**Finding the best ML Engineer**

In this use case, we would like to create the best algorithm to scan a Resume, analyze it and determine if it is a good match to a specific job description.

The exercise will focus on finding the best Machine Learning Engineer and you will have to train a model that will give a score to each candidate that applied for the job.

### Data

The data you are given is the following:

* *Job description.pdf*: Description of Job role, Preferred Skills, Educations.
* *train.csv:* List of candidate file name and corresponding scores
* *test.csv*: List of candidate file name
* *trainResumes/.pdf*: List of resume to be used for the training based of File Names
* *trainResumes2/.pdf*: List of resume to be added to the training set
* *testResumes/.pdf*: List of resume to be scored based on File Names

### Guidance:

### What type of information is relevant in the CVs to match the job description?

### Can you summarize through a visualization the type of information that is relevant to the job description (e.g. key words with high impact, key words with low impact)?

### What feature can you create from the CV to build your model?

### Is this a classification model or a regression model? What models/algorithms will be suitable for such task?

### How can you use embedding for this use case?

### Scoring:

### You need to score the 20 CVs in the test folder and provide the TOP 3 CVs that match the job description along with the scores.