

Drug Recommendation System

Instructor: Dr. Ying Lin

Team Members: Akshita Chandiramani, Ruchita Jadhav, Sanjeev Ramasamy Seenivasagamani

Problem & Objectives

The Drug Review Dataset from the UCI Machine Learning Repository provides patient reviews on specific drugs along with related conditions. Our aim is to automatically classify drug reviews into patient's condition and predict drug ratings to recommend drugs to patients.

Through this project we will be able to map the patient's conditions to it's probable disease/sickness through online User Interface created in R-Shiny. The application also provides drug according to patient's conditions.

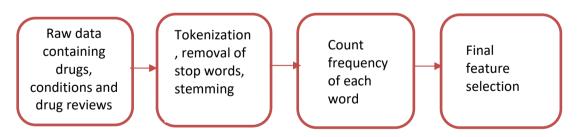
Patient is able to see available drug options along with the reviews from other patients.

Data Description

The Drug Review Dataset from the UCI Machine Learning Repository provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. The data was obtained by crawling online pharmaceutical review sites.

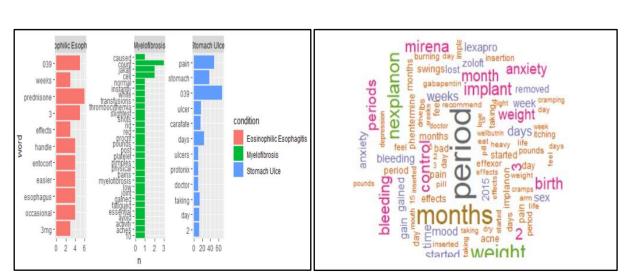
The dataset consists of 6 features and has about 161297 records. It consists of data ranging from February 2008 to December 2017.

Data Collection and Pre-processing



Extraction of Patient's condition from Reviews

Word Cloud on Reviews



Model Description

Type of Algorithm	Model	Features	Evaluation
Classification	k-Nearest Neighbour	Reviews, Tokenized key words	Accuracy
	Tree-bagging	Reviews, Tokenized key words	Accuracy
	Naive Bayes	Reviews, Tokenized key words	Accuracy
Clustering	k-Means	Reviews, Tokenized key words, document term matrix	Sum of square error
	HAC	Reviews, Tokenized key words, document term matrix	-

Model Comparison metrics:

Classification:

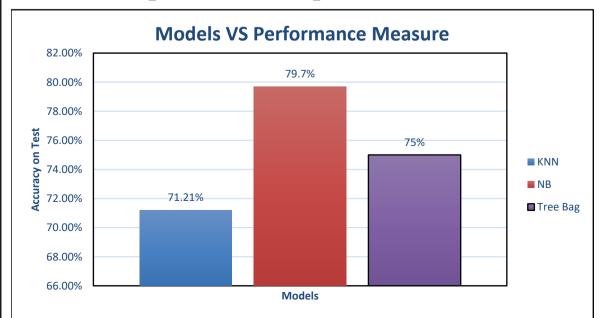
Training and Testing split: Train Data: 70% Test Data: 30%

Performance Measure: Accuracies

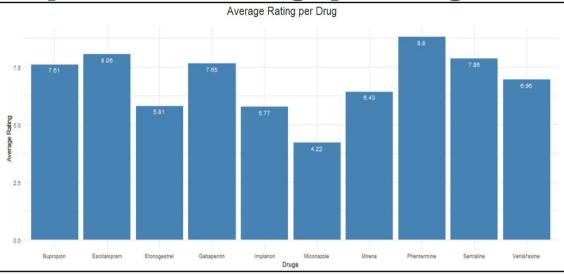
Clustering:

Performance Measure: Sum of square error

Results - prediction performance

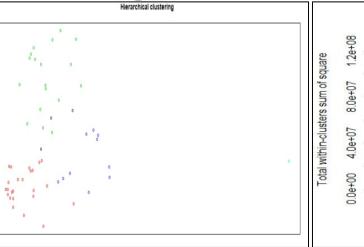


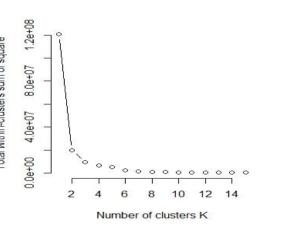
Graph of User Ratings per Drug



Classification Model(Naïve Bayes)

Clustering Model





Conclusions:

A person could use our platform to input the symptoms and in return our model will analyze patient's sickness and present the condition. Based on this condition the application will also able to provide the appropriate drug to the user. The application also shows the reviews of the drugs and ratings from different users fetched from the dataset.

Future Work:

Prediction of rating using sentiment from user reviews.

Reference:

https://cfss.uchicago.edu/notes/supervised-text-classification/

https://cai.tools.sap/blog/text-clustering-with-r-an-introduction-for-data=scientists/