Prediction of Amazon Best Seller Book Prices

Prepared By

Tamader AboAlhassan

Afrah Almoutiri

Moneera Alfulaij

Abstract

The goal of this project is to apply Linear Regression on data scraped from Amazon Best Seller Book website, to predict the price of book based on the features of that book (book name, author, rating, customer rated, year and genre). I worked with data that I have scraped from Amazon.com using Beautiful Soup. Then we import csv file of Amazon Best Seller Book then combined it with web scripting data. After that, the data was combined into a single Pandas data frame, in order to achieve the goal of this project. After cleaning the data, I started to explore the dataset and split the data into train/validation, and test data. Next, Linear Regression model was used to train the data and to achieve the goal of this project. Finally, I communicated my work using a PowerPoint presentation to show my findings.

Design

By training the data using Leaner regression models, we could predict the price of a book, based on the features of that book (author, rating, customer rated etc.)

Algorithms

The methodology used in this project is: Problem understanding, data validation, data exploration, data visualization, feature engineering, training and modeling the data.

Data

The data is scraped from of Amazon Best Seller Book and contains information about Book from2010-2021. An individual sample is a singular book (a row in the dataframe). The dataset has 700 rows and 9 columns.

Tools

- requests and BeautifulSoup for web scraping
- pandas for data manipulation
- statsmodels and sklearn for linear regression modeling
- matplotlib. pyplot, seaborn, and stats for plotting
- Python and Jupyter Notebook

Communication