**Introduction**

**an introduction/overview/executive summary section that describes the dataset and summarizes the goal of the project and key steps that were performed**

In this report, our goal is to predict the movie rating by using a machine learning algorithm. The data set is coming from MovieLens with about 10M rows of userId, movieId, rating, timestamp, title and genres. Movies are released from early 20th century till 2008. Movie rating are made by user from 1995 to 2009.

Data cleaning is applied to the original data following with data exploration. 4 major effects are identified. Our approach is using normalization to these global effects on baseline rating and regularization (by tuning parameter on lambda) to penalize large estimates that come from small sample size.

1. Movie specific effect
2. User specific effect
3. Genre specific effect
4. Rate per Year specific effect

The evaluation of algorithm is based on root mean squared error (RMSE) of predicted rating against actual rating. Algorithm is trained on train set and being test on test set. Final RMSE is presented basing the on the final hold-out validation set with result in the tier of “RMSE < 0.86490”.

**Method**

**a methods/analysis section that explains the process and techniques used, including data cleaning, data exploration and visualization, insights gained, and your modeling approach**

**1. Data Cleaning**

edx data set contains 6 columns (userId, movieId, rating, timestamp, title and genres).

**2. Data Exploration**

**3. Modeling Approach**

**Result**

**a results section that presents the modeling results and discusses the model performance**

**Conclusion**

**a conclusion section that gives a brief summary of the report, its limitations and future work**