MovieLens project

Andrea Roetti

June 11, 2020

Introduction

The MovieLens project goal is to build up a prediction system for movies ratings based on a historical series of records. The dataset we will use as base to create our model is called **edx**, while the dataset where we will measure the efficiency of our model, as if we did not know the given ratings, is called **validation**. In order to get started, we first load some libraries that will be useful for the analysis.

```r
library(tidyverse)
```

We therefore read the two datasets that we have previously already dowloaded as .rds files and saved locally for reasons of size: a file bigger than 100MB is not possible to push to Github, therefore not possible to be kept in the project folders syncronised with it. We know that using complete path rather than relative path is not recommended, in fact below we also provide, as comment the path to the datasets as they were in the "data" folder within our working directory.

```r
edx <- readRDS("C:/Users/roetti/Documents/MovieLens/edx.rds")
# edx <- readRDS("data/edx.rds")
validation <- readRDS("C:/Users/roetti/Documents/MovieLens/validation.rds")
# edx <- readRDS("data/validation.rds")
```