

About Dataset

[Real or Fake]: Fake Job Description Prediction

This dataset contains 18K job descriptions out of which about 800 are fake. The data consists of both textual information and meta-information about the jobs. The dataset can be used to create classification models which can learn the job descriptions which are fraudulent.

Task:

1. Create a classification model that uses text data features and meta-features and predict which job description are fraudulent or real.
2. Identify key traits/features (words, entities, phrases) of job descriptions which are fraudulent in nature.
3. Run a contextual embedding model to identify the most similar job descriptions.
4. Perform Exploratory Data Analysis on the dataset to identify interesting insights from this dataset.