Capstone Project 1 Proposal

Dataset

Drug review dataset from Drugs.com via the UCI Machine Learning Repository. The dataset includes the name of the drug, name of the condition, patient review, 10-star patient rating, date of review entry, and number of users who found the review useful with over 215,063 instances stored in .tsv files.

Problem

What patient reviews result in the lowest drug ratings? We'll read text from patient reviews into a model, identify words that indicate good or bad ratings, and use model to predict ratings.

Outcome

Build a machine learning model that can predict what drugs will receive good or bad ratings.

Deliverables

Code for machine learning model and findings in a GitHub Repository, as well as a presentation