

## PIC 16A - Project Proposal

### 1.a. What is the project about?

The objective of this project is to implement a movie/tv-show recommendation algorithm based on quantitative/qualitative data. The algorithm would recommend similar TV shows and movies based on an initial input.

### 1.b. What are the inputs and outputs to the system?

- Input: A list of movies/TV-shows of your preference
- Output: Similar movies/TV shows to the input

### 1.c. What are the intermediate steps that will happen to your dataset or user?

- Clean up the data (removing rows with null values)
- Label non-numerical values with label encoder
- Try multiple models (tree, SVM, perceptron, KNN, etc) for higher accuracy

## 2. For each member of the group, list at least 1 concrete learning goal you wish to achieve through this project.

- **Rabiga**
  - Learn about / implement a process system which filters tag words based on relevance and matches the most similar ones
- **Euibin**
  - Usage of different multiple models and their advantages/disadvantages

## 3. Please add the link and describe them in a sentence or two if there are any:

- **Public datasets**
  - <https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data>
  - <https://www.kaggle.com/datasets/satpreetmakhija/netflix-movies-and-tv-shows-2021>
  - <https://grouplens.org/datasets/movielens/25m/>
- **Online tutorials**
  - <https://www.kaggle.com/code/mfaaris/netflix-recommendation-both-movie-and-tv-show>