

Drug Review Seniment analysis

Tudor Andrei Dumitrascu

May 04, 2022

Goal

Predict the rating of a product based on user reviews

Dataset

- ▶ CSV file containing:
 - ▶ Drug name (categorical)
 - ▶ Condition (categorical)
 - ▶ Text review (text)
 - ▶ Rating (numerical: 1-10)
 - ▶ Date
 - ▶ Useful Count (numerical)

Main approaches

- ▶ Machine learning methods
- ▶ Transformer models

ML models

- ▶ Text feature extractors:
 - ▶ Bag of words
 - ▶ TF-IDF
- ▶ Models:
 - ▶ Random Forest
 - ▶ Decision Tree
 - ▶ KNeighbourClassifier
- ▶ Metric:
 - ▶ F1 Score

Results - mean F1 Score

- ▶ RF + TF-IDF(100 features, (1,2)ngrams)
- ▶ Train: 91%
- ▶ Test: 71%

DL Models

- ▶ Bert based models
 - ▶ Bert Uncased
 - ▶ PubMedBERT

Mistakes and Issues

- ▶ DL training crashed multiple times
- ▶ Used wrong subset

Questions