



Take-home assignment: Machine Learning Engineer

Problem statement:

You are given a dataset of 20 sample resumes with the personal details masked out. Your job is to come up with a hypothesis for a generalised technique to extract certain information from all of them. The information which is to be extracted is stated below.

You can use an ensemble of machine learning algorithms and other techniques which you think can help in making the predictions better. Your focus should be to come up with an algorithm/technique which can generalise for most of the resumes. You can take the help of any AWS/GCP services in your approach.

What to extract:

1. Names of organizations they have worked in and the years they were working in them. Certain resumes might not have this information, so you can ignore them.
2. Names of institutions and colleges they have studied in and the associated years.

Details for Submission:

You can submit your idea/proposal in a doc format. You can also provide a POC (proof-of-concept) if required in a jupyter notebook.

Dataset:

[ML Assignment test dataset](#)