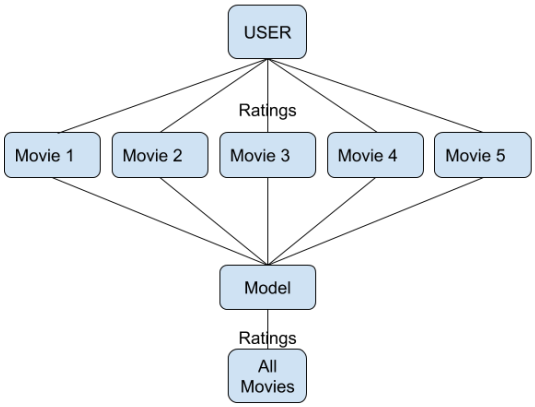
Brereton Conway, Aldin Llolla, Daniel Pinto
CPE 695
Project Proposal

The purpose of this project is to create a recommendation system utilizing the MovieLens database. This database includes 20 million user ratings across 27,000 different movies. It also provides more information such as tags, the relevance of these tags, and genres. The task is to analyze a user's ratings for various movies to predict their enjoyment of other movies, such that if they enter a movie that is in our database we can accurately predict a rating which they might give it.

Our model will be considered supervised since our database provides ratings as classifications for the movies. We will provide an initial base model that can classify/ predict a new user's rating for a specific movie they input after they provide us with a base set of ratings to work with. The process will be as follows: the user will first enter ratings for some number of movies and a final movie without a rating, then our model will process the data the user has provided us with, and finally it will predict what the user will rate that final movie. Once this base model is finished, we will attempt to add more factors such as weighing different attributes more (i.e. genre) or weighing a rating more if the person who provided the input has more ratings in our system (i.e. they are an experienced movie rater). Similarly, data can be obtained from IMDb where more attributes related to the movie can be added. A basic example would be if a user continuously rates a specific director's movies highly it can be guessed that they would also rate another one of his/her movies highly. We will be running our model against itself (using cross-fold validation for example) to test which created model will perform best. By the end of this project we hope to have created the most efficient model which we can base a user interface around. Hopefully, this application can be used by our friend group in order to figure out what to watch for movie night! Below is a very high-level basic overview of how the system should work in the end.

Our group will divide up the work as follows: Daniel and Aldin will take charge of preprocessing of data and the creation of scripts to connect useful information from IMDb to our data. Brereton will be tasked with model creation and we will all complete testing together.


USER
Ratings
Movie 1
Movie 2
Movie 3
Movie 4
Movie 5
Model
Ratings
All Movies