Movie Recommendation System

Tara Greenwood, Natasha Kacoroski,
Werlindo Mangrobang, John Naujoks, Anna Zubova

Data

- Movie ratings
- Demographic user data
- Movie information

ALS model

ALS model based on available data on movie ratings

- Dropped Timestamp data
- Trained ALS on available ratings
- Replaced predictions with NaNs with average ratings per user to test performance
- Evaluated predictions, RMSE 0.87

Predictions with K-mean clustering

Clustering model based on demographic information

- K-means clustering to generate **8 demographic clusters**
- Model predicts rating based on **mean movie rating by other users** in the cluster
- Output data: **movie ID**s with **rating score**, **user ID** and **cluster** that user belongs to

Combining 2 models

1. Apply **ALS model** to come up with predictions
2. Replace **NaNs** with predictions from **cluster model** (average rating among cluster)
3. If there still NaNs, fill with **cluster average**
4. Extract **10 movies with the highest predicted rating** for a particular user

Thanks!

Questions?