

Capstone 2 Proposal
Book Classification
Evan Hintz

This project will focus on unsupervised machine learning models to classify various book texts into genres. Since book genres already exist I'm very curious to see the patterns that a machine learning model will pull out of texts as opposed to attempting to classify them into pre-existing themes. The end goal is to be able to offer companies a new and potentially better scheme for labeling books which will aid in consumer recommendation systems. Theoretically, an unsupervised model will be able to see patterns that may not be included in the traditional book genres.

The data used will be collected from Project Gutenberg, an online library of open-source texts (expired copyrights, or granted rights). A selection of various genres will be used, cleaned, and analyzed.

As stated, the final product will ideally be an improved source of information for book recommendations.

Project Gutenberg: <https://www.gutenberg.org/>