## 

## Group members:

Usama Iftikhar Butt

Gulfam Sidhu

## Section:

CS-17

## Project:

Youtube recommendation System

## Submitted to:

Prof. Waqas Haider Khan Bangyal

Due to file size we are unable to upload on lms. So please download project from here. <https://github.com/PseudoDevelopers/Youtube-Recommendation-System>

## Libraries required

First run this command to install required libraries.

pip install pandas numpy nltk sklearn django

# Directory structure

* There is a directory called preprocessing. It is used to clean & preprocess the data.
* The second directory is processing. It contains all the AI algorithms.
* The third important directory is datasets. And it contains the csv files.
* All the other directories are of Django.

# How to run

* Open the root directory of project in cmd.
* Then run python preprocessing/Preprocessing.py. It will generate a new file called preprocessed.csv in datasets directory.
* Now run the server using command python manage.py runserver.
* Now open localhost link in browser with a port number given in cmd.
* Now click on any video. A new page with that video will open.
* Now scroll down a little bit. And you will see recommended videos.
* You can also search videos. Or click on a particular tag to see recommendations.

# Dataset

The dataset used is available here [kaggle.com/datasnaek/youtube-new](https://www.kaggle.com/datasnaek/youtube-new)

### Columns

There are 16 different columns. Including video title, description, tags, likes, dislikes, number of comments etc.

### Rows

There are over 40000 rows. But actually many rows are duplicated. Unique rows in the dataset is about 6000.

# Process explanation

* In preprocessing we use different techniques to clean the data. Like converting all textual data to lower case. Removing all special characters etc.
* And in recommendation we use cosine similarity algorithm to match the similar word in tags of different videos.