**Date:25-feb-2024**

**Task name: Resume classification**

You have this dataset:

<https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset>

**Problem Statement: text classification**

## **About Dataset**

### **Context**

A collection of Resume Examples taken from livecareer.com for categorizing a given resume into any of the labels defined in the dataset.

### **Content**

Contains 2400+ Resumes in string as well as PDF format.  
PDF stored in the data folder differentiated into their respective labels as folders with each resume residing inside the folder in pdf form with filename as the id defined in the csv.

**Inside the CSV:**

* **ID:** Unique identifier and file name for the respective pdf.
* **Resume\_str:** Contains the resume text only in string format.
* **Resume\_html:** Contains the resume data in html format as present while web scrapping.
* **Category:** Category of the job the resume was used to apply.

**Present categories are:**

HR, Designer, Information-Technology, Teacher, Advocate, Business-Development, Healthcare, Fitness, Agriculture, BPO, Sales, Consultant, Digital-Media, Automobile, Chef, Finance, Apparel, Engineering, Accountant, Construction, Public-Relations, Banking, Arts, Aviation

**Your Task:**

1. **You have to convert this dataset into Huggingface data format**
2. **After converting this dataset Push it to the HF Hub**
3. **Feel free to use any LLM from HF & fine-tune a Text Classification Model on top of this dataset (use the dataset from your HUB), which will able to classify these resume categories**

**Submission Process:**

Use Google Colab for this task

till friday Before 5:00

Submit by this form: <https://forms.gle/MDKrRrZ8iDbeW88z7>

**Date:10-March-2024**

**Task name: Try Redis Vector DB**