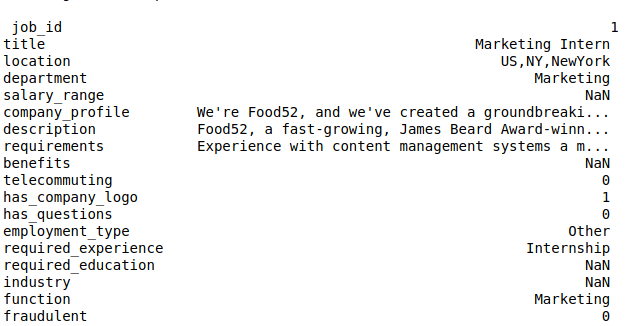
COMS3007: Machine Learning Assignment 2020

Group Members

|  |  |
| --- | --- |
| Bakang Majova |  |
| Maccauley Marota |  |
| Thabo Rachidi | 1632496 |

1. Our dataset has just under 18 000 job descriptions and around 800 of them are fraudulent. The data set contains 13 string/text columns, 4 boolean(0/1) columns and an ID column. We are using this dataset to create classification models that can learn whether a job description is real or fraudulent.

|  |  |  |
| --- | --- | --- |
| Attribute | Description | Type |
| Title | The title of the job posting | String/Text |
| Location | Geographical location of the job posting. | String/Text |
| Department | Corporate department (e.g. Sales). | String/Text |
| Salary range | Indicative salary range. | String/Text |
| Company profile | A brief company description. | String/Text |
| Description | The detailed description of the job posting. | String/Text |
| Requirements | Enlisted requirements for the job opening | String/Text |
| Benefits | Enlisted offered benefits by the employer | String/Text |
| Telecommuting | True for telecommuting positions | Boolean |
| Company logo | True if company logo is present | Boolean |
| Questions | True if screening questions are present | Boolean |
| Employment type | The type of employment for the job posting (Full-type, Part-time, Contract, etc.) | String/Text |
| Required Experience | The experience required for the job posting (Executive, Entry level, Intern, etc.) | String/Text |
| Required Education | The education required for the job posting (Doctorate, Master’s Degree, Bachelor, etc.) | String/Text |
| Industry | The industry that the job posting belongs to (Automotive, IT, Health care, Real estate, etc.) | String/Text |
| Function | The function the job posting belongs to (Consulting, Engineering, Research, Sales etc.) | String/Text |
| Fraudulent | Classification attribute | Boolean |

  
Sample Data point