**Machine Learning Engineer Task: Resume Categorization**

**Following these steps for working:**

In this study, we analyze a resume dataset to categorize all resumes with respective categories.

First of all, we create a Jupyter Notebook and split the given CSV file into two CSV such as (*resume\_data* and *resume\_test\_data*).

After Splitting we use resume data for Exploratory Data Analysis and Training the various models such as the Machine learning model (Random Forest Classifier, Logistic Regression, K-Nearest Neighbor and Support Vector Machine) and Deep Learning Approach (ANN and LSTM) via Jupyter Notebook to get comparison best model for our testing dataset.

After doing my First step, when I got my best classifier model. Then I convert this trained model into a “*best\_clf.pkl*” pickle file. and also convert our vectorization model “*tfidf.pkl*” which is indeed our testing data.

Thirdly, I use Script to collect all test pdf that is mentioned in “*resume\_test\_data*” into a “*test\_data*” folder. so that when I run my *script.py* are automatically categorize all pdf into respective folders.

Now the final task is to create a “*script.py*” by using pickle files which automatically scraps all pdf in our “*test\_data*” folder and categorizes it with the respective folder name.

Now if you run the script.py and put it into the path where are “*test\_data*” folder (test resume) is located. then this script.py categorizes all pdf corresponding to their categories names.

**Summarized Instructions to run this full project:**

1. Run Notebook and collect pickle files and *resume\_test\_data* (*test\_resume\_data.csv*).
2. Collection all pdf into a single “*test\_data”* folder with respect to the ID of *test\_resume\_data.csv*
3. Go to any code editor (VS code) create a new folder, put pickle files and script.py files
4. Run script.py and put the input path of the *test\_data* folder.
5. After completing you get a categorized folder wise all pdf files.

Alhamdulillah, it was a nice journey …