

Spot The Scam

Fraudulent Job Posting Detector

Alice Karakachian, David Pu, David Nguyen,
Sai Ananya Channamsetti, Lakshmi Gayathri Divakaruni



Introduction

The Problem:

- Technology has increased accessibility to information about employment opportunities
- However, with this, there has also been an increase in fraudulent job postings
 - Collecting personal information, trying to take job seekers' money through upfront fees

Our project:

- Aim: Improve the accuracy of the identification of fraudulent job postings
- Our solution: full-stack application that applies both classical and transformer models to this modern problem

The screenshot displays the 'Spot the Scam' application interface. At the top, there's a navigation bar with 'SPOT THE SCAM' and user options 'Score' and 'Review'. Below this, a 'Model Risk Dashboard' section is visible. The main content area is divided into two panels. The left panel, titled 'Score a job posting', contains a form for job details. The right panel, titled 'Model snapshot', shows performance metrics for the 'linear_svm_C1.0' model.

Score a job posting (AWAITING SUBMISSION)

Paste details from a listing. We keep the text local until you submit for scoring.

Job title *

Remote Accounts Payable Specialist (Immediate Start)

Description Requirements Company profile Benefits

Role overview

We are expanding our remote accounting pod to support new enterprise contracts. You will process invoices, maintain vendor ledgers, and support monthly close for North America clients.

Location Employment type

Remote Contract

Required experience Required education

2+ years Associate Degree

Industry Function

Accounting Finance

Model snapshot

Validation and held-out test scores from the most recent training run.

linear_svm_C1.0

classical - tfidf+tabular Test ECE: 0.006

F1 VAL: 0.811 TEST: 0.789

PRECISION VAL: 0.939 TEST: 0.92

RECALL VAL: 0.713 TEST: 0.691

ROC AUC VAL: 0.988 TEST: 0.981

PR AUC VAL: 0.872 TEST: 0.845

Brier VAL: 0.012 TEST: 0.014

GRAY-ZONE POLICY

Width 0.10

Lower bound 0.50

AI Assistant

Powered by Google Gemini

Clear History

Experience with backend frameworks (Next.js, FastAPI), databases (PostgreSQL), and deployment (Vercel, AWS) is a major plus. You enjoy working closely with designers and customers to iterate on product vision. You might have a portfolio of beautiful side projects or significant contributions to open-source projects.

Apply if you've built: Complex web apps, interactive dashboards, or have a GitHub profile full of clean, well-documented code.

11:33:33 PM

Here's a breakdown of the job posting, based on the provided pipeline analysis:

1. Fraudulent Activity Assessment:

- **Overall Assessment:** The pipeline identifies the job posting as legitimate with a low fraud probability of 1.9%. This suggests the system views the posting as genuine.
- **Reasoning:** The analysis weighs various factors, concluding that the positive signals outweigh the red flags, leading to the "legitimate" classification.

2. Key Red Flags and Positive Signals:

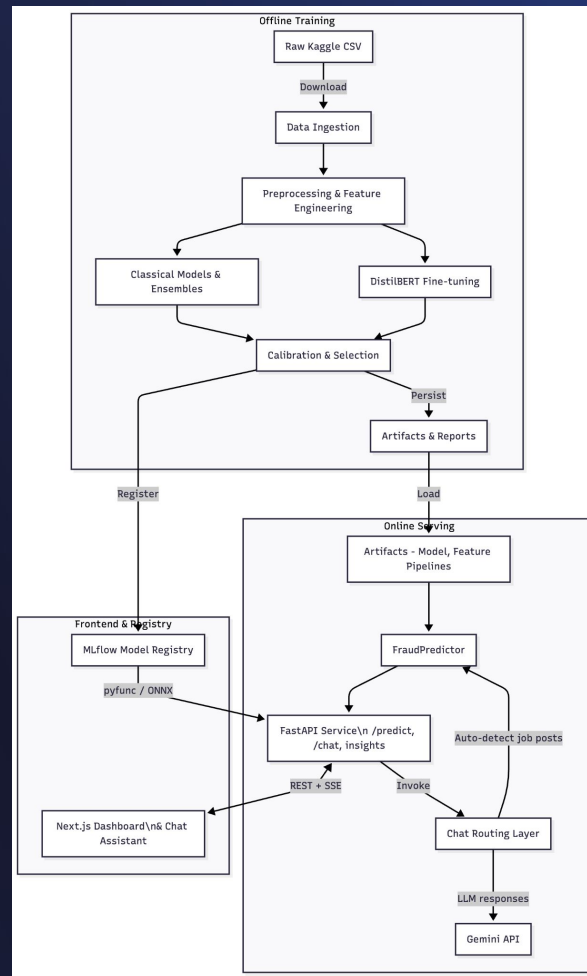
- **Red Flags (Minor):**
 - has_company_Logo

Ask a question about job fraud detection...

Press **Enter** to send, **Shift + Enter** for new line

Methods

- Kaggle job-post datasets merged. Text fields combined into a single text_all feature
- Classical models: Logistic Regression, Linear SVM, LightGBM, XGBoost
- Transformer model: fine-tuned DistilBERT on text_all
- All candidates tuned and ranked by validation F1; best calibrated model evaluated on test
- Gemini API chatbot wraps the model using conversation history, job fields, and model outputs



Results

Table 1: Performance of `ensemble_top3` on validation and test splits.

Split	F1	Precision	Recall	ROC AUC	PR AUC	Brier
Validation	0.8561	0.9297	0.7933	0.9890	0.9053	0.0103
Test	0.7721	0.8537	0.7047	0.9863	0.8659	0.0143

- Validation winner: calibrated classical ensemble `ensemble_top3`, selected over transformer baselines by validation F1
- Generalizes well: strong F1 on validation and test with high ROC/PR AUC and low Brier (see table)

References

- [1] Sultana Umme Habiba, Md Khairul Islam, and Farzana Tasnim. A comparative study on fake job post prediction using different data mining techniques. In *2021 2nd international conference on robotics, electrical and signal processing techniques (ICREST)*, pages 543–546. IEEE, 2021.
- [2] Vijay Madaan, Neha Sharma, Raghubeer Singh Bangari, and Srinivas Aluvala. Fraudulent job posting detection using logistic regression. In *2024 International Conference on Information Science and Communications Technologies (ICISCT)*, pages 1–6. IEEE, 2024.
- [3] Gayathri Malaichamy, Cristina Hava Muntean, and Anderson Augusto Simiscuka. Online job posting authenticity prediction with machine and deep learning: Performance comparison between n-gram and tf-idf. In *International Conference on Deep Learning Theory and Applications*, pages 143–162. Springer, 2024.
- [4] K Ramakrishna Reddy, G Indrani, N Pavan Kumar, and K Vamshi Krishna. Fake job posting detection using machine learning algorithms. In *2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pages 822–827. IEEE, 2025.
- [5] K. Suparna and R. Anil Kumar. A comparative study on fake job post prediction using different datamining techniques. *International Journal of Information Technology and Computer Engineering*, 12(3):317–323, 2024

Thanks !



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)