

# INSY 5336: PYTHON PROGRAMMING

## FINAL PROJECT REPORT

NAME: NISARG PAWAR

STUDENT ID: 1001720812

### Objectives:

- Write a python program that fetches movie reviews for the top 50 movies from 3 major websites: IMDB, Metacritics, and Rotten Tomatoes.
- Store the data in a comma separated file and in SQLite database
- Compare the review style across the 3 websites by calculating the cosine similarity score between the keywords used in the reviews of these 3 websites.
- Compare the top 50 movies that are considered top rated across the 3 websites by calculating the cosine similarity score between the genres of movies.

### Project Details:

**Data Constraints:** There have been no filters for the movies picked, all of them are from the listing mentioned in the project document. However, there is one constraint on the reviews link on each movie.

- a) **For IMDB:** 'spoiler=hide' since reviews with spoilers did not let you read them openly. So we drop all reviews with spoilers.
- b) **For Rotten tomatoes:** 'type=user' to toggle the reviews to audience reviews instead of critic.
- c) **For Metacritic:** no such filter required since they have a user-review dedicated page.

### Project Observations and Outcomes:

Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. The smaller the angle, higher the cosine similarity.

**For Movie Reviews:** In our case we can see that the cosine similarity between imdb and rotten tomatoes is the highest, second is between imdb and metacritic and third between metacritic and rotten tomatoes. Each of them have a similarity score close to 1 so we can say they are heading in a very similar direction in terms of movie reviews.

**For Movie Genre:** We have the highest cosine similarity between IMDB and Metacritic. Second between IMDB and Rotten Tomatoes, and Third between Rotten Tomatoes and Metacritic. This tells us that the users of IMDB and Metacritic have more in common in terms of movie genres. Whereas the users of Rotten Tomatoes and Metacritic have relatively less common genres of movies.