Project Proposal

Problem statement (Hypothesis)

Can we forecast the sentiment based on the patient reviews on drug products and their medical conditions?

Context

This dataset collects patient reviews detailing experiences with specific medications for various medical conditions. The challenges lies in extracting the generanized conclusions, and forecast rating based on the medical conditions while considering the potential biases present in the dataset.

Criteria for Success

The data science team will deliver the data driven insights within three weeks, meeting the following criteria:
Accuracy reaches at least 80% in sentiment classification analysis.

Scope of the Solution

The data science team will provide the generalized conclusions of sentiment classification analysis within three weeks.

Constrains

Computational resource limitations in some modeling tasks (such as neural networks) pose a challenge, particularly for new data scientists unfamiliar with cloud services.

Stakeholders

Healthcare Providers: Interested in understanding patient sentiments and treatments to enhance patient care strategies.

Pharmaceutical Companies: Seeking insights from patient experiences to improve drug efficacy and minimize side effects in their products.

Data Scientist Teams: Engaged in analyzing the dataset to give insights for identification medicine effectiveness, side effect and rating prediction.

Data sources

https://archive.ics.uci.edu/dataset/462/drug+review+dataset+drugs+com (doi:10.24432/C5SK5S), the dataset creators: Surya Kallumadi, Felix Grer.

Approach

1. Data Wrangling: Collect, clean the dataset.
2. Exploratory Data Analysis (EDA) and featuring engineering: Visualize features, inspect each feature and correlation with others, the positive, negative, and neutral sentiment distributions, what is the most frequently occurring words in different sentiments.
3. Modeling: Apply suitable machine learning techniques for sentiment analysis.

4. Data Story: Combine insights from analysis into a clear narrative with impactful visualizations.

Deliverable

A GitHub repo containing the work of each step including: A slide deck and A project report