AI-Powered Fake Job Posting Detection

Visesh Bentula

Adilakshmi Chimakurthi

Societal Problem Being Addressed

Job seekers increasingly rely on online platforms such as job portals and social media to find employment. However, these platforms are also exploited by scammers who post fraudulent job listings to extract personal information, financial details, or scam money from vulnerable applicants. These fake postings disproportionately affect recent graduates, unemployed individuals, and those with limited digital literacy.

The societal implications include identity theft, emotional harm, economic losses, and erosion of trust in online employment platforms. There is a strong need for an automated, intelligent system that can help identify potentially fraudulent job postings and protect users—especially those in low-resource or high-risk situations.

Proposed AI-Driven Approach

We propose building a machine learning model to automatically classify job postings as **real** or **fake**. The model will analyze the features including **job company name**, **job description**, and other metadata to identify fraudulent listings.

A key addition to our approach is the **use of company names** as a strong signal for classification. By analyzing patterns across company names (e.g., frequent occurrences in fake listings, missing or suspicious naming patterns), we aim to improve the model's ability to flag potentially

fraudulent postings. For example, fake listings may use generic or misleading company names like “Global Solutions Pvt” or lack any verifiable web presence.

Data Availability

We will use the **Fake Job Postings** dataset from Kaggle, which contains ~10,000 entries with the following relevant fields:

- title, company_name, description, requirements, benefits, company_profile
- telecommuting, salary_range, has_company_logo, employment_type
- Labeled column: fraudulent (1 = fake)

This dataset not provided sufficient variety and depth to train models using both **textual** and **structured features**.

Dataset from Kaggle: https://www.kaggle.com/datasets/srisaisuhassanisetty/fake-job-postings?resource=download

Technical Methods to Be Used

- **NLP techniques** for analyzing title, description, company_profile
- **Feature engineering** for company_name patterns (e.g., word frequency, token patterns, length, known brand list)

  Existing model

- **Supervised Learning Models**: Logistic Regression, Random Forest with the description column

New model:

- **Supervised Learning Models**: Logistic Regression, Random Forest with the company name column
- **Imbalance Handling**: Synthetic data generation

Anticipated Impact & Ethical Considerations

- **Impact**: Help job seekers avoid scams, reduce harm to vulnerable groups, and enhance trust in job platforms
- **Ethical Concerns**:
    - Avoid false positives that may harm small or new companies