



Gute's Books

New Market Strategy

Who We Are

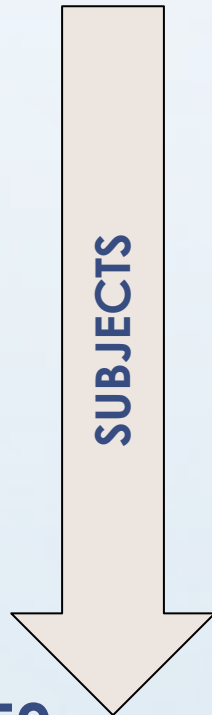
- Established retailer launching an online store
- Inventory: Over 4,000 books and counting
- Competitive Advantage: Digitized Library

New Strategy

- Improve search capabilities
- Improve recommendations based on previous book interests

Data Preparation

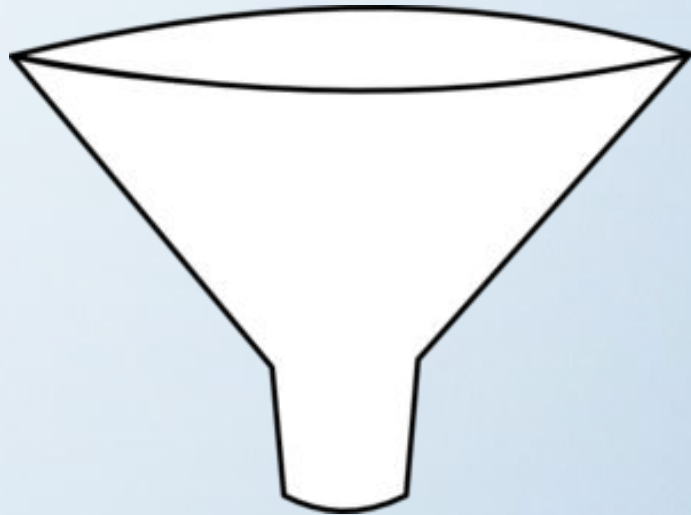
250+



SUBJECTS

~50

Anthropology, Economics,
Sociology, Psychology



Social Sciences

Classification

- Purpose
 - Classify subjects of new books
 - Assign reading levels to books
 - Recommend books based on previous book interest
- Subject classification models included:
 - SVM
 - Naïve Bayes
 - Random Forest
 - Maximum Entropy
- Maximum Entropy was the best model (highest F-score, precision, and recall)

Assigning Reading Level

- Analyze contents of each text
- Produce a readability score
- Coleman-Liau Index
 - Characters per word, not syllables
 - Computers more accurately count characters

Filter Selections

- Subject
- Children's or Adult
- Reading Level

Deployment

- Previous models used as layers for the recommendation engine
- Recommendations based on distance measures of subject and reading level



GUTE'S BOOKS

**Treva Winlock
Derek Kopecky
Ian O'Connor
Patrick Passafiume**

100 Text Sample

