

2020 Programming

Competition Abstract

Programming Competition Abstract

# **Problem Background:**

Congratulations! You and your team finally landed on an internship at **Netflux**.

**Netflux** is the world’s leading streaming entertainment service with millions of subscribers. In recent times, there has been a record increase in subscribers due to the COVID19 outbreak. With most of the world stuck at home, people are spending more time watching streaming video than ever before. The abundant availability of movies and increase in demand led to a new problem where the users had a hard time finding the movies they actually want to watch. This is where Recommendation Systems come into play. Recommendation Systems are a type of **information filtering systems** that improve the quality of search results and provide items that are more relevant to the search item.

Your team’s task is to build this recommendation system that is practical and well detailed.Use this **opportunity** to challenge yourselves and develop valuable solutions for your dream company, **Netflux.**

# **The Challenge:**

Build a recommendation engine that would take information from the user and provide a list of top 5 movies that the user would enjoy. Use the two datasets **tmdb\_5000\_credits.csv** and **tmdb\_5000\_movies.csv** containing the following features:

The first dataset (**tmdb\_5000\_credits.csv**) is in the following format:

* movie\_id - A unique identifier for each movie.
* cast - The name of lead and supporting actors.
* crew - The name of Director, Editor, Composer, Writer etc.

The second dataset (**tmdb\_5000\_movies.csv)** is in the following format:

* budget - The budget in which the movie was made.
* genre - The genre of the movie, Action, Comedy ,Thriller etc.
* homepage - A link to the homepage of the movie.
* id - This is in fact the movie\_id as in the first dataset.
* keywords - The keywords or tags related to the movie.
* original\_language - The language in which the movie was made.
* original\_title - The title of the movie before translation or adaptation.
* overview - A brief description of the movie.
* popularity - A numeric quantity specifying the movie popularity.
* production\_companies - The production house of the movie.
* production\_countries - The country in which it was produced.
* release\_date - The date on which it was released.
* revenue - The worldwide revenue generated by the movie.
* runtime - The running time of the movie in minutes.
* status - "Released" or "Rumored".
* tagline - Movie's tagline.
* title - Title of the movie.
* vote\_average - average ratings the movie received.
* vote\_count - the count of votes received.

Your program should be able to take an input from the user and recommend the top 5 movies. In the current case, you are only provided with the datasets that describe the content of the movies and you can decide how to use/manipulate these datasets in any ways possible. You don’t have to use all the features/metadata to recommend a movie. You also have the freedom to decide what input you would like to get from the user. Further, you can decide what you would like to output, you can just display the movie titles or add a description. Use your creativity and technical skills to solve this problem. Take the program’s CPU and memory usage into consideration and select the most appropriate algorithm to address the problem.

In your report and presentation, use your knowledge of object oriented design methodology, design patterns and UML design tools to clearly communicate your solution and choices.

**Note:** You can use any programming language you are comfortable with as long as you have the correct environment set up.

***For rubric and deliverables information refer to the 2020 Programming Rule Book.***

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*All the best for your project and a full time offer\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***