



# Surprise Software Presentation

DSC 672: Data Science Capstone  
Ross Pollock | [rmpollo1@gmail.com](mailto:rmpollo1@gmail.com)



# Surprise

Scikit for Recommender Systems

Surprise is a python library designed to allow users to develop and analyze recommender systems.

Surprise comes with many built in algorithms and dataset as well as evaluation metric and strategies.

Surprise is designed specifically for explicit rating data rather than implicit data.

<http://surpriselib.com/>



# Surprise: Dataset & Readers

- Reader Class
  - Defines Data Format
  - Rating Scale (1-5 default)
  - Line / Row Format (user item rating)
  - Separator
  - Skip Lines
- Dataset
  - Load Builtin
    - Movielens 100k & 1m
    - Jester
  - Load from File
  - Load from Dataframe
  - Load from Folds



# Surprise: Algorithms

- Algo\_base
  - Base Class for recommender systems
  - Allows users to develop their own recommender systems
- Baselines
  - Normal Predictor
  - Baselines Only
- Neighborhood Methods
  - **KNN** (K Nearest Neighbors)
  - KNN With Means
  - KNN With Z-Score
  - KNN With Baselines
- Matrix Factorization Models
  - **SVD**
  - NMF (Non-Negative Matrix Factorization)
  - SVD++
- Others
  - Co-Clustering
  - Slope One



# Surprise: KNN & SVD

- KNN

- k | Max Number of Neighbors
- min\_k | Minimum Number of Neighbors
- sim\_options | Similarity
  - Mean Squared Distance
  - Cosine
  - Pearson (optional Baseline)
  - User / Item Based
  - Min Support
  - Shrinkage for pearson with baseline

- SVD

- N\_factors
- N\_epochs
- Biased (use baselines)
- Initialization
  - Mean & Std.
- Regularization
  - L2
  - Can set for each part or all
- Learning Rate



# Surprise: Evaluation

- Strategies (model\_selection)

- Train Test Split
- Cross Validation
  - KFold
  - Repeated KFold
  - Shuffle Split
  - Leave One Out
  - Predefined KFold

- Metrics (accuracy)

- RMSE - Root Mean Squared Error
- MSE - Mean Squared Error
- MAE - Mean Absolute Error
- FCP - Fraction of Concordant Pairs



# Surprise: Parameter Search

- Grid Search CV
- Random Search CV



# Surprise: Saving & Loading

- Dump Module
  - Dump
    - Save Predictions and / or Model
  - Load





# Tutorial Link

<https://youtu.be/c-2qiJLtc3o>