detecting recruitment scams in Bangladesh. It discusses training behaviour, test performance, baseline comparison, ablation results, threshold sensitivity, fraud indicators, generalization across platforms and languages, and deployment considerations, with emphasis on minority class detection, regional adaptability, and practical use.

5.1 Training Dynamics

Epoch-wise performance is provided in Table 6 and Figure 7. As shown in Table 6, Training loss was reduced from 0.55 to 0.18; validation loss from 0.45 to 0.27. The model stabilized at epoch 3 (91.5% validation accuracy), with only marginal gains thereafter (less than 1.2 pp), consistent with no extreme overfitting.

Table 6: Epoch-Wise Training and Validation Performance

Epoch	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)	Pattern
1	0.55	0.45	77.5	82.5	Initial

Epoch	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)	Pattern
2	0.40	0.38	85.0	88.5	Rapid
3	0.28	0.29	90.0	91.5	Stabilization
4	0.23	0.28	92.0	92.5	Marginal
5	0.18	0.27	93.0	92.7	Convergence

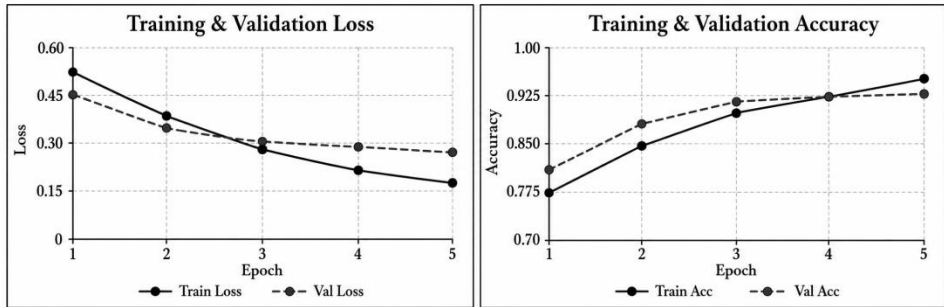


Figure 7: Training and Validation Performance Over 5 Epochs

5.2 Test Set Performance

The held-out test set (2,000 samples: 1,717 legitimate, 283 fraudulent) yielded excellent results, as reported in Table 7 and Figure 7.

Table 7: Test Set Performance by Class

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Real Jobs	94.0	96.0	95.0	1,717
Fake Jobs	89.0	85.0	87.0	283
Weighted Avg	93.5	95.6	94.5	2,000
Overall Test Accuracy: 95.6%				

5.3 Comparative Analysis

Table 8 shows performance on baselines that have been tested on the same test set.

Table 8: Comparative Performance Across Models

Model	Acc (%)	Prec, (Fake)	Recall (Fake)	F1 (Fake)	Time
Rule-Based	68.2	45.3	62.0	52.4	N/A
RF+TF-IDF	87.3	72.5	68.8	70.6	8 min

Model	Acc (%)	Prec, (Fake)	Recall (Fake)	F1 (Fake)	Time
Bi-LSTM	89.8	76.2	73.5	74.8	45 min
BERT-Base	92.1	82.5	79.2	80.8	75 min
BERT-Multi (Std)	93.4	85.1	81.7	83.4	82 min
Proposed	95.6	89.0	85.0	87.0	90 min

Key stepwise gains: (1) Rule-based to ML: +19.1 pp; (2) ML to Deep Learning: +2.5 pp; (3) Deep Learning to Transformer: +2.3 pp; (4) English to Multilingual BERT: +1.3 pp; (5) Standard to Enhanced BERT: +2.2 pp. The proposed model detects 85.0% of fraudulent postings and outperforms standard BERT-Multilingual by 3.3 percentage points on fraud recall, a statistically significant difference (McNemar test, $p = 0.003$).

5.4 Ablation Study

Table 9 presents the contribution of each component through ablation testing. Removing class weighting has the largest impact, reducing fraud F1 by 8.8 points. Enhanced dropout contributes approximately 2.5 points and multilingual pre-training provides a 5.9 point advantage over English-only BERT.

Table 9: Ablation Study Results

Configuration	Test Acc (%)	Fake F1 (%)
Proposed (Full)	95.6	87.0
No Class Weighting	93.8	78.2
Standard Dropout	94.9	84.5
BERT-Base Backbone	92.3	81.1

5.4.1 Performance Analysis

For legitimate postings, the model achieves a recall of 96.0% and precision of 94.0% with only 48 false positives (2.8% error rate), acceptable for platforms with human moderators. For fraudulent postings, the model correctly identifies 85.0% of scams with 89.0% precision, missing approximately 40 genuine fraud cases (14.1%). Despite the 6.07:1 class imbalance, the minority class F1-score of 87.0% remains within 8 points of the majority class score of 95.0%, confirming the effectiveness of the weighted loss function.

5.5 Threshold Analysis

Table 10 shows the performance of the model at varying levels, decision thresholds, providing the practitioners with the flexibility of calibration.

Table 10: Classification Performance Across Thresholds

Threshold	Acc. (%)	Prec. (Fake)	Recall (Fake)	F1 (Fake)
0.3	92.1	71.2	94.3	81.2
0.4	94.5	82.5	89.8	86.0
0.5	95.6	89.0	85.0	87.0
0.6	95.8	92.1	80.2	85.7
0.7	95.3	94.8	74.6	83.5

Table 10 shows the precision recall trade-off across decision thresholds. A threshold of 0.3 increases fraud recall to 94.3% but lowers precision to 71.2%, creating many false positives. In contrast, a threshold of 0.7 improves precision to 94.8% but misses 25.4% of scams. Therefore, threshold selection should depend on platform risk: 0.4 for high-risk Telegram and WhatsApp channels, 0.5 for Facebook and LinkedIn, and 0.6 to 0.7 for lower risk platforms such as BDJobs.

5.6 Linguistic Pattern Analysis

Attention weight analysis reveals the tokens most strongly associated with fraudulent classifications. Table 11 presents the top 10 tokens by mean attention weight across the fraudulent class in the test set, derived by averaging the attention scores from the final transformer layer's [CLS] token over all 283 fraudulent test samples.

Table 11: Top Fraud-Indicative Tokens Mean Attention Weight

Token / Phrase	Mean Attention Weight	Feature Category
earned the money at home	0.88	Unrealistic promise
bkash	0.87	MFS payment request
registration fee	0.84	Upfront payment demand
whatsapp only	0.81	Informal contact signal
income earned online	0.79	Vague income claim
urgent hiring	0.77	Urgency trigger
no experience	0.76	Unrealistic eligibility
nagad send koren	0.74	Bangla MFS phrase

Token / Phrase	Mean Attention Weight	Feature Category
limited seats	0.71	Artificial scarcity
call now	0.68	Pressure tactic

The dominant fraud signals fall into three linguistic categories: (1) Financial exploitation cues, direct references to bKash, Nagad, or registration fees; (2) Urgency and scarcity language, phrases such as “urgent hiring” and “limited seats”; and (3) Unrealistic promise phrases, vague income claims common in work-from-home scams. Notably, Bangla-script MFS references (e.g., “nagad send koren”) carry high attention weight, confirming the value of multilingual pre-training for capturing code-mixed fraud signals.

5.7 Generalization Analysis

This subsection evaluates the model beyond overall test set performance by examining results across recruitment platforms and language groups. Because scam patterns vary between formal portals, informal channels, and code-mixed contexts, platform wise and language wise analysis provides a practical view of model reliability and highlights areas needing more data or domain adaptation.

5.7.1 Platform-Specific Performance

Table 12 illustrates that there are notable performance differences across platforms. Formal platforms such as BDJobs and LinkedIn achieve 96 to 97% accuracy, while informal channels such as Telegram and WhatsApp achieve 91 to 92%, reflecting the greater variability of unmoderated content.

Table 12: Performance by Data Source Platform

Platform	Samples	Accuracy (%)	Fake F1 (%)
BDJobs	800	97.2	90.5
LinkedIn	500	96.8	89.1
Telegram	350	92.1	82.3
WhatsApp	150	91.5	80.7
Facebook	200	93.7	84.6

5.7.2 Language-Script Sub-group Analysis

Table 13 reports classification performance disaggregated by the linguistic composition of each posting.

Table 13: Performance by Language Composition

Script Composition	Samples (Test)	Accuracy (%)	Fake F1 (%)
English-only (42.5%)	850	96.8	88.9
Code-mixed Bangla English (47.5%)	950	95.1	86.4
Bangla-only (10%)	200	91.5	79.3

The model performs lower on Bangla only postings, with a Fake F1 of 79.3% compared with 88.9% for English only postings. This 9.6 percentage point gap is likely due to the limited Bangla only data in the training set and weaker Bangla representation in BERT Multilingual. Adding more Bangla only data and fine tuning with BanglaBERT could improve performance.

5.8 Real-World Deployment Considerations

Three deployment modes may be considered for the proposed framework: a real-time API for screening newly posted jobs, periodic batch processing for large collections of advertisements, and a browser extension for alerting users to potentially suspicious postings. The appropriate deployment configuration should be selected according to platform resources, workload, and moderation capacity. A tiered review strategy may also be adopted, in which low-risk postings are processed automatically, uncertain cases are referred for brief human review, and high-risk cases receive detailed examination. To improve transparency and user trust, future implementations should provide a clear “Reason for Flagging” alongside each prediction, such as an unrealistic salary claim, a request for payment through bKash, or WhatsApp-only contact information. Such explanations may be generated using post-hoc interpretability techniques, including SHAP, to help moderators understand the factors influencing the model’s decision.

5.9 Limitations

The data window, which spans February 2024 to November 2025, means that the static model may experience data drift after deployment as scammers adapt their tactics. The dataset is also skewed toward formal channels, which account for approximately 65% of the collected postings, leaving informal channels underrepresented. Inter-annotator agreement was strong, with a kappa value of 0.82, although some disagreement remained and regional language coverage was limited. The model also failed to detect approximately 15% of fraudulent

postings and may remain vulnerable to adversarial examples and deliberate evasion attempts. Additional practical challenges include integration with existing recruitment systems, limited model interpretability, and privacy concerns associated with scanning user-generated messages.

6. Conclusion

This study proposed an enhanced BERT Multilingual framework for detecting fraudulent online job postings in Bangladesh. Using a Bangladesh specific dataset of 20,000 annotated postings from formal portals, social media, and informal channels, the model achieved 95.6% accuracy and an 87.0% fraud class F1 score.

The results show that regional adaptation improves scam detection compared with rule based, traditional machine learning, LSTM, English only BERT, and standard BERT Multilingual models. Key fraud indicators included mobile financial service references, WhatsApp only contact, urgency language, unrealistic salary claims, and Bangla English code mixed expressions. Ablation results further confirm that class imbalance handling is essential for detecting fraudulent postings.

Overall, the framework offers a scalable basis for recruitment scam detection in emerging digital labour markets, although future effectiveness will require continued adaptation, broader linguistic coverage, and responsible platform integration.

7. Future Works

Future research will improve robustness, adaptability, and deployment readiness. The dataset should be expanded with more Bangla only postings and informal channel data to reduce language and platform imbalance. Additional structural and behavioural features, such as employer profile age, posting frequency, link safety, account verification, and communication patterns, may also improve fraud detection. Multimodal inputs, including images, documents, and screenshots, can help identify visually deceptive scams.

For deployment, knowledge distillation, pruning, and INT8 quantization can make the model lighter for low resource platforms and mobile use. Explainable AI methods such as SHAP or attention based interpretation can provide human readable reasons for flagged postings. Future studies may also explore active learning, continual learning, federated learning, and cross lingual transfer to keep the system updated while preserving privacy and extending it to other South Asian labor markets.

Acknowledgement

The authors acknowledge the Department of Information and Communication Technology, Bangladesh University of Professionals, for academic and institutional support during this study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability

The dataset contains job posting text and platform specific metadata collected from public sources. For privacy, platform compliance, and fraud signal protection, the full dataset is not publicly released. An anonymized version or processed features may be available from the corresponding author upon reasonable request and ethical approval.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AI Declaration

AI assisted tools were used only for grammar checking and writing clarity. All data, figures, tables, analysis, and scientific conclusions were produced by the authors.

References

- Ahmed, S., Rahman, M., & Hossain, K. (2021). Mobile financial services fraud in Bangladesh: Patterns and prevention strategies. *Journal of Financial Crime*, 28(3), 742–758. <https://doi.org/10.1108/JFC-09-2020-0189>
- Akhter, N., Tuba, S. A., & Khatun, A. (2025). A systematic analysis of machine learning and deep learning techniques to detect fake news. *Journal of FST*, 3(1), 171–184. <https://doi.org/10.64494/JFST/v3i1/na/2025/08/171-184>
- Akram, N., Irfan, R., Al-Shamayleh, A. S., Kousar, A., Qaddos, A., Imran, M., & Akhunzada, A. (2024). Online recruitment fraud (ORF) detection using deep learning approaches. *IEEE Access*, 12, 109388–109408. <https://doi.org/10.1109/ACCESS.2024.3435670>

- Bangladesh Telecommunication Regulatory Commission. (2024). *Internet subscribers* from <https://btrc.gov.bd/pages/static-pages/6922e0a3933eb65569e27f59>. [Access Date: 13 June 2026]
- The Business Standard. (2023, June 20). Financial fraud cases rise in Bangladesh. <https://www.tbsnews.net/bangladesh/crime/financial-fraud-cases-rise-bangladesh>. [Access Date: 13 June 2026]
- Chen, L., & Wang, Y. (2022). BERT-based fraud detection in online marketplaces. *Expert Systems with Applications*, 192, Article 116385. <https://doi.org/10.1016/j.eswa.2021.116385>
- The Daily Star. (2023, March 15). Employment scams surge 281% in Bangladesh. <https://www.thedailystar.net/news/bangladesh/crime-justice/employment-scams-surge-281-percent>. [Access Date: 13 June 2026]
- Das, R., & Islam, M. (2022). Cybercrime trends in Bangladesh: A comprehensive analysis. *International Journal of Cyber Criminology*, 16(1), 45–62.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- The Economic Times. (2025). Vietnam job seekers face rising online fraud threats. <https://economictimes.com/news/international/vietnam-job-seekers-fraud>. [Access Date: 13 June 2026]
- Gupta, A., & Sharma, R. (2018). Online recruitment fraud: A comprehensive survey. *Computers & Security*, 78, 56–72. <https://doi.org/10.1016/j.cose.2018.06.002>
- Hassan, M. M., Ahmed, N., & Karim, R. (2023). Digital recruitment practices in Bangladesh: Opportunities and vulnerabilities. *Asian Journal of Business Research*, 13(2), 89–108. <https://doi.org/10.14707/ajbr.230>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

- Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1746–1751). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1181>
- Kumar, R., Singh, P., & Gupta, V. (2023). Transformer models for fraud detection: A systematic review. *ACM Computing Surveys*, 55(4), 1–35. <https://doi.org/10.1145/3532213>
- Liu, Z., Chen, H., & Wang, X. (2021). Multimodal fraud detection using deep learning. *Information Processing & Management*, 58(5), Article 102678. <https://doi.org/10.1016/j.ipm.2021.102678>
- Nguyen, T. H., Le, V. D., & Tran, M. K. (2023). Overseas employment scams targeting Vietnamese workers: Patterns and prevention. *Asian Journal of Criminology*, 18(2), 156–175.
- Okonkwo, C. N., & Azubuike, E. (2021). Advance fee fraud in Nigeria: Evolution and countermeasures. *African Journal of Criminology*, 14(3), 78–95.
- Palash, M. I. A., Diganto, A., Fatan, O. N., Taher, K. A., & Al Nahian, M. J. (2022). Detection of fake job postings on online using convolutional neural network. *Journal of FST*, 1(1), 207–217. <https://journal.bup.edu.bd/Journals/Article/Download/2090>
- Patel, H., & Kumar, S. (2021a). Machine learning approaches for online job fraud detection. *Expert Systems with Applications*, 173, Article 114658. <https://doi.org/10.1016/j.eswa.2021.114658>
- Patel, H., & Kumar, S. (2021b). Detecting fraudulent job postings using ensemble methods. *Journal of Information Security and Applications*, 59, Article 102824. <https://doi.org/10.1016/j.jisa.2021.102824>
- Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4996–5001). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1493>
- Sharma, A., & Patel, R. (2022). Work-from-home scams in India during COVID-19: Patterns and detection. *Indian Journal of Cybersecurity*, 8(2), 112–128.

- Taneja, K., Vashishtha, J., & Ratnoo, S. (2025). Fraud-BERT: Transformer-based context-aware online recruitment fraud detection. *Discover Computing*, 28, Article 9. <https://doi.org/10.1007/s10791-025-09502-8>
- Vidros, S., Koliass, C., Kambourakis, G., & Akoglu, L. (2017). Automatic detection of online recruitment frauds. *Future Internet*, 9(1), Article 6. <https://doi.org/10.3390/fi9010006>
- Zhao, Y., Chen, X., & Liu, H. (2020). Predicting fraudulent job postings using contextual features. *Information Systems*, 89, Article 101456. <https://doi.org/10.1016/j.is.2020.101456>