

Pill Recommendation System

Made by - Aditi Goyal (B19EE003) , Darshit K Jain (B19EE024) , Harsh Rajiv Agarwal (B19EE036)

Code Structure -

1. Data product - We deployed the findings of our analysis for visualization purposes and so that it can have real world utility value. Check it out here - <https://darshitjain04.github.io/PRML-Project-Frontend/>
2. [Pill_Recommendation.ipynb](#) - This is the main file with all the preprocessing, EDA and Machine learning and Deep Learning Models.
 - a. Installing libraries and Dependencies
 - b. Importing dataset - the dataset is fetched from google drive.
 - c. Exploratory Data analysis - all types of relations between the data are plotted to understand the nuances and hidden interpretations. Pi charts are added to show distribution of ratings.
 - d. Data preprocessing - Basic data information, cleaning up the data using removal of stopwords, removing non-letters, stemming etc.
 - e. Dividing into test and train and transforming using Count Vectorise
 - f. Applying Machine Learning models
 - i. SVM
 - ii. Random Forest
 - iii. Multinomial Naive Bayes
 - iv. LGBM
 - v. Linear Regression
 - g. Applying Deep learning Models
 - i. Neural Net 1
 - ii. Neural Net 2
 - iii. LSTM - Long Short Term Memory Model
 - iv. CNN
 - h. Applying Harvard Sentiment Dictionary Analysis - Analysing emotions of a review using the Harvard emotional dictionary which has opinionated words was another way to include the sentiments of a review. Using the values of Harvard Dictionary we find out the number of positive and negative words in a review
 - i. Classifier Combination - Voting - We need to combine all the predictions from the array of classifiers and predictive methods used. Each ML/DL model was given a weight in proportion to its accuracy with respect to other ML/DL models. Predictions from LGBM and Harvard Sentiment Prediction were added directly since they represent separate

individualistic predictions. All these 4 components were summed and multiplied by the usefulCount of the respective review.

3. [Emotional Analysis.ipynb](#) - This contains the emotional analysis done on the reviews using NRC Lexicon Library.
 - a. It contains the same preprocessing as the above file.
 - b. Post that NRC Lexicon library is explored - Using the review for each drug besides sentiment analysis (Positive/Neutral/Negative) , Emotion analysis was also performed to understand human behaviour and underlying opinion. The *NRC Emotion Lexicon* is a list of 14k unigrams and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).
 - c. Reviews are passed to the library functions to get the emotion scores.
4. Frontend code -
 - a. The frontend was made using React JS. The code relating to all the website deployment is added in this folder.
 - b. It is a stand alone application. It fetches data from the 2 .csv files(total predictions from all models and the emotion analysis files) and manipulates them to display results.

How to run -

1. Run the [Pill Recommendation.ipynb](#) file first.
 - a. The SVM code keeps crashing hence those cells should be avoided while running.
 - b. LSTM takes about 1.5 hrs to complete running.
 - c. The predictions from all the models are collected and stored in a .csv file.
 - d. The final prediction scores calculated are also stored in a .csv file at the end.
2. Run the [Emotional Analysis.ipynb](#) file after that.
 - a. It is a completely separate entity from the [Pill Recommendation.ipynb](#) file. The results from both the files are used to predict data based on the reviews and rating as shown on the deployed website.
 - b. It takes 6 hours to run.
3. View the end result at - <https://darshitjain04.github.io/PRML-Project-Frontend/>