

Hướng dẫn cài đặt project Phân loại phim

1. Các bạn có thể chạy trực tiếp dự án qua colab bằng file.ipynb

```
!gdown https://drive.google.com/uc?id=1UvHAQn7Ypzj4KonaSZJNew01Zl_U-aHl

Downloading...
From (original): https://drive.google.com/uc?id=1UvHAQn7Ypzj4KonaSZJNew01Zl_U-aHl
From (redirected): https://drive.google.com/uc?id=1UvHAQn7Ypzj4KonaSZJNew01Zl_U-aHl&confirm=t&uid=eab40e64-0fcs
To: /content/INT3405-projects.zip
100% 105M/105M [00:03<00:00, 31.5MB/s]

[2]
!unzip INT3405-projects.zip
!cp -r 'INT3405E Final project' project
%cd project
!unzip ml1m.zip
```

2. Nếu như sử dụng trên các nền tảng khác trên window

1. Cần install 1 số thư viện vào môi trường bằng các lệnh sau
 - Pip install gdown
 - Pip install torchmetrics
 - Pip install opencv-python
 - Pip install pandas
 - Pip install numpy
 - Pip install torch
 - Pip install nltk
 - Pip install matplotlib
 - Pip install torchvision

Tải 1 file zip về bằng câu lệnh

```
!gdown https://drive.google.com/uc?id=1UvHAQn7Ypzj4KonaSZJNew01Zl_U-aHl

Downloading...
From (original): https://drive.google.com/uc?id=1UvHAQn7Ypzj4KonaSZJNew01Zl_U-aHl
From (redirected): https://drive.google.com/uc?id=1UvHAQn7Ypzj4KonaSZJNew01Zl_U-aHl
To: /content/INT3405-projects.zip
100% 105M/105M [00:03<00:00, 31.5MB/s]
```

Sau đó trong file colab đang sử dụng nhưng dòng lệnh linux chúng ta nên đổi những câu lệnh này

```
!unzip INT3405-projects.zip
!cp -r 'INT3405E Final project' project
%cd project
!unzip ml1m.zip
```

Chúng ta sẽ sử dụng

```
import zipfile
import shutil
import os

# Giải nén tập tin INT3405-projects.zip
with zipfile.ZipFile('INT3405-projects.zip', 'r') as zip_ref:
    zip_ref.extractall()

# Sao chép thư mục 'INT3405E Final project' sang thư mục mới có tên là
'project'
shutil.copytree('INT3405E Final project', 'project')

# Di chuyển sang thư mục 'project'
os.chdir('project')

# Giải nén tập tin ml1m.zip
with zipfile.ZipFile('ml1m.zip', 'r') as zip_ref:
    zip_ref.extractall()
```

Sau khi giải nén thì chúng ta có được database bên trong folder “project”

Bây giờ có thể chạy toàn bộ project 1 cách bình thường như trên colab

Nhưng lưu ý những cái đường dẫn trong linux và window cũng khác nhau

```
movies_train = pandas.read_csv('/content/project/content/dataset/movies_train.dat', engine='python',
                                sep='::', names=['movie_id', 'title', 'genre'], encoding='latin-1', index_col=False).set_index('movie_id')
movies_test = pandas.read_csv('/content/project/content/dataset/movies_test.dat', engine='python',
                               sep='::', names=(property) str: StringMethods[Series, DataFrame], index_col=False).set_index('movie_id')
movies_train['genre'] = movies_train.genre

movies_train = pandas.read_csv('E:\\học máy\\phanloaiphim\\project\\content\\dataset\\movies_train.dat', engine='python',
                                sep='::', names=['movie_id', 'title', 'genre'], encoding='latin-1', index_col=False).set_index('movie_id')
movies_test = pandas.read_csv('E:\\học máy\\phanloaiphim\\project\\content\\dataset\\movies_test.dat', engine='python',
                               sep='::', names=['movie_id', 'title', 'genre'], encoding='latin-1', index_col=False).set_index('movie_id')
```

Chúng ta nên đổi lại những dấu “/” thành “\” và xem đã đúng đường dẫn đến database chưa.