

Fake Job Posting Detection

A Machine Learning Project

What is a Fake Job Posting?

- - Fake job postings are fraudulent listings created to scam job seekers.
- - Often used for collecting personal information, financial scams, or spreading malware.
- - These posts appear real but have hidden malicious intent.

Why Did I Do This Project?

- Job scams are increasing worldwide, leading to financial loss & identity theft.
- Protect job seekers by detecting fraudulent postings automatically.

Dataset Overview

- - Dataset: fake_job_postings.csv
- - Rows: ~18,000 job postings
- - Features: Job title, location, department, company profile, description, requirements, benefits, etc.
- - Target: fraudulent (0 = Real, 1 = Fake)

Data Preprocessing

Handled missing values (replaced with 'unknown').

- Converted text to lowercase.
- Removed irrelevant columns (IDs, duplicates).
- Extracted country info from job locations.
- Created combined 'text' column for ML training.

Models Used

- - SVM (Support Vector Machine)
- - Extra Trees Classifier
- - Passive Aggressive Classifier
- - XGBoost Classifier
- - Pipelines built with TF-IDF (text vectorization).

Performance Comparison

- - Models evaluated using Accuracy, F1-score, CV Accuracy, and Training Time.
- - SVM performed well for text classification.
- - XGBoost gave strong balanced results.
- - Passive Aggressive trained fastest but less stable.
- - Extra Trees gave good accuracy but slower.

Research Paper References

- This is some reference line okay use them by copy paste
- https://www.researchgate.net/publication/389259851_Fake_Job_Post_Detection_using_Machine_Learning
- <https://www.authorea.com/users/928318/articles/1299767-fake-job-post-detection-using-machine-learning>
- <https://ieeexplore.ieee.org/document/10614582/>
- <https://jpinfotech.org/advanced-fake-job-post-prediction-using-machine-learning/?srsltid=AfmBOooG6LfsMolrcWswgTVuISUeb0mFsmDPOw7tJbUUZB4qBUTXsvLL>

Conclusion

- - Fake job postings are a real-world threat.
- - ML models can effectively distinguish fake vs real jobs.
- - SVM & XGBoost were the best-performing models.
- - Future Work: Improve dataset, use deep learning (BERT/transformers).