IMP FINAL PRESENTATION

GROUP 35

# BEWARE

## Battling Employment Woes and Recruitment Exploit

A Data-Driven Solution to Combat Fake Job Postings

Faculty Advisor: Prof. Snehal Shekatkar

SME: Prof. Arnab Chakrabarti

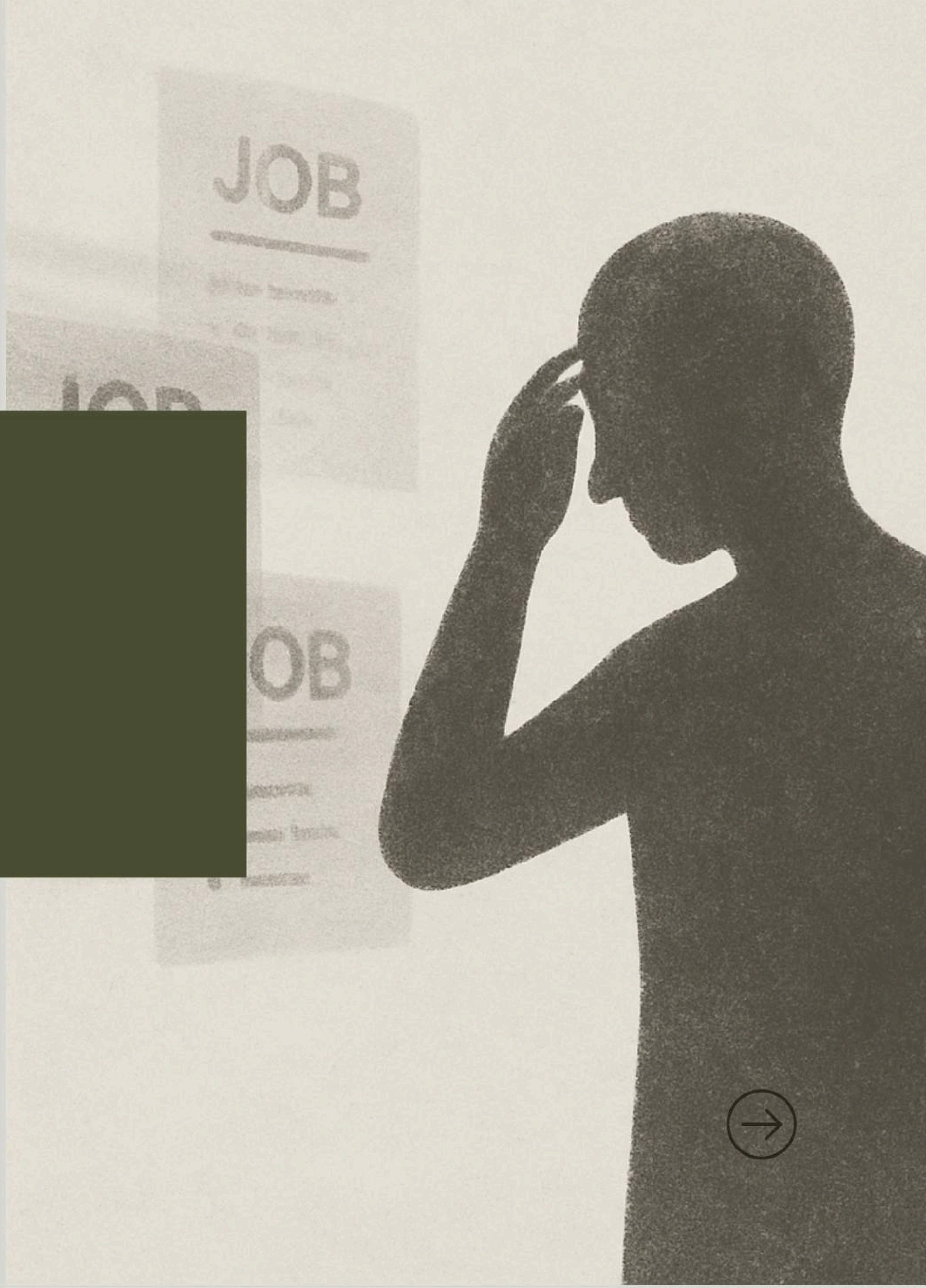
THE FAKE JOB POSTING CRISIS

# WHAT'S HAPPENING:

- Fresh graduates encounter fraudulent job postings on LinkedIn, Naukri, Internshala, WhatsApp
- Victims lose money, time, and trust in legitimate hiring platforms
- Despite being widespread, few tools exist to both detect AND prevent these scams

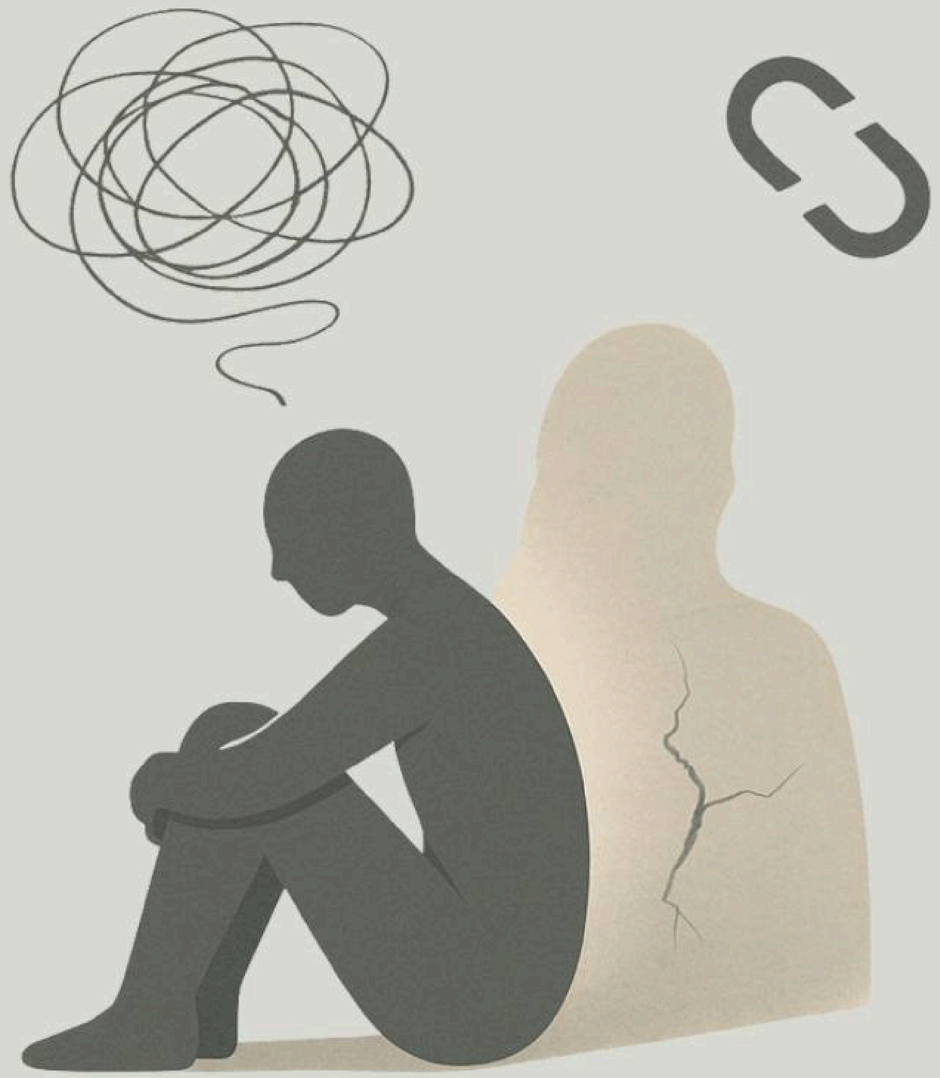
# WHO'S AFFECTED:

- Students and fresh graduates
- First-generation job seekers
- Anyone navigating digital job markets



THE HUMAN COST:

# WHY THIS MATTERS - REAL IMPACT




**Financial Loss** :
- $86 million lost in the US alone (FTC, 2022)
- Average victim loses ₹15,000-50,000 (estimated from victim reports)

**Time Wasted** :
- Hours spent on fake applications
- Delayed entry into real employment

**Psychological Damage** :
- Loss of confidence
- 40% reduction in trust toward job platforms (Ansar & Hussain, 2025)
- Some withdraw from job searching entirely (Ravenelle et al., 2022)

WHAT WE SET OUT TO DO:

# OUR OBJECTIVE

To address the growing threat of fraudulent job postings, our project focuses on developing practical, data-driven solutions that safeguard job seekers and enhance transparency in digital hiring platforms. We set out not only to detect scams but also to understand their impact and empower students with the tools to stay safe.


BEWARE

# WHY INTERDISCIPLINARY?

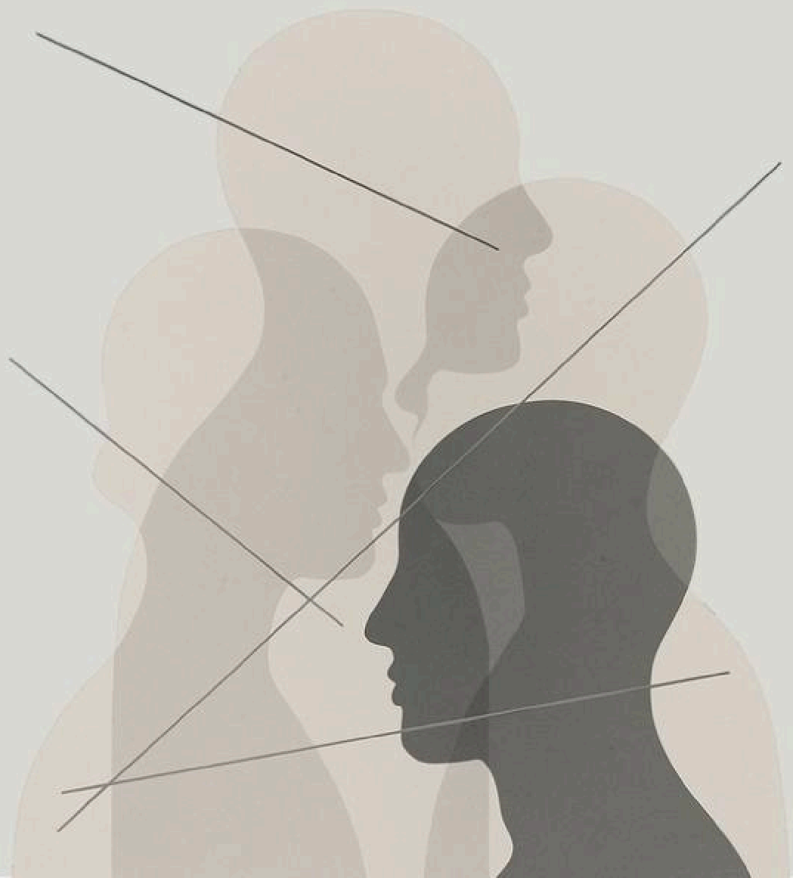
The Challenge Requires:

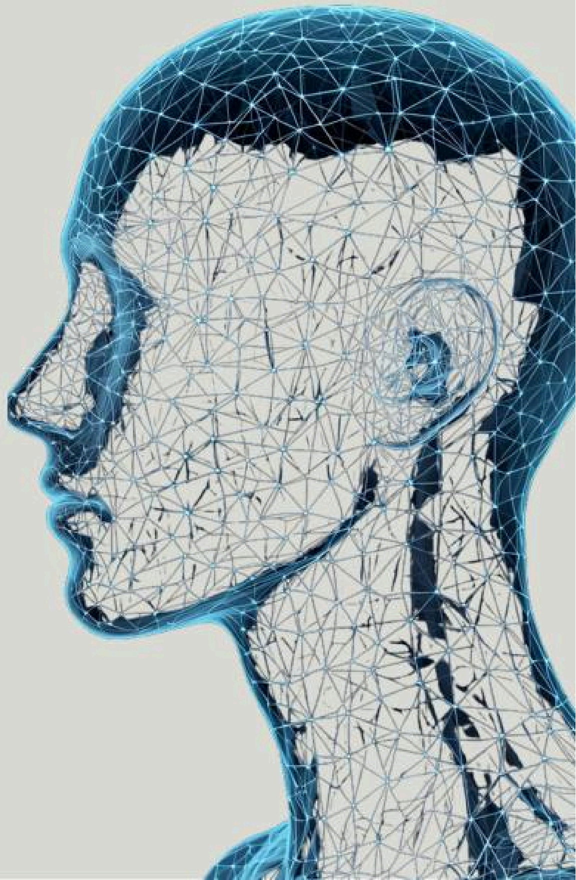
- Technology to detect patterns at scale
- Data Science to understand fraud signals
- Economics to quantify real-world impact
- Marketing to reach and educate audiences
- Design to make solutions accessible

Our Insight: You can build the most accurate model in the world, but if students don't know it exists or can't use it, you've solved nothing.

## INTERDISCIPLINARY SUCCESS METRICS

| Objective | Disciplines Involved | Success Indicator |
|---|---|---|
| Accurate Detection | CS, Data Science | 96.4% fraud recall |
| User Adoption | Marketing, Design | 22K views, 800 interactions |
| Economic Impact | Economics, Analytics | ₹5.7-19.2L prevention/month |
| Accessibility | Design, Marketing | 50+ on-ground engagements |





THE SCALE OF THE PROBLEM

# KEY INSIGHTS & DATA ANALYSIS

**Key Statistics**:

- 1 in 5 fresh graduates encounter fraudulent job postings in their first year
- $86 million lost to employment scams in the US (2022, FTC)
- ₹220+ million USD lost in first 6 months of 2024 alone
- Job frauds are the fastest-growing cybercrime in India (CERT-In)

**Our Combined Dataset:**

- 27,880 total job postings analyzed
- 39% fraudulent (10,866) vs 61% legitimate (17,014)
- Sources: Fake Job Posting dataset + EMSCAD dataset (2014-2016)

Higher fraud percentage than real-world (~4%) helps the model learn fraud patterns more effectively

**Statistical Insights**:

| **Fraud Indicator** | **Presence in Frauds** |
|---|---|
| Payment requests | 68% |
| Fake partnerships | 29% |
| Dodgy contact details | 34% |

Fraud detection requires **multiple features combined**, not just one warning sign

QUALITATIVE
FINDINGS

**Local Scam Tactics We Identified :**

- UPI payment references
- Walk-in interview promises
- WhatsApp-based recruitment
- Regional language mixing
- "Work from home" with upfront fees

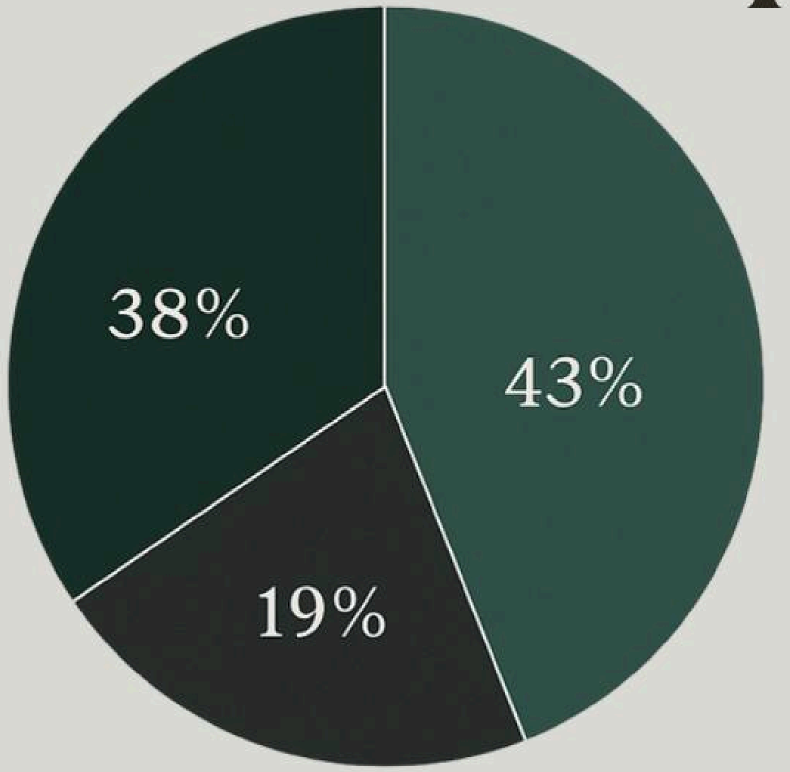
**Direct Costs:**

- Registration fees, training costs, security deposits
- Average victim loss: ₹15,000-50,000

**Hidden Costs (Research by Ravenelle et al.) :**

- Wasted time on fraudulent applications
- Delayed entry into legitimate employment
- 40% reduction in willingness to use job boards after scam exposure
- Development of extreme behaviors:
- Mass applying without discernment, OR
- Complete withdrawal from job searching

# THE THREE TYPES OF FRAUD WE FOUND




41% Obviously fraudulent

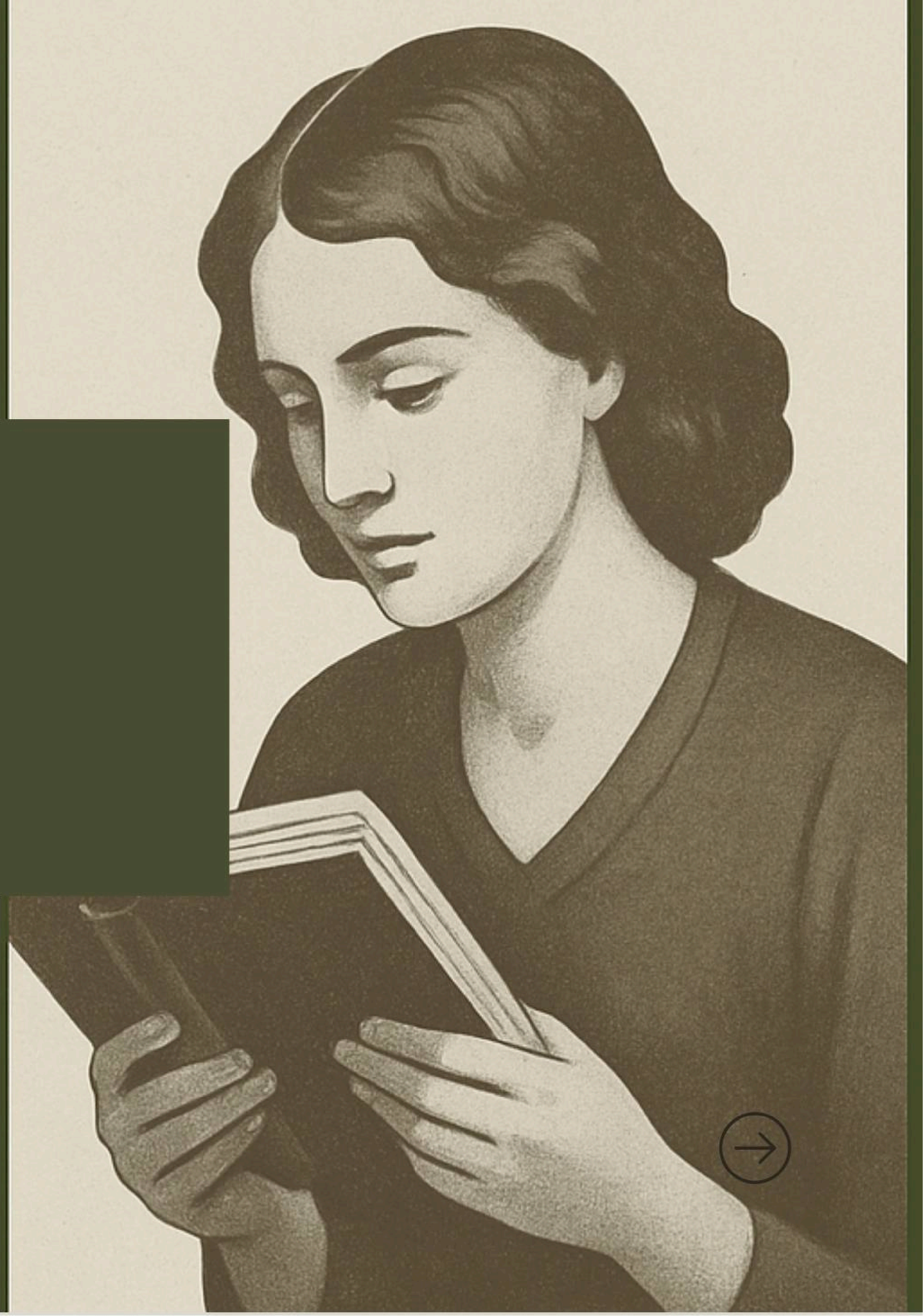
38% Professionally written but suspicious

19% Indistinguishable from real postings

WHAT WORKED

# LITERATURE REVIEW

| Model | Performnce | Key Insight |
|---|---|---|
| Gradient Boosting (Naude et al.) | F1: 0.88 | Empirical rule features crucial |
| Deep Neural Network (Swetha et al.) | 98% accuracy | Categorical features reduce complexity |
| Random Forest (Vidros et al.) | Strong baseline | EMSCAD dataset foundation |



WHAT WAS MISSING

# GAPS WE ADDRESSED

- Outdated Data - Most models trained on 2014-2016 data
- Western-Centric - Limited research on the Indian job market fraud
- No User Interface - High accuracy models with no accessible tool
- Missing Awareness - Technical solutions without an education component

Our Solution: Addresses all four gaps simultaneously

# OUR 3-COMPONENT SOLUTION ARCHITECTURE

**Machine Learning Model** :

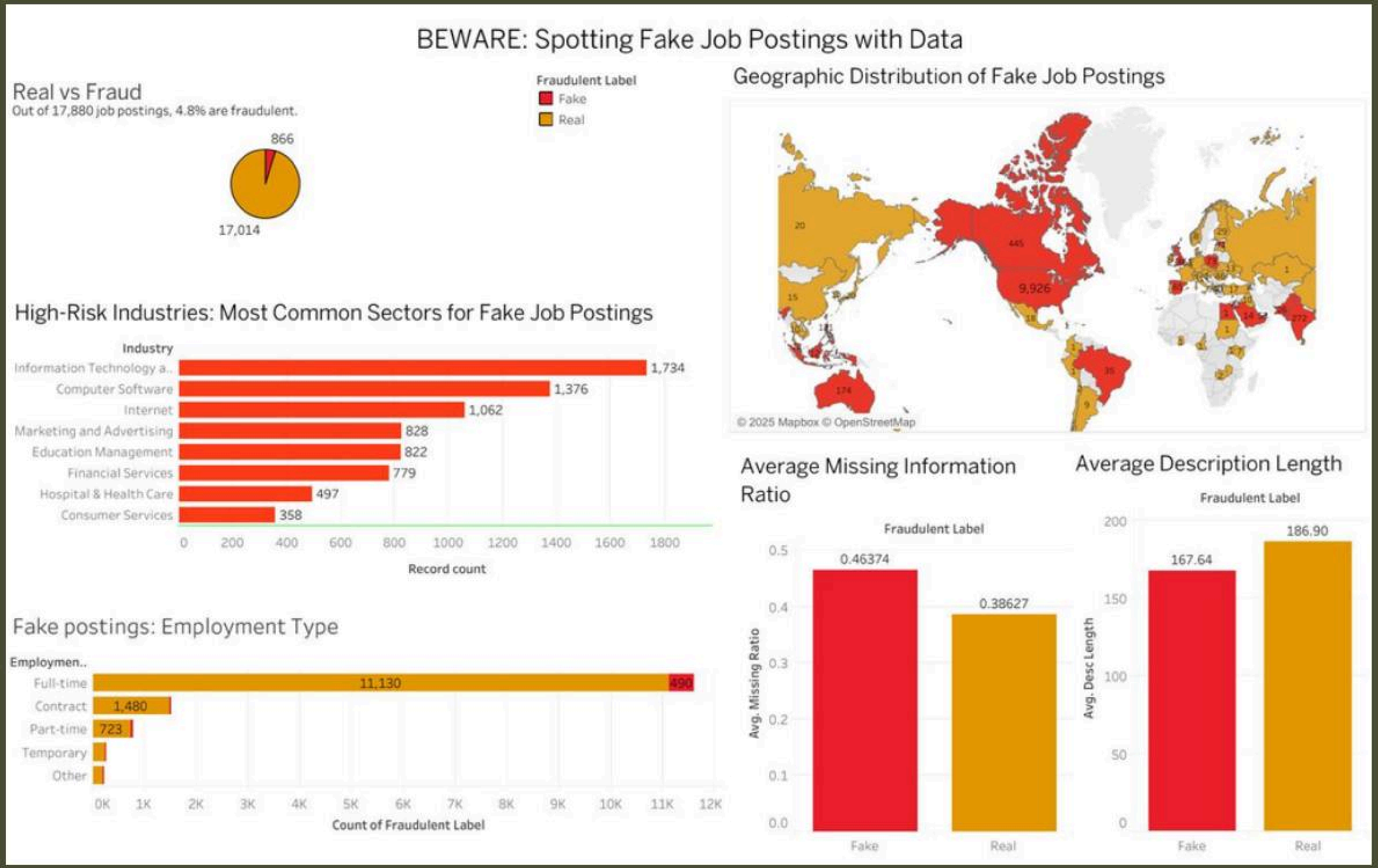
- Core fraud detection using LightGBM
- 427 features analyzed per job posting

**BEWARE Website**:

- Simple text/image input
- Clear risk assessment output

**Analytical Dashboard** :

- Industry-wise fraud distribution
- Geographic concentration mapping



# DEVELOPMENTAL PROCESS

**43 Engineered Fraud Indicators :**

- Payment detection - "registration fee", "security deposit", UPI mentions
- Fake endorsements - "authorized partner", false company tie-ups
- Scam job types - "data entry", "form filling", "copy-paste work."
- Income guarantees - "earn ₹5000/day", "assured payout."
- Suspicious contact - WhatsApp-only, no professional email, shortened URLs
- Writing style - Excessive punctuation, ALL CAPS, readability scores

**384 Semantic Embeddings:**

- Generated using SentenceTransformer (all-MiniLM-L6-v2)
- Captures meaning beyond explicit keywords

**Why LightGBM Won :**

| Criteria | LightGBM | BERT | Random Forest |
|---|---|---|---|
| Training Time | 3 Mins | 45 Mins | 8 Mins |
| Fraud Detection | 96.4% | 94.1% | 91.4% |
| False Alarms | 0.1% | 0.8% | 2.1% |
| Explainability | SHAP Compatible | Limited | Medium |

# FINE-TUNING THE MODEL

**Challenge :**

Finding optimal hyperparameters from thousands of combinations

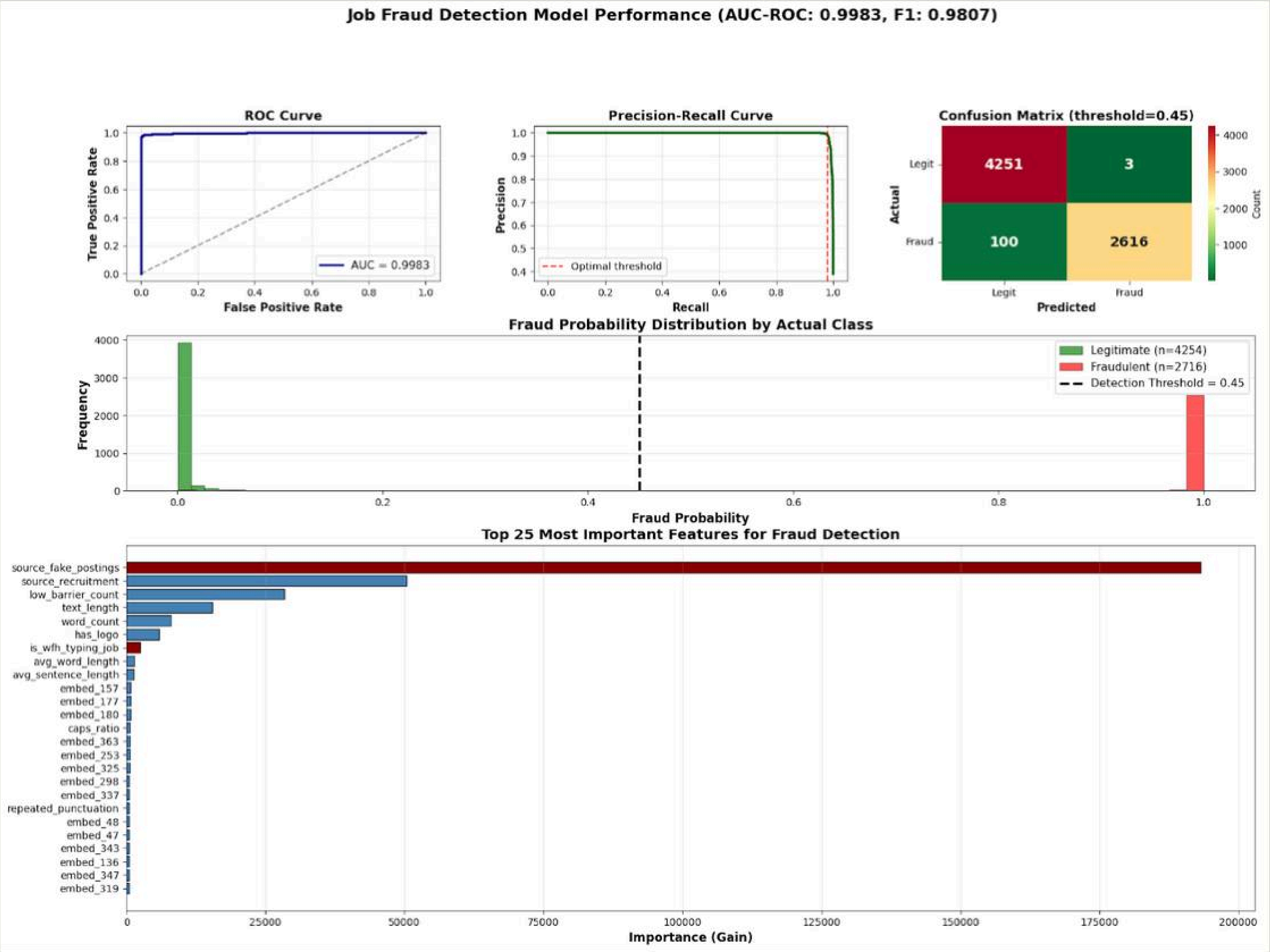
**Solution :**

Optuna framework for automated optimization

- Intelligent parameter search (not random)
- Balanced AUC and fraud recall
- Avoided time-consuming grid search

**Threshold Selection:**

- Not just a technical detail - a design choice
- Final threshold: 0.45 (maximizes fraud detection while minimizing false alarms)
- Adjustable based on user risk tolerance



# BUILDING THE INTERFACE

**Input Methods Designed :**

Text & Image

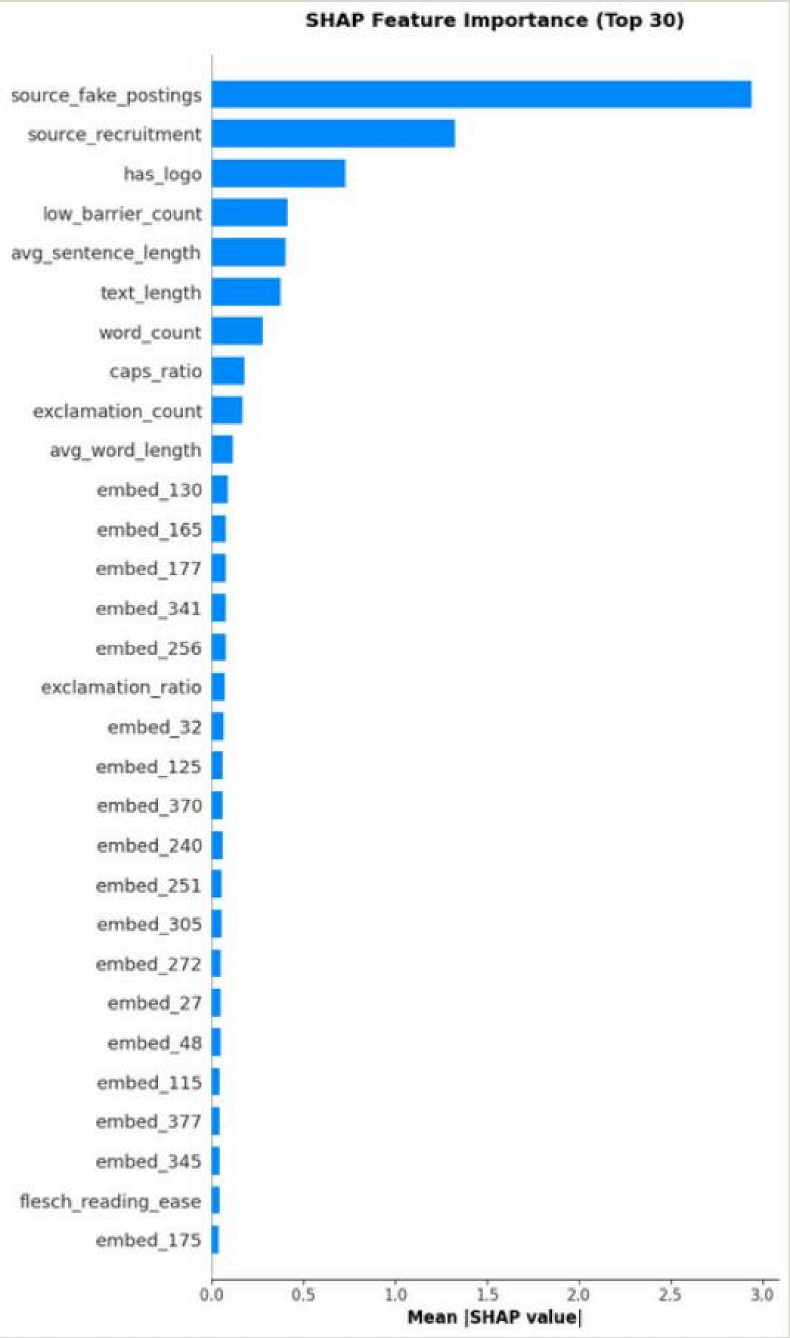
- **Text input:** Copy-paste from job portals, emails, WhatsApp
- **Image input:** OCR extraction from screenshots, flyers, photos

**Output Design Philosophy :**

Make technical insights accessible to everyone

**What Users See:**

- **Visual risk gauge** (Low → Moderate → High → Critical)
- **Clear verdict** (Fraudulent vs Legitimate)
- **Confidence level** (How certain is the model?)
- **Red flag system** (What exactly triggered the alert?)
- **Action guidance** (What should you do?)
- **SHAP explanations** (Why did the model decide this?)



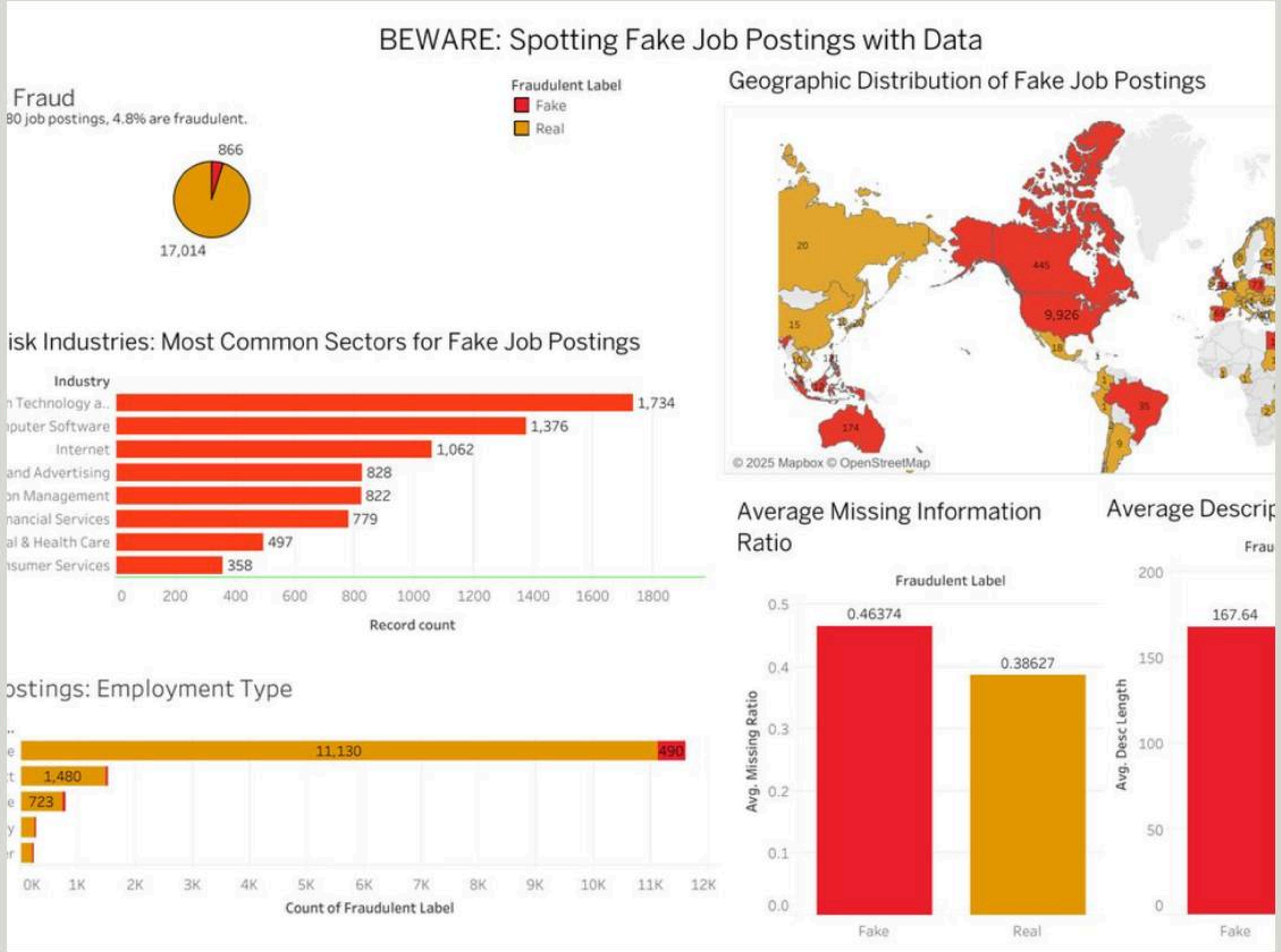
# DASHBOARD CREATION

**What It Shows :**

- Fraud distribution across industries
- Prevalence by job type (full-time, remote, contract)
- Geographic hotspots for fraud activity
- Description completeness patterns
- Missing information ratios

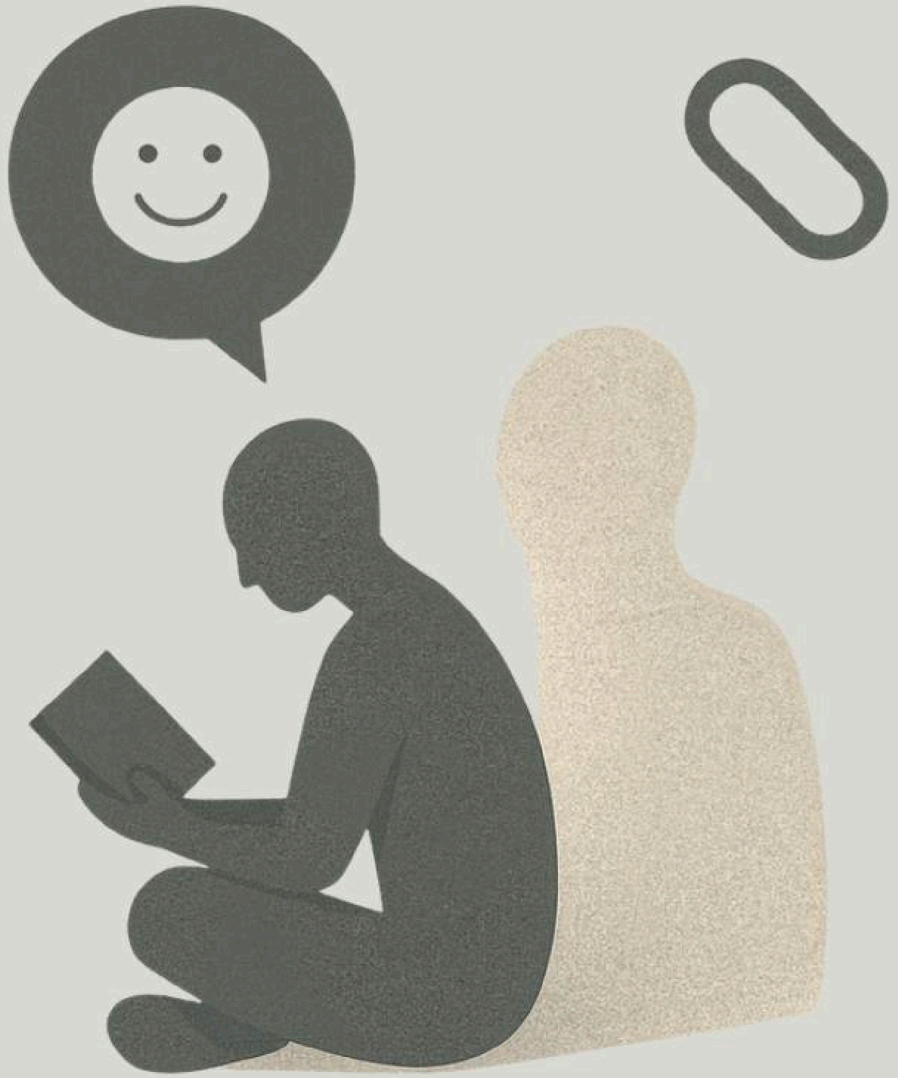
**Why It Matters :**

- Validates what the model learned
- Helps users understand broader patterns
- Supports policy recommendations



TECHNICAL IMPLEMENTATION CHALLENGES:

# REAL PROBLEMS WE SOLVED



**OCR Variability** :
- Problem: Image quality affects text extraction
- Solution: Accept ±3-7% probability fluctuation, added confidence indicators

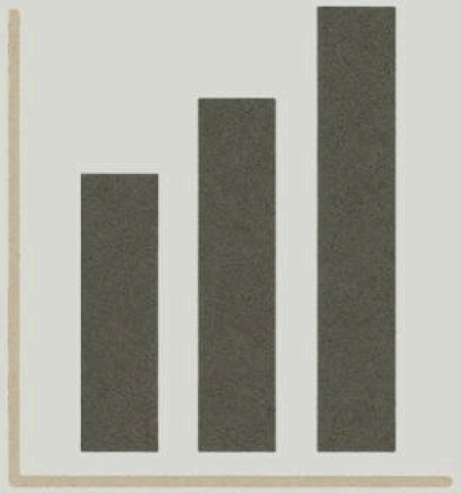
**Dataset Imbalance** :
- Problem: 39% fraud rate vs 4% real-world rate
- Solution: Documented limitation, planned for threshold adjustment in deployment

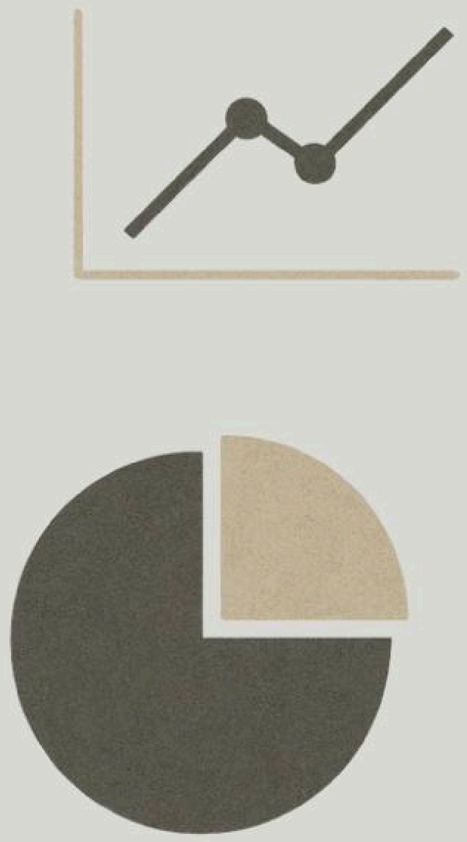
**Language Limitation** :
- Problem: The Model only understands English
- Future Work: Multi-language support for Hindi, regional languages

**Evolving Fraud Patterns** :
- Problem: Scams adapt over time
- Solution: Built a retraining pipeline for regular updates

HOW WE KNOW IT WORKS

# VALIDATION & TESTING RESULTS



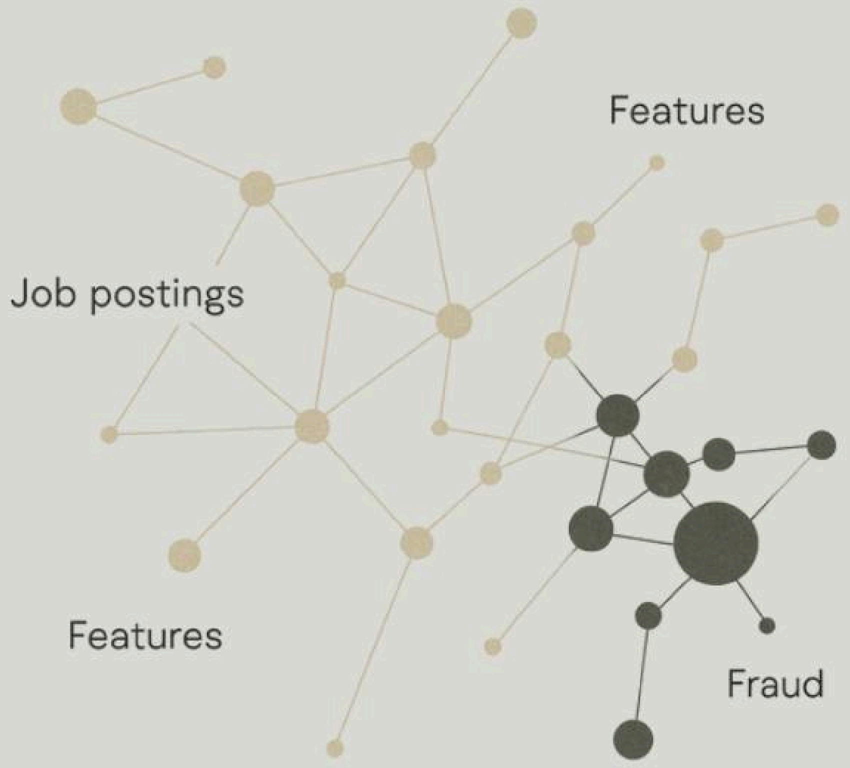


**Final Model Performance :**
**(**Test Set: 6,970 postings)

- **AUC-ROC: 0.9980** (near-perfect discrimination)
- **Accuracy: 98.6%** (6,886 correct out of 6,970)
- **Fraud Recall: 96.4%** (caught 2,618 out of 2,716 frauds)
- **False Positive Rate: 0.1%** (only 3 legitimate jobs flagged)

**Real-World Impact Projection :**
For 1,000 monthly users with 4% fraud rate (40 scams)

- **38-39** frauds prevented per month
- **₹5.7-19.2 lakh** in losses prevented monthly



# THE CAMPAIGN

HYPERLOCAL @ FLAME UNIVERSITY

**Bite-size, audience-focused storytelling on Instagram**

**Core features:**

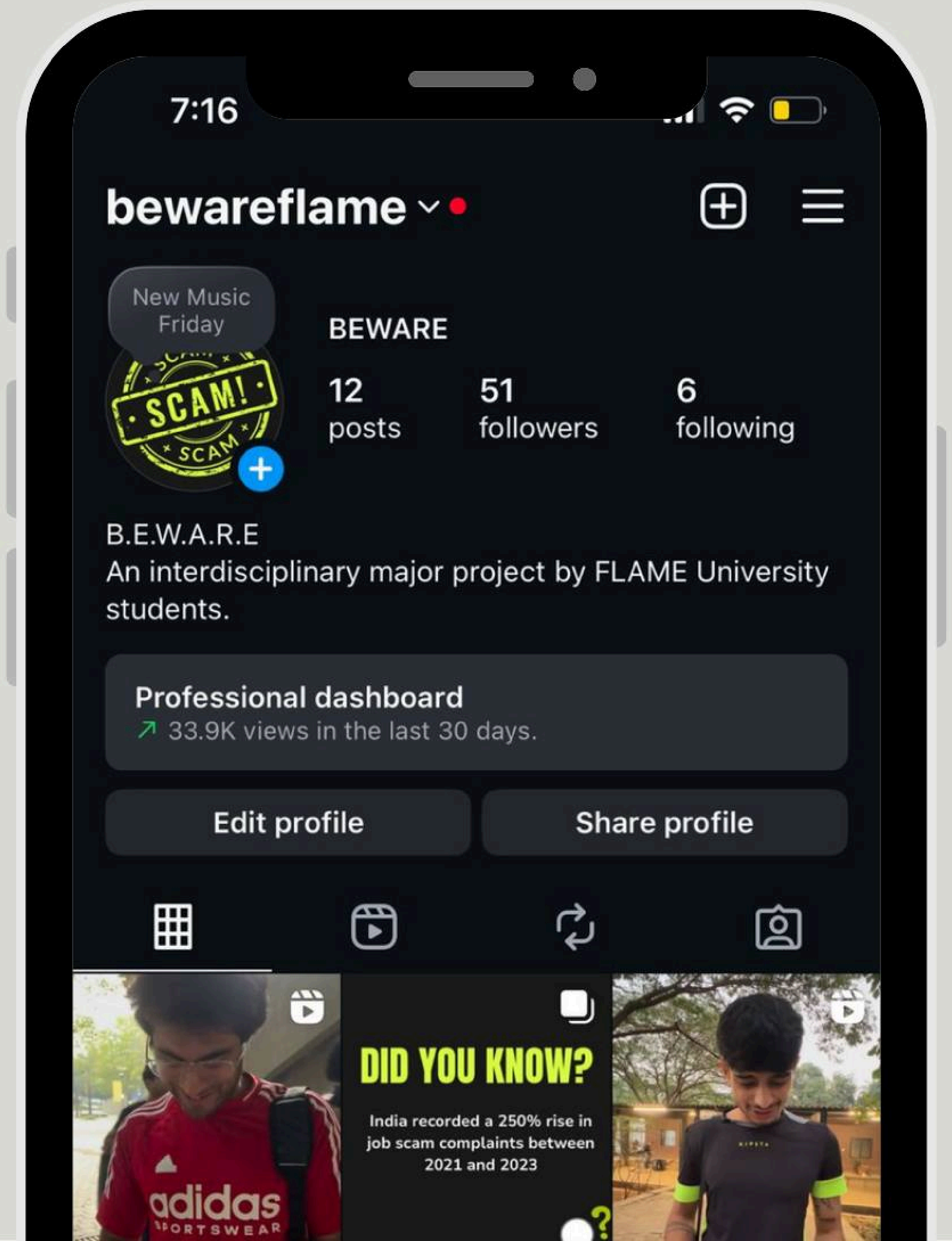
- Verified statistic-based infographics
- Informative carousel guides
- Job scam-related memes and reels
- Student testimonials
- Interactive UGC centric posts
- On-site activation content - interactive campus stall

**Primarily targeting undergraduate students who are :**

- Undergraduates & fresh graduates
- Campus job seekers
- Internship applicants

**OUR DIGITAL FOCUS**

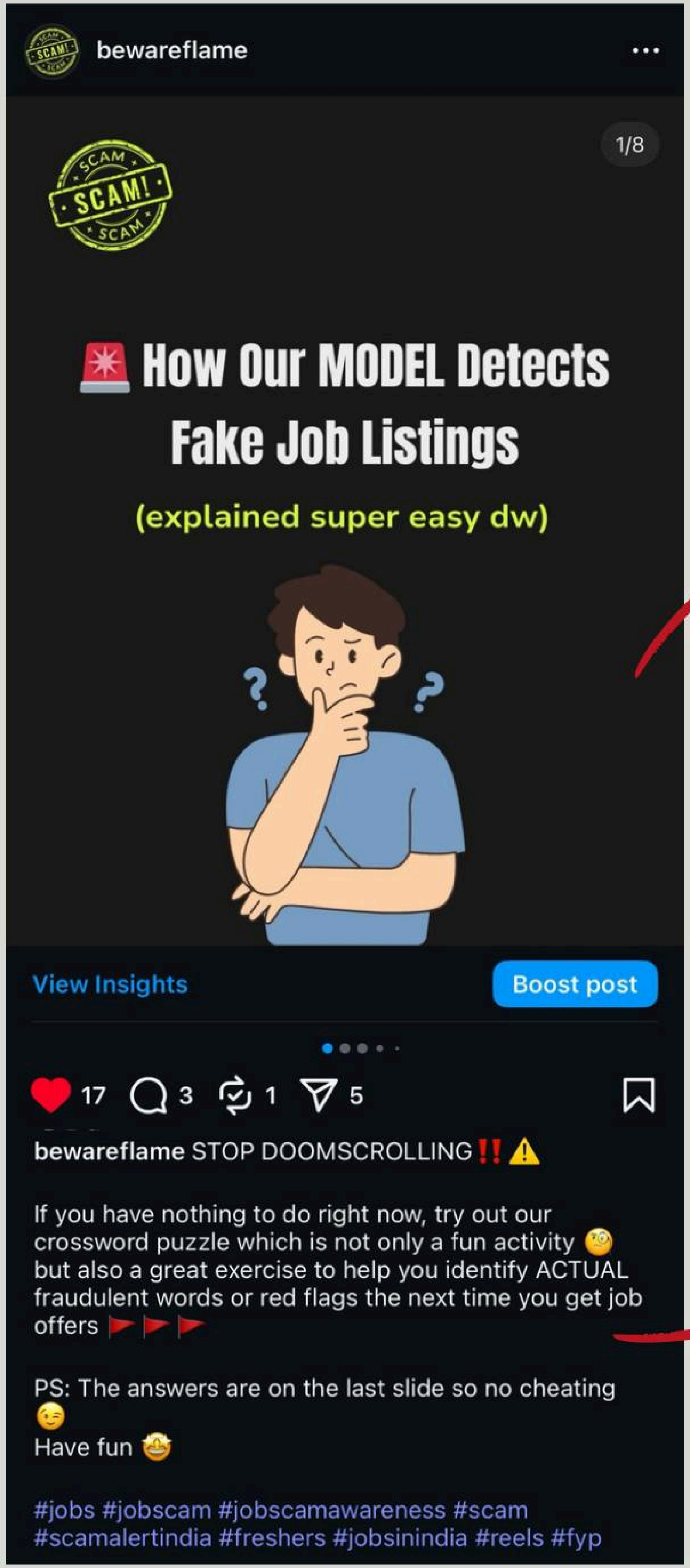
- Commonly used by the target group
- Easy tool for reach and impact measurement
- High sharability scope of content
- Compatible with various content formats



# IMPLEMENTATION

**Our process:**

- Outlined tentative content calendar
- Strategised brand identity for maximal impact and engagement
- Brainstormed for key content bucket creatives
- Planned on-site activation - logistics and design
- Gained insights and feedback from the FLAME Career Services Office (CSO)




## DESIGN CHOICES (TO REFLECT BRAND IMAGE & TONE)

HEADING FONT: **Anton**

CONTENT FONT: Nunito Sans

CAMPAIGN COLOURS

**We kept our language friendly and energetic - to resonate with our target group**

# RESULTS & EVIDENCE

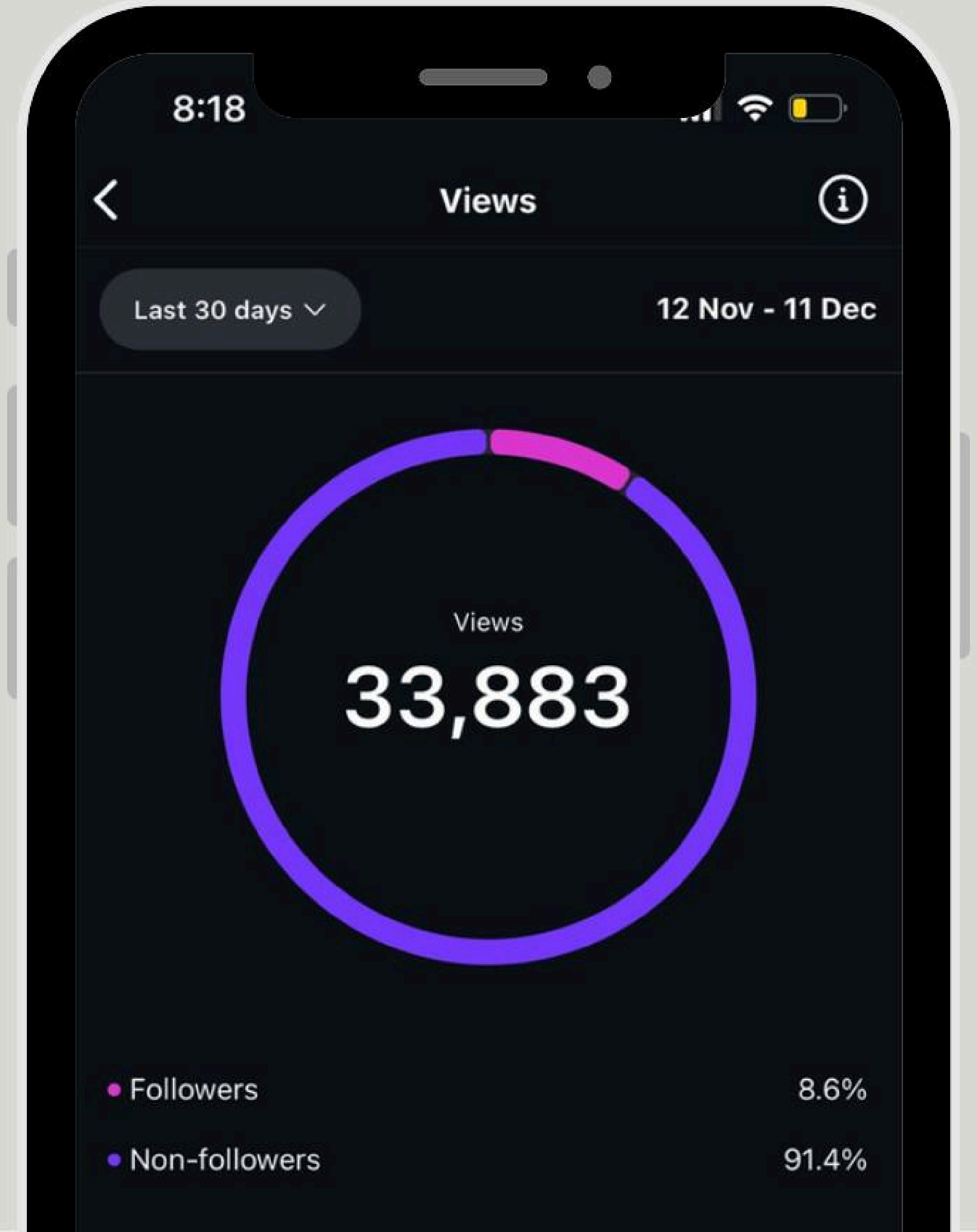
**Objectives Addressed :**
- Create accessible scam-awareness content.
- Help students make more informed job-related decisions.

**What We Did :**
- 12 total posts: carousels, reels, infographics, UGC (crossword), testimonials.
- On-ground activation: awareness STall near the mess.

**Quantitative Evidence :**
- 33,000+ views (since late November).
- 1,100+ interactions (likes, comments, shares, saves).
- 50+ followers gained within a month.

**Qualitative Evidence :**
- High-engagement reels → strong shareability + interest.
- Crossword UGC + stall conversations → active participation & discussion.
- Testimonials + reaction videos → increased awareness & intent to verify listings.

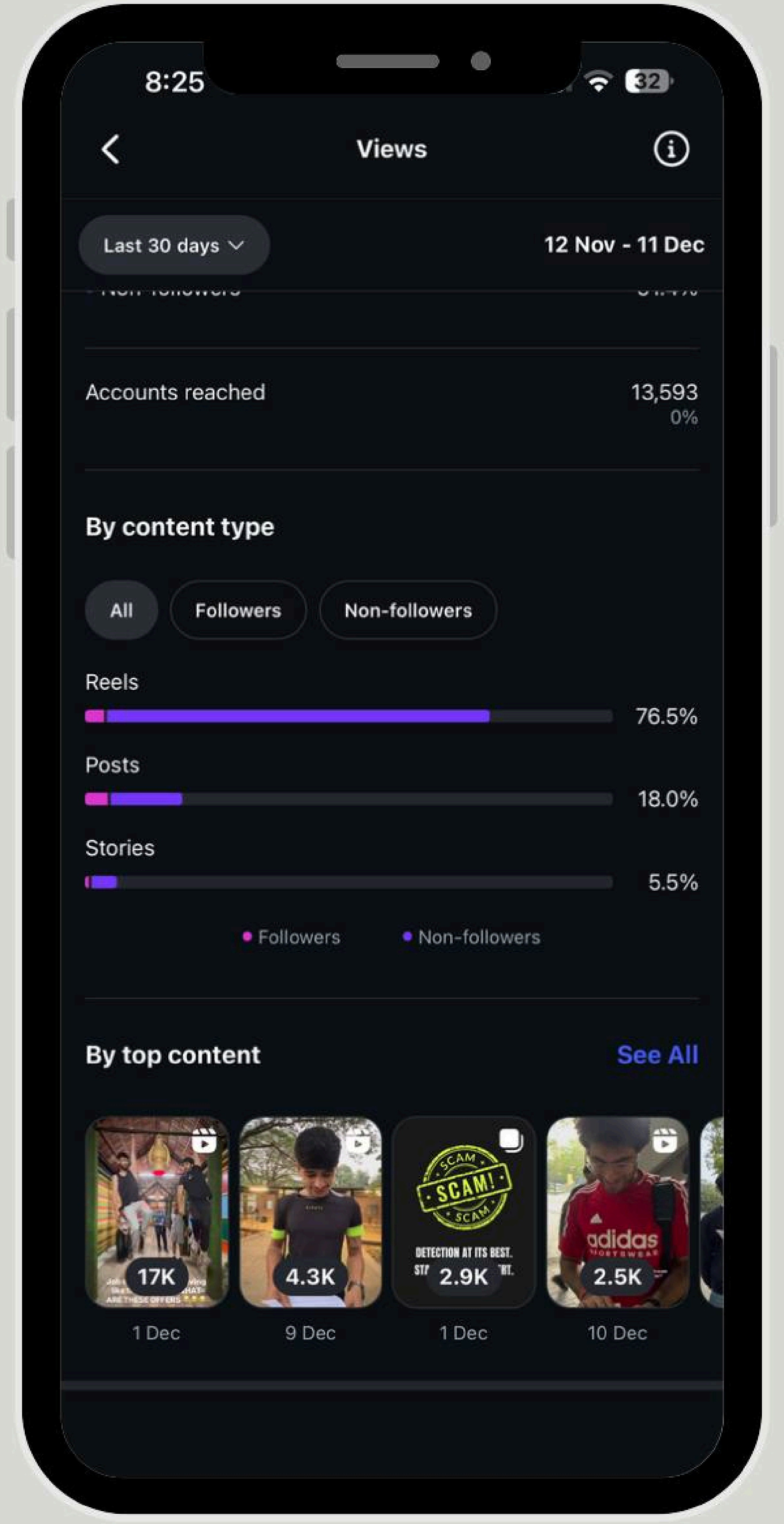
**Limitations :**
- Short campaign duration.
- Campus-centric audience.
- Small qualitative sample & self-selection bias.
- Engagement does not necessarily equal behaviour change.

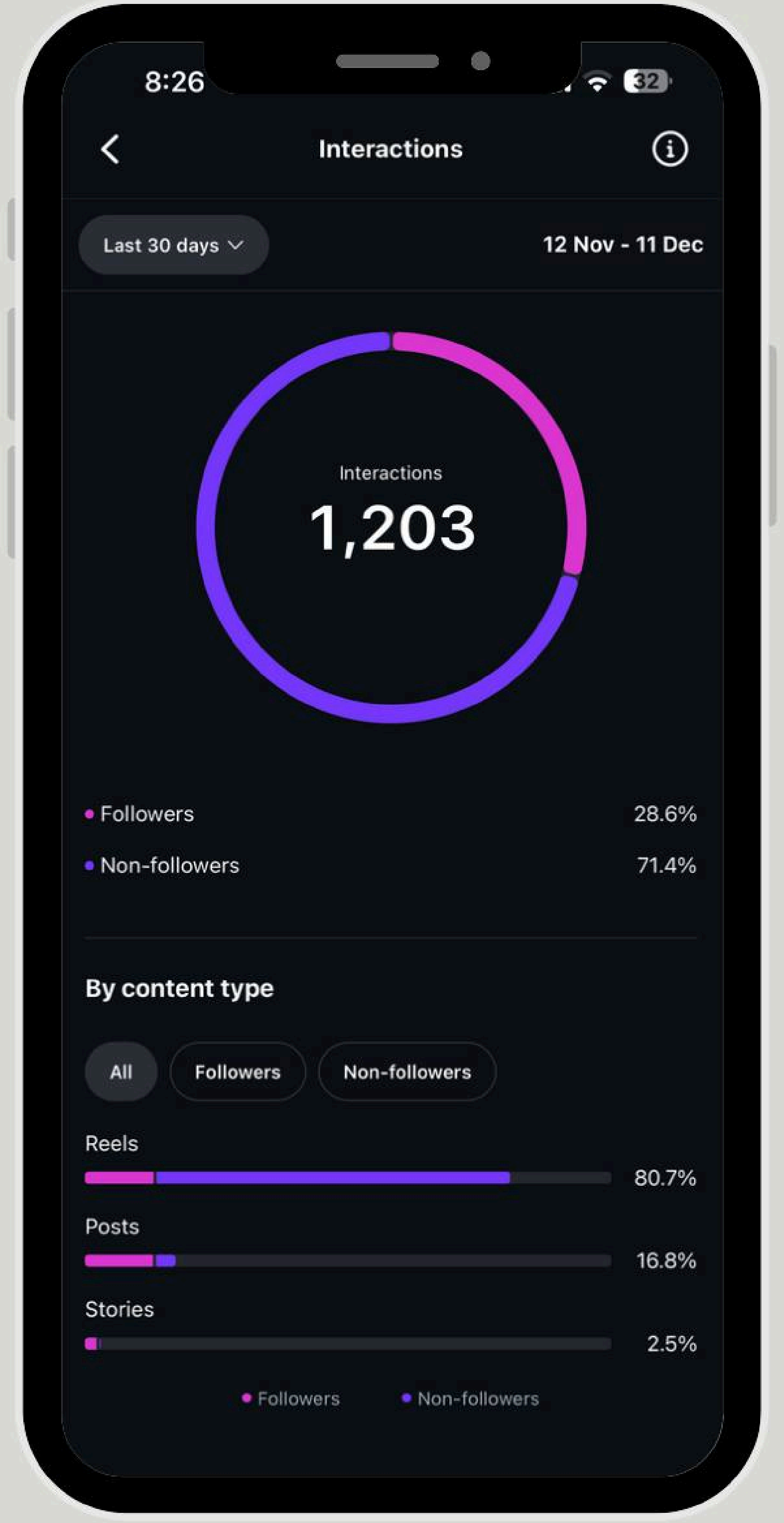


# RESULTS & EVIDENCE

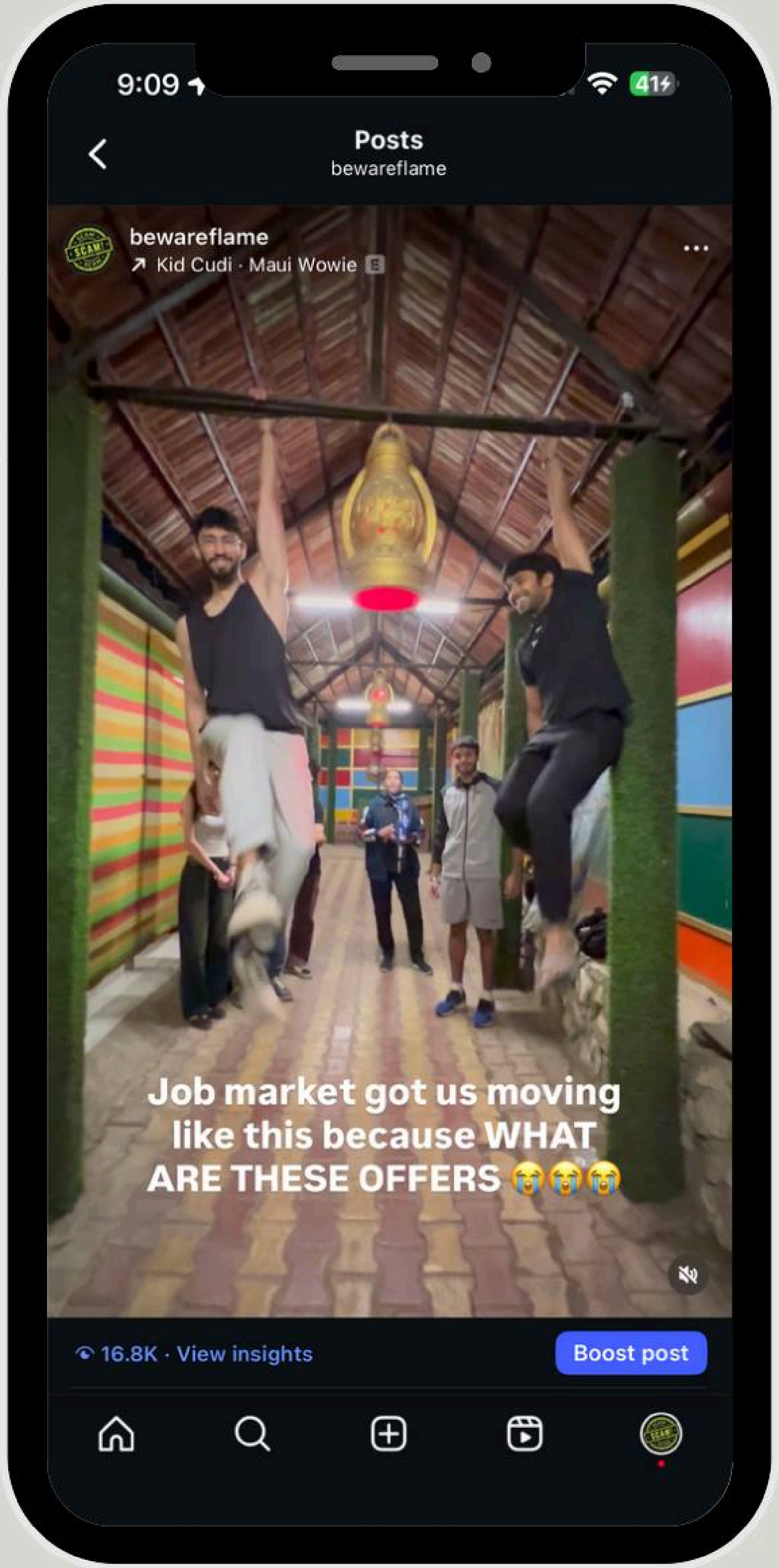
In less than a two-week window, our strategic content push delivered significant organic reach and engagement. The response validated both the need for scam-awareness resources and the effectiveness of our communication approach.

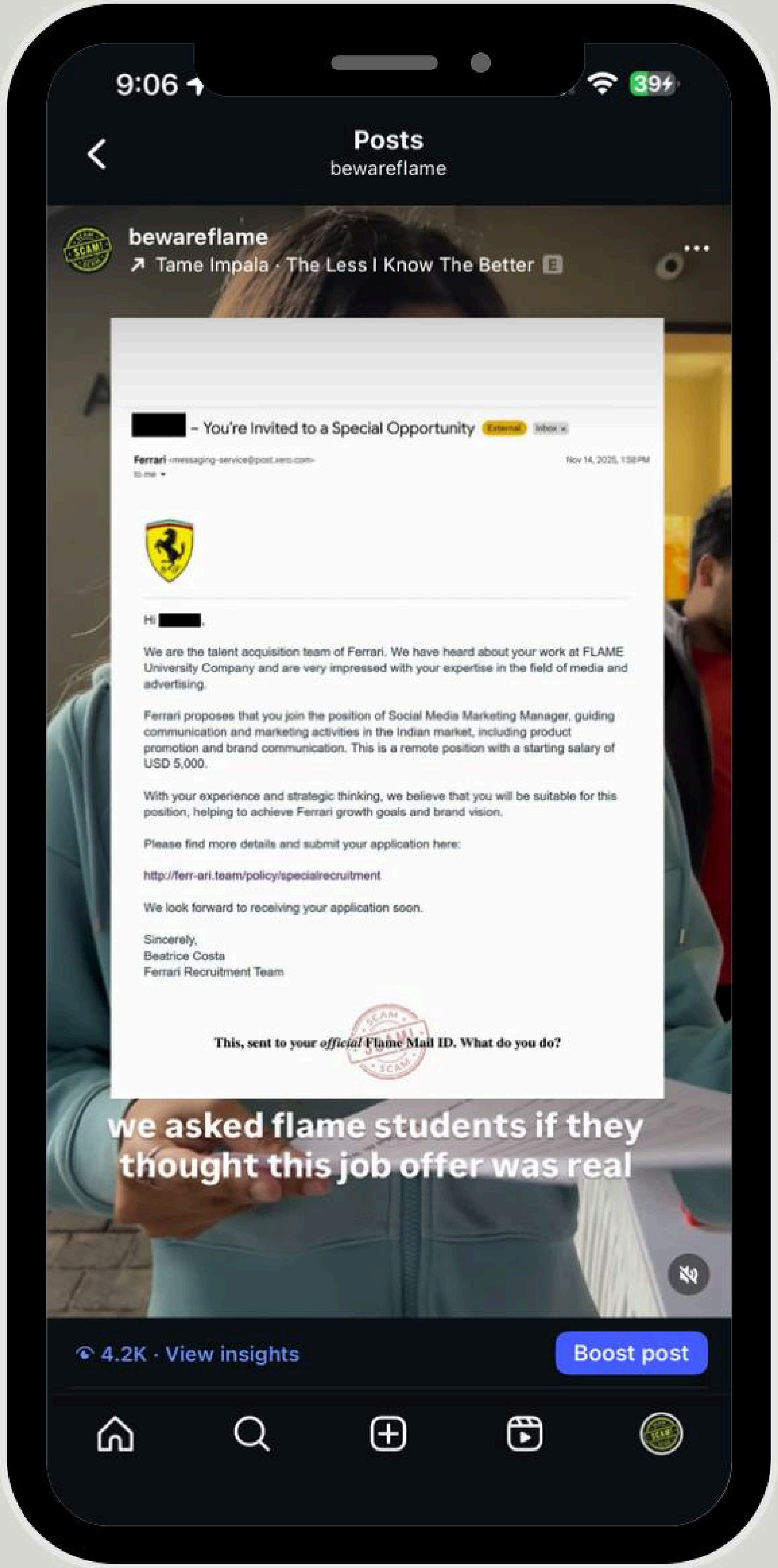
Limitation We Acknowledge: Engagement ≠ behavior change (yet to be measured long-term)

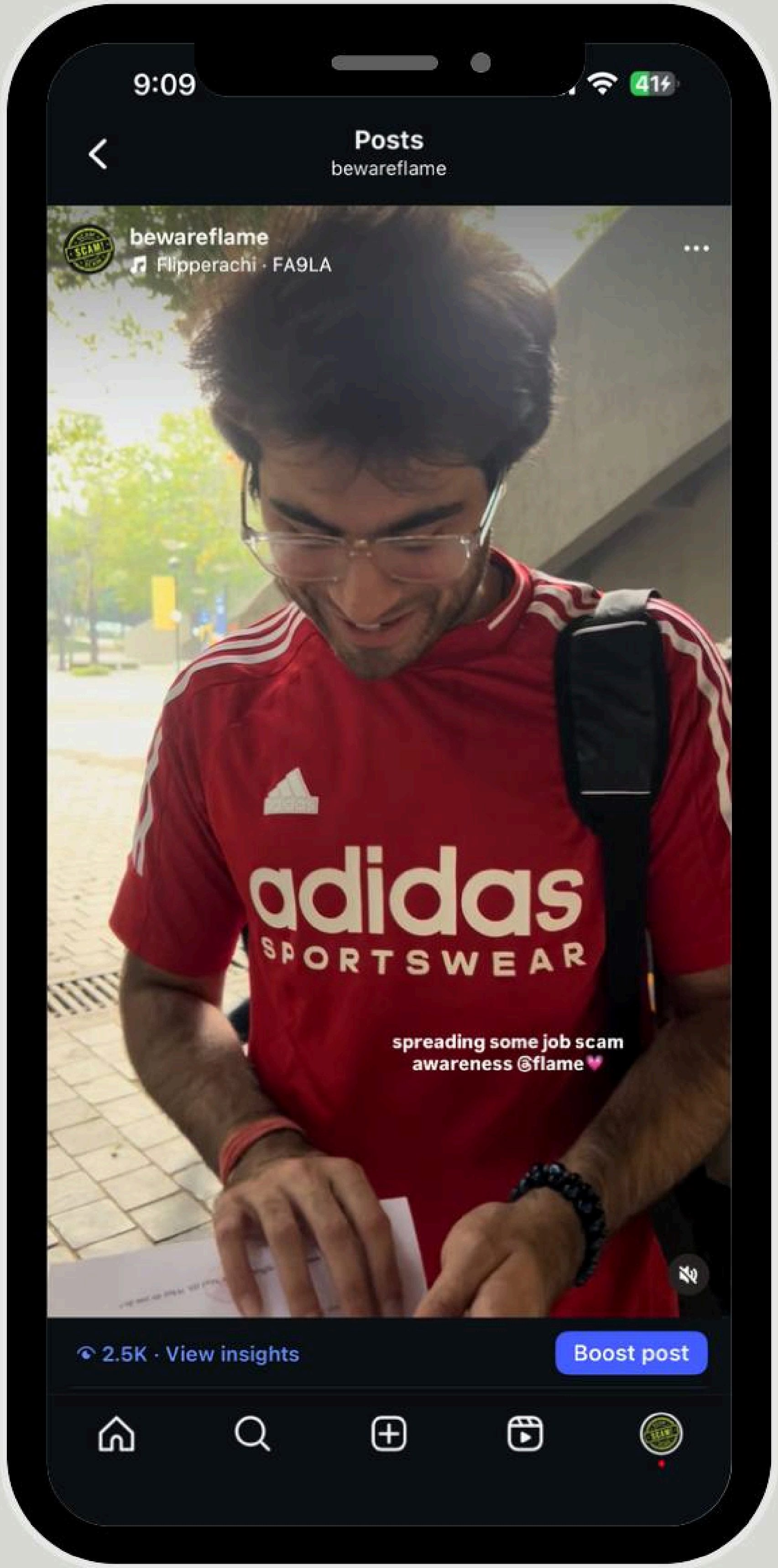


# CAMPAIGN IMAGES




IMP FINAL PRESENTATION

GROUP 35

# THANK YOU

@bewareflame on instagram

by Aaryan, Ameesha, Dhriti, Moksha, Neha, Rachna, and Shanaya

# REFERENCES :


1. Federal Trade Commission (FTC). (2022, 2024). Employment and Job Scam Reports.
2. Indian Computer Emergency Response Team (CERT-In). Job fraud cybercrime statistics.
3. Indian Ministry of Labour. (2023). Study on fraudulent job postings encountered by fresh graduates.
4. Alharby, F. (2019). An Intelligent Model for Online Recruitment Fraud Detection. Journal of Information Security. https://doi.org/10.4236/JIS.2019.103009
5. Ansar, S., & Hussain, S. (2025). Effect of Online Job Scams on Trust and Dependence on Digital Platform. https://doi.org/10.63468/jpsa.3.3.86
6. Naudé, M., Adebayo, K.J. & Nanda, R. (2023). A machine learning approach to detecting fraudulent job types. AI & Society, 38(3), 1013–1024. https://doi.org/10.1007/s00146-022-01469-0
7. Ravenelle, A., Janko, E., & Kowalski, K. C. (2022). Good jobs, scam jobs: Detecting, normalizing, and internalizing online job scams during the COVID-19 pandemic. https://doi.org/10.17615/91b3-hg17
8. Swetha, K., Reddy, M. T., Sravani, K., & Subramanyam, B. (2023). Fake Job Detection Using Machine Learning Approach. Journal of Engineering Sciences, 14(02).
9. Vidros, S., et al. (2017). EMSCAD Dataset for Online Recruitment Fraud Detection.