DATA 620 Final Project Proposal

7/11/2021

Authors: John Mazon, LeTicia Cancel, Bharani Nitalla

Data Source

https://www.kaggle.com/rohanharode07/webmd-drug-reviews-dataset

WebMD dataset that includes data for various drugs, the condition the drug was meant to treat, possible side effects, user reviews and some demographic information about the reviewers. We plan on exploring the relationship between the condition to be treated and the possible side effects of the drug. If there is time we can do a second analysis of the relationship between the ease of use rating, the effectiveness rating and the overall user satisfaction rating.

The data also includes written user reviews which will be used for the text processing portion of this project.

Hypothesis

We suspect that drugs with a high ease of use rating will also have a high satisfaction rating. We also suspect to see repeat side effects such as nausea, headache, and dizziness.

Preliminary Plan & Concerns

The preliminary plan is to explore different networks since we have many datapoints that are can be analyzed for relationships. Once we have useful networks we can then further explore them based on the demographic data such as gender and age range.

Roles/Responsibilities

John Mazon – Improve network analysis code, assist with final review

LeTicia Cancel – Final project proposal document, first draft of network analysis code, assist with final review

Bharani Nittala – Sentiment analysis code, assist with final review