End to End Machine Learning project

## Tutorial 1: GitHub & Code setup

* Steup a new environment
* Setup.py
* Requirements.txt
* Steps:

1. Create a new repository “mlproject”
2. Create a new folder in which entire project will be developed
3. Copy the path of the folder & open anaconda prompt
4. Code . this will open vscode
5. First we need to make sure that we are in sync with the GitHub repository
6. Open new terminal in vs code and create new environment

(select cmd in terminal and use below command)

conda create -p venv python==3.8 -y

activate the virtual environment

conda activate venv/

1. Clone the repository and sync with GitHub

* Initialize git repository (git init)
* Add README.md file, before adding create README.md file (either in git or in vs code) (in README.md file we can write descriptions, what are all the steps required)
* Add README.md file in GitHub repository

git add README.md

commit the file,

git commit -m "First commit"

* Push the file to git repository

git branch – M main

* Add origin (so it is in sync with the GitHub repository)

git remote add origin <https://github.com/AD/mlproject.git>

git remote – v

to push the data into github repo we need to set git global

(check git global config and add username and email of your own)

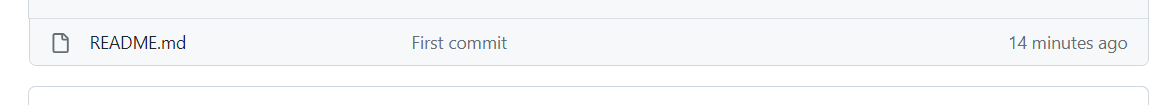
git config --global user.name "John Doe"

git config --global user.email johndoe@example.com

* Push the data

git push -u origin main

by this now we can see the file in GitHub repository.



* Create a new .gitignore file in GitHub repository, select python and commit changes

git pull   
  
so all the updation will be done in vscode as well.

* Setup a setup.py and requirements.txt

(setup.py will be responsible for making or ML application as a package, which can be used further.)

from setuptools import find\_packages , setup #find all the packages

setup (

    name = "mlproject",

    version = "0.0.1",

    author= "yourname",

    author\_email= "youremail id",

    packages=find\_packages()

    install\_requires = ['pandas', 'numpy', 'seaborn '] #automatic installation can be done

)

* Create a new folder “src” in vs code (source) and inside it create \_\_init\_\_.py
* When in setup.py find\_packages() is running it will go and find in how many files have \_\_init\_\_.py running. Considering source as a package , it will try to build it.

1. We create a function , when we need to install many packages,

get\_requirements () #which will take requirements.txt it should be able to read all those files.

def get\_requirements (file\_path:str)-> List[str]:

    '''

    this function will return the list of requirements

    '''

    requirements = []

    with open (file\_path) as file\_obj:

        requirements = file\_obj.readlines()  #the line will get readed, but \n will get added

        requirements = [ req.replace("\n","") for req in requirements.txt]  #replace \n with blank

A screenshot of a computer

Description automatically generated

Here -e. will automatically trigger setup.py

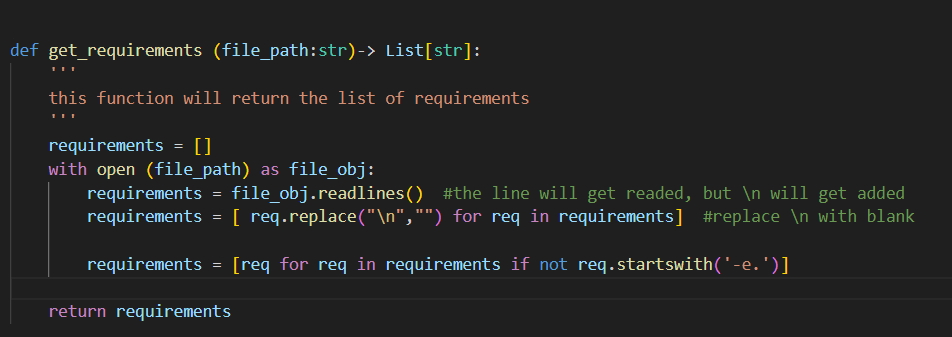
When we install requirements.txt the setup.py will also run.

A black background with white text

Description automatically generated

To solve the above problem use :

pip install -r requirements.txt



A screen shot of a computer screen

Description automatically generated

A screenshot of a computer program

Description automatically generated

Build is happening,

mlproject.egg-info will tell us the packages which are getting installed.

We have also created a source folder and build the packages.

Add the files – git add .

git commit -m “setup”

git push -u origin main

## Tutorial 2: Project structure, Logging, & Exception Handling

* Create components folder in “src” in that folder again create \_\_init\_\_.py file.
* Create data\_ingestion.py file (which will be for reading the data)
* Create data\_transformation.py file (for data transformation)
* Create model\_trainer.py file (for model training )
* Create pipeline folder in “src” and create train\_pipeline.py file
* Create “\_\_init\_\_.py” and “predict\_pipeline.py” file in pipeline folder.
* Create logger.py, exception.py and utils.py file in “src” folder.

(check exception handling python documentation)

* Create function & class of your won to do exception handling (exception.py)

#we are going to write our own exception

############ exception handling ############

import sys

#the sys module in python provides various functions and variables that are used

#  to manipulate different parts of the python runtime environment

def error\_message\_details(error, error\_detail: sys):  #error\_details will be present in sys

    \_, \_, exe\_tb = error\_detail.exc\_info()

    file\_name = exe\_tb.tb\_frame.f\_code.co\_filename  #to get the filename

     #execution info, this will give us three info we are interested in the last one.

    #on which line exception has occured, in which file exception  has occured

    error\_message = "Error occured in python script name [{0}] line number [{1}] error message [{2}]".format(file\_name, exe\_tb.tb\_lineno, str(error))

    return error\_message

    #whenever error occurs we are going to call this function.

#created own exception class

class CustomException(Exception):

    def \_\_init\_\_(self, error\_message, error\_detail:sys):

        #inheriting from exception

        super.\_\_init\_\_(error\_message) #inherit exception class

        self.error\_message = error\_message\_details(error\_message, error\_detail= error\_detail)

#whenever we raise custom exception, it is inheriting from parent exception, whatever error msg is coming,

# it will initialize and gets assigned to a class variable

    def \_\_str\_\_(self):

        return self.error\_message   #to print the error message is called.

* Similarly do in logger.py file.

Logger.py file will be used to log the information.

Create log file.

import logging

import os

from datetime import datetime

#create log file

LOG\_FILE = f"{datetime.now().strftime('%m\_%d\_%Y\_%H\_%M\_%S')}.log"

logs\_path = os.path.join(os.getcwd(), "logs", LOG\_FILE)

#whatever logs will be created it will be respect to current working directory.

#logs folder will get created,

os.makedirs(logs\_path, exist\_ok=True) #even when there is file, keep on appending it.

LOG\_FILE\_PATH = os.path.join(logs\_path, LOG\_FILE)

#whenever we want to create the log, we have to set this up in basic config

#give the file name, where you want to store it

#format

#which level

logging.basicConfig(

    filename = LOG\_FILE\_PATH,

    format = "[%(asctime)s] %(lineno)d %(name)s - %(levelname)s- %(message)s",

    level = logging.INFO,

)# any print msg will use this config, wrt msg

If \_\_name\_\_ == “\_\_main\_\_”:

Logging.info(“Logging has started”)

If we run the logger.py file, it gets executed successfully, and logs gets created as below”:

A screenshot of a computer

Description automatically generated

* Similarly check for exception.py file

(add below code in exception.py)

import logging

if \_\_name\_\_ == "\_\_main\_\_":

    try:

        a = 1/0

    except:

        logging.info("Divide by zero ")

        raise CustomException

when run , an error message printed:

A screenshot of a computer program

Description automatically generated

Solved it by :

 try:

        a = 1/0

    except Exception as e:

        logging.info("Divide by zero ")

        raise CustomException(e, sys)

another error occurred:

A screen shot of a computer program

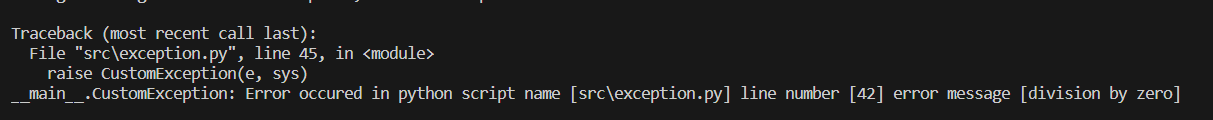
Description automatically generated

Changed super to super() as below

 super().\_\_init\_\_(error\_message) #inherit exception class

        self.error\_message = error\_message\_details(error\_message, error\_detail= error\_detail)

we get the following exception.



So the custom exception has been recorded in the logging.(check again)

We are getting the custom exception. (find why it is not logged in logging file)