

MovieLens Project Report

Author: Pranjal
Date: 2025-08-17

Introduction

This report explores the MovieLens 10M dataset. We build models to predict movie ratings and evaluate them using RMSE.

Data Preparation

The dataset was loaded from ratings.dat and movies.dat, then merged into one dataset (edx). We created a validation set (10%) to evaluate the models.

Models and Results

Model 1: Mean Only - We predicted all ratings using the global mean.

Model 2: Movie Effect - We adjusted predictions by accounting for average rating differences for each movie.

Model 3: Movie + User Effect - We further adjusted predictions by accounting for user biases.

RMSE Results

Model	RMSE
Mean Only	≈ 1.05
Movie Effect	≈ 0.94
Movie + User Effect	≈ 0.865

Conclusion

We built multiple models to predict movie ratings using the MovieLens dataset. The Movie + User Effect model performed best with the lowest RMSE.