# MovieLens

## theSherrinford

## 20/06/2020

This is a pdf document generated from attached R Markdown file.

The aim is to use the provided data to build a model which would ultimately help predicting the ratings for unknown data set.

Here's a part of how the provided data looked like.

```
##   userId movieId rating timestamp                            title
## 1      1     122      5 838985046               Boomerang (1992)
## 2      1     185      5 838983525               Net, The (1995)
## 4      1     292      5 838983421               Outbreak (1995)
## 5      1     316      5 838983392               Stargate (1994)
## 6      1     329      5 838983392 Star Trek: Generations (1994)
## 7      1     355      5 838984474       Flintstones, The (1994)
##                           genres
## 1               Comedy|Romance
## 2          Action|Crime|Thriller
## 4   Action|Drama|Sci-Fi|Thriller
## 5          Action|Adventure|Sci-Fi
## 6 Action|Adventure|Drama|Sci-Fi
## 7          Children|Comedy|Fantasy
```

Movie rating is considered to be consisted of three parts. 1. The average rating calculated with arithmetic mean of all available ratings. This value is calculated to be

```
## [1] 3.512465
```

2. Effet of movie bias calculated with all the ratings available for that movie. Here's a part of what we've got

```
## # A tibble: 6 x 2
##   movieId     bm
##     <dbl>  <dbl>
## 1       1  0.415
## 2       2 -0.307
## 3       3 -0.365
## 4       4 -0.648
## 5       5 -0.444
## 6       6  0.303
```

3. User specific effect calculated using all the available ratings gives by the specific user. Here's a part of the same

```
## # A tibble: 6 x 2
##   userId      bu
##    <int>   <dbl>
## 1      1    1.68
```

```
## 2           2 -0.236
## 3           3  0.264
## 4           4  0.652
## 5           5  0.0853
## 6           6  0.346
```

Using this model, movie ratings are calculated for movies is validation set and the RMSE is calculated as

```
## [1] 0.8653488
```