P-380: Project – Resume Classification

**Business objective- The document classification solution should significantly reduce the manual human effort in the HRM. It should achieve a higher level of accuracy and automation with minimal human intervention**

**Sample Data Set Details: Resumes and financial documents**

**Acceptance criteria:**

To build the best model which gives the maximum performance, and need to deploy the model with either Flask/ stream lit

**Milestones: Phase 1**

| **Milestone** | **Duration** | **Task start - End Date** |
| --- | --- | --- |
| Kick off and Business Objective discussion | 1 day | 20 – Mar – 2024 |
| Data set Details | 2 days | 22 – Mar – 2024 |
| EDA | 1 Weeks – 1 ½ week | 26 – Mar -2024 |
| Model Building | 1 Week – 1 ½ week | 2 – Apr - 2024 |
| Model Evaluation | 1 week | 9 – Apr - 2024 |
| Feedback | 9 – Apr - 2024 |
| Deployment | 1 Week | 16 – Apr - 2024 |
| Final presentation | 1 day | 16 – April - 2024 |

Protocols:

1. All participants should add here to agreed timelines and timelines will not be extended
2. All the documentation – Final presentation and python code to be submitted before the final presentation day
3. All the participants must attend review meetings

Agenda:

* + - 1. Pull all the resumes into a single data frame
      2. Create a new column call category pointing towards their work category
      3. Label encode that category
      4. Perform text preprocessing on the “Resumes” column as illustrated in the image below
      5. EDA