**PRASHANT \_ROBOFLOW PROJECT**

* Login into Roboflow Environment and Prepare the project as per use-case.
* Prepare your dataset (image data) for training purposes
* Roboflow asks for training the images (i.e. Draw bounding boxes on your dataset)
* Generate the API KEY

**CODE INFO**

* Labels are created (headers, paragraphs, tables, etc.)
* Converting pdf into jpg for prediction
* Pdf images are stored in the working environment as jpg or png
* Raw images stored into *segmentation\_run\_data\rfx\_pdf\_folder*
* Predictions images stored in *segmentation\_run\_data\runs*
* Predictions images having bounding boxes on them and that bounding box labels (coordinates) are stored in the same location as the *Labels* folder
* Sending those prediction images and labels to cv2 for cropping the images as per the bounding box.
* Then using tesseract to extract the data from images and tag it using lab\_dict.
* Return the pages in JSON format.