

Judging a Book by its Characteristics

Brandon McNeil

Abstract

The goal of this project was to analyze book titles from Goodreads.com to see if we can predict a book's average user review using a linear regression model.

Data

For this project, we used web scraping techniques to gather 13,000+ records from Goodreads.com, gathering key features available to us on the website.

Algorithms:

We used a simple linear regression to try to predict user scores. We used a KFold method of cross validation to conclude that this was the best model to use for our project. Our final model yielded a R-Squared of **0.184**. While these results

Tools

- BeautifulSoup and Selenium packages for web scraping.
- Pandas package for data manipulation.
- Matplotlib, Seaborn and YellowBrick packages for plotting.
- Sklearn for Linear Regression Analysis.