Data Science – Final Project Proposal

Business Issue:

Help users of the website, Book-Crossings, find books that they are likely to enjoy.

Potential Impact (if based on a book sales company):

Increase user discovery and satisfaction, resulting in overall increased sales or retention

Data Source:

Book-Crossing Dataset

<http://www2.informatik.uni-freiburg.de/~cziegler/BX/>

Objective:

Build an algorithm that predicts a user rating of any book, based on user similarities in ratings.

Hypothesis:

Using an algorithm to predict a user’s rating of a specific book will prove to be more accurate than the null

Approach:

Use User-User Collaborative Filtering technique to predict the rating a user would give a book that they have not previously rated. Per John’s suggestions in Slack, make use of the scikit SURPRISE package, to simplify the technical aspect of model building.

Success Metric:

Use the following error metrics to determine how well the model performs: RMSE, Precision and Recall