Set up github

1. New environment
2. Project structure (setup.py)
3. Requirements.txt

Create an environment to store all the packages you install(**venv**), after creating it activate it

Clone the git hub repo and sync it

Follow the step in github:

echo "# mlproject" >> README.md

git init

git add README.md

git commit -m "first commit"

git branch -M main

git remote add origin https://github.com/ArjunT254/mlproject.git

git push -u origin main

**Git ignore**- some of the unwanted files wont get committed

**Requirements.txt**- This file will contains name of all the packages that I need to install while implementing the project

**Setup.py** – responsible for creating ml application as a package and which allows us to deploy it as a module

**\_\_init\_\_.py** - this file ensures that the src folder is found as a package, so that you can import it

**Find packages()** inside setup.py will check how many folders have \_\_init\_\_ inside it and consider it as a package and build it, once it is build you can import it.

-e. in requirement.txt will make setup run automatically

**Components**

* Data Ingestion: Read data from database/ file location. This is basically a py file which contains code to read data. Divide the data into train and test, create a validation set
* Data transformation: A py file which contains code for data transformation, ex- converting categorical data to numerical data, one hot encoding, label encoding
* Model trainer: py codes to train the model, confusion matrix, r2 ad value , push model pickle file into cloud

**Pipeline**

Train\_pipeline : This will call the previously mentioned components to train the model

Predict\_pipline: To predict output for the new data

**Src**

Logger- logging

Utils- Functionalities which are used in entire application, ex-save model in cloud, read dataset from db, mongo db client

Exception – exception handling