

HPC CLOUD HACK 2018

Team **Belkin**

Nishant Shah

Rayan Ahmed

Beemnet Alemayehu

Objective

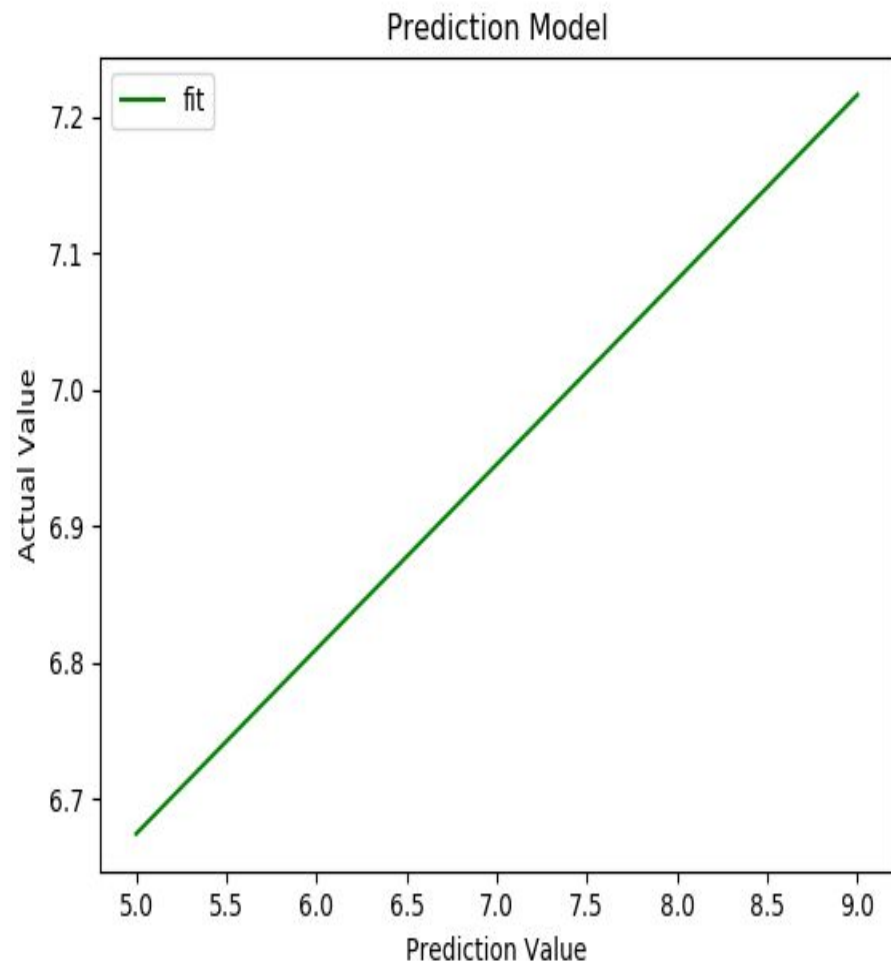
Create a Machine Learning model to predict overall movie ratings based on ratings of different individual factors

Goals

- Gather data
- Filter data into usable form
- Create a Machine Learning model
- Feed the data to the model
- Compare the prediction with a test data
- Use linear regression to find relation with prediction and actual value

Results

- All factors were directly related to overall movie ratings but some more than others
- Plot rating has the highest correlation (of the factors used) with overall movie ratings
- Model prediction (used 90% of the data for training) gave us 8.75% error
- The prediction model can help movie makers choose what aspect of the movie to prioritize
- Movie companies can use this model to make estimation of how successful their movie is going to be



Prediction	Actual
7.6	7
7.6	7.5
7.6	6.6
7.6	7.9
7.6	7.8
7.6	6.8
7.6	7.4
7.4	5.3
7.6	7.4
7.2	7.9
5.0	6.3
7.6	6.7
8.0	6.5
8.0	6.2
7.6	5
7.6	7
7.6	6.3
7.6	5
7.9	8
7.6	7.6
7.6	7.9
7.6	8.3
8.1	6.7
7.6	5.6
7.6	7.6
7.6	7.2
7.6	6.3
7.6	7.1
7.6	6.2
7.6	8.1
5.0	7

Obstacle and Solutions

- Could not get data from IMDB and Rotten Tomatoes APIs directly
 - Had to look for data already compiled
- Transferring files to the Google Cloud
 - Had to generate keys using Putty

THE MENTOR EFFECT

BEFORE

AFTER



Thank You!

This hack was made possible with great support from our mentors and all the awesome folks here with us today