

SUMMARY

ABSTRACT

This project is designed to analyse reviews for depression drugs using unsupervised machine learning algorithms.

DESIGN

- Data preprocessing
- Exploratory data analysis
- Topic modeling
- Topic interpretation

DATA From UCI machine learning repository.

Data points : 12,164

Attributes : reviews, ratings, drug name and condition

METHODOLOGY

- Change data to lowercase , remove punctuations, stop words and lemmatization
- Vectorize by TFidf
- Topic modeling using NMF algorithm
- Topic interpretation using top word and top document expressions
- Analyse the ratings with the average topic expression
- Analyse drug brands using average topic expressions

Results

20 topics were generated

Topic names

- Length of time on medication
- Sleep problems
- Weight gain
- Dose of drugs
- Anxiety
- Nausea
- Panic attack
- Changes
- Sex drive

TOOLS

Python - Data analysis, modeling

TABLEAU - Data visualization

COMMUNICATION : Tableau Sheets

