

HW2:Predicting Fake Job Posting

*Antara Tewary (G01413546), Ankit Kumar (G01436204)***1. Introduction**

This assignment aims to create a system that flags suspicious job postings on Indeed.com, addressing the challenge of fraudulent postings that mislead the efforts of job seekers. We are using Fake Job Prediction Dataset, which contains 18000 job descriptions out of which 800 are fraudulent samples. The goal is to create an automated model which prioritizes which posting to review based on the likelihood of it being fraudulent.

2. Data Preparation

- *Label Encoding and Filtering Invalid Data:* The target column fraudulent was cast to integer type to represent legitimate (0) and fake (1) job postings. Any records with invalid values were filtered out.
- *Handling Missing Values:*