Team -7 : Fake Job Post Detection

Team Members:

1. Pedireddi Siri Satya (21B21A4433)
2. Nekkanti Yajna Sri Sahithi (21B21A4409)
3. Eerella Usha Prasuna (21B21A4423)
4. Nagam Manisha (21B21A4428)
5. Chintapalli Vyshnavi Priya (21B21A4436)

Why we choose this project:

1. Increasing Job Scams:

- Many job seekers fall victim to fraudulent job postings, losing money, time, and sometimes personal data.
- Scammers use fake job postings to steal identities, trick applicants into paying fees, or conduct phishing attacks.

2. Automating Fraud Detection:

- Traditional manual screening of job posts is inefficient and prone to errors.
- NLP (Natural Language Processing) can help automate the detection of fake job posts based on textual patterns, linguistic features, and semantic analysis.

3. Enhancing Job Portals and Security:

- Job portals like LinkedIn, Indeed, and Glassdoor can integrate NLP-based fake job detection to filter out scams.
- Employers and job seekers will benefit from a safer hiring environment.

Abstract:

The increasing number of fraudulent job postings on online job portals has become a serious concern, leading to financial losses and identity theft among job seekers. This project, Fake Job Post Detection Using NLP, aims to develop a machine learning-based model that can automatically identify fake job listings by analyzing their textual content.

Using Natural Language Processing (NLP) techniques, the system extracts key features from job descriptions, such as job titles, company names, required qualifications, and job responsibilities. These features are then analyzed to detect patterns commonly associated with fraudulent job postings, such as vague job descriptions, unrealistic salary offers, and requests for sensitive information.

The model is trained using a labeled dataset containing real and fake job postings. Machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Deep Learning techniques are employed to classify job postings as real or fake. The project also incorporates data preprocessing steps like tokenization, stop-word removal, and vectorization to enhance the accuracy of the classification model.

Technologies:

This project involves various technologies and tools for data preprocessing, machine learning, and NLP-based text analysis. The key technologies used are:

1. Programming Language:
Python – Used for data processing, model training, and implementation due to its rich ecosystem of NLP and machine learning libraries.

2. Natural Language Processing (NLP) Libraries:
NLTK (Natural Language Toolkit) – Used for text preprocessing, tokenization, stop-word removal, and stemming/lemmatization.
spaCy – Alternative NLP library for efficient text processing.
TextBlob – Used for sentiment analysis and linguistic processing.

3. Data Processing & Feature Engineering:
Pandas – Used for handling and processing structured job post data.
NumPy – Used for numerical computations.

Process of the project:

Process of Fake Job Post Detection Using NLP
1. Data Collection
2. Data Preprocessing
3. Feature Engineering & Text Representation
4. Model Selection & Training
5. Model Evaluation
6. Deployment & Real-Time Detection
7. Conclusion