If any of the code is not runnable. There are separate home works of each task and separate projects with dependencies in the google drive link below:

<https://drive.google.com/open?id=1_rQ6LnD7BDD5TmH3VLCRdjca-WADoIs3>

This folder contains 4 python files.

1)Data Collation(collection\_Recall.py) :

All the queries we used to collect the data and API Recall and precisions are calculated in this one.

Run it as it is

2)Boxplots.py:

This file contains all the diagrams, boxplots and Histograms related to project.

3)SVM\_LG.py

This file contains logisticregression, linear-svm and non-linear svm code.

You can comment any two from main method before running.

Linear SVM will take around 10 mins to run.

4)kmeans.py:

This file is related to kmean clustering.

You can run them as it is. No need to change anything it will generate 3 files in cluster\_text\_project folder and then you can create word cloud from those online.

The project data is in “Final\_Project.txt” file in Project\_Data folder.