Lucas Reynolds
Springboard, Data Science Career Track
January 2020 Cohort
Capstone Project 1

Project Proposal

Problem

Beyond specific trials, patient feedback on medication often doesn't get reported back to health providers or pharmaceutical companies. The majority of patient reviews and ratings are reported online. The data from these online reviews could provide insight as to which drugs receive the best ratings and why. What are patients saying about medication with high ratings versus those with low ratings? Can we use patient reviews to predict the rating a drug will receive?

Dataset

Drugs.com is the largest, most used, website with independent information on medications. The site provides a database of peer-reviewed information on various medicines as well as user ratings and reviews gathered from across the web. We will use a dataset including 215,063 rows of ratings and review data from Drugs.com available via the UCI Machine Learning Repository.[1]

Process

1. Scrape dataset from UCI Machine Learning Repository
2. Perform exploratory data analysis
3. Use Python to wrangle and clean data
4. Build model to predict rating based on review keywords
5. Estimate performance and make additional recommendations

Deliverables

Wrangled dataset, along with documentation and code for exploratory analysis and predictive model will be provided in a GitHub Repository. We will also walk through our process, results, and recommendation in a presentation.

---

[1] https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29