Patient's Condition Prediction Based on Drug Review

Imam Hasan Araf

Presentation Agenda

Business Problems and Goal

The Business Problem

- Pharmaceutical companies to better anticipate the outcome of drugs
- Personalized treatment plan and drug recommendation for patients

Goal

- Predict the condition of the patient based on the drug review
- Recommend a classification model

Bottom Line Up Front

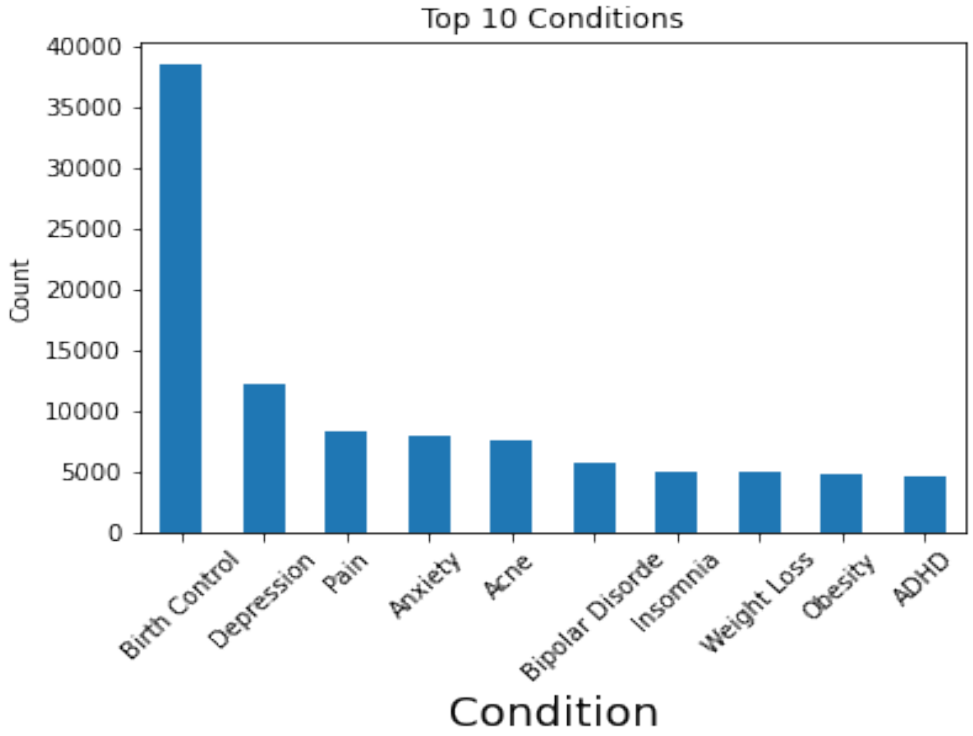
Logistic Regression provided the best results.

Data Understanding

- UCI ML Drug Review Dataset
- Shape: (215603,7)
- 7 features are:

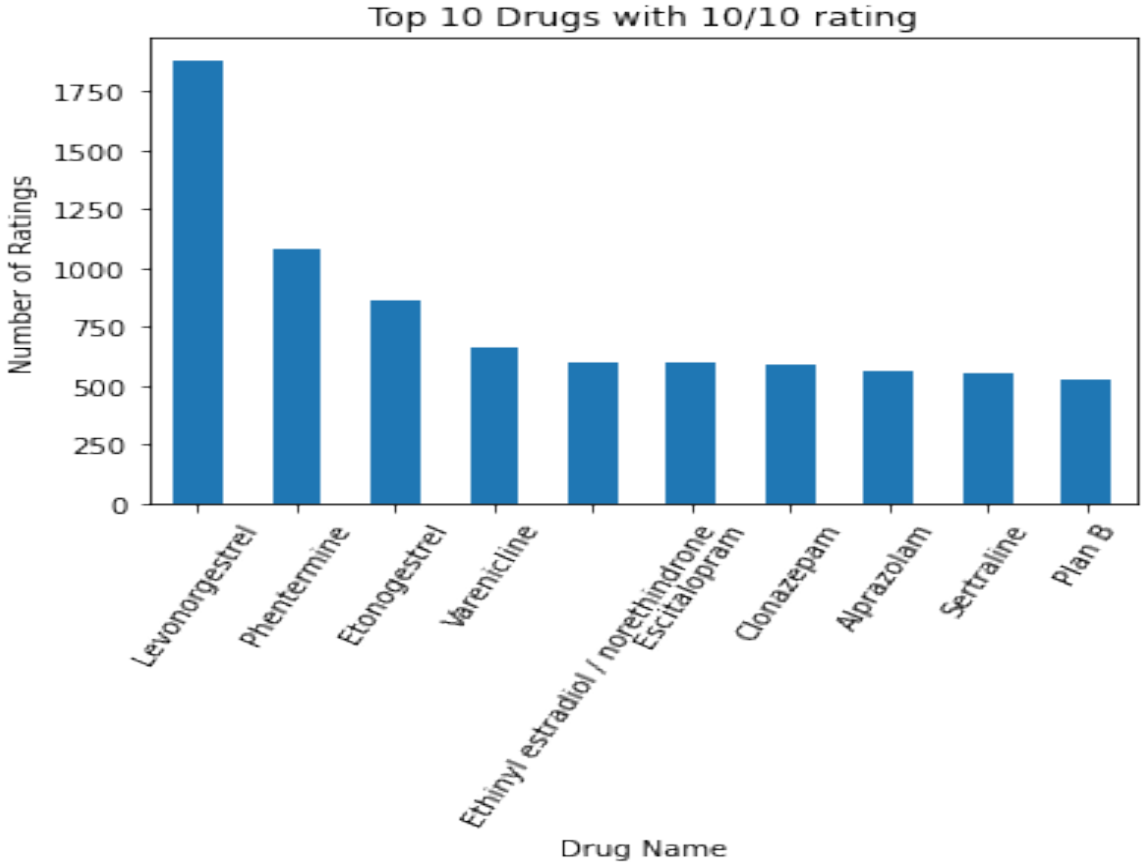
  uniqueID, drugName, condition, review, rating, date, usefulCount
- Target feature: “condition”
- Data Preprocessing:
  1. Missing value removal
  2. Bag of Words model to vectorize the review feature

Exploratory Data Analysis



Top 5 conditions:

Birth Control, Depression, Pain, Anxiety, Acne

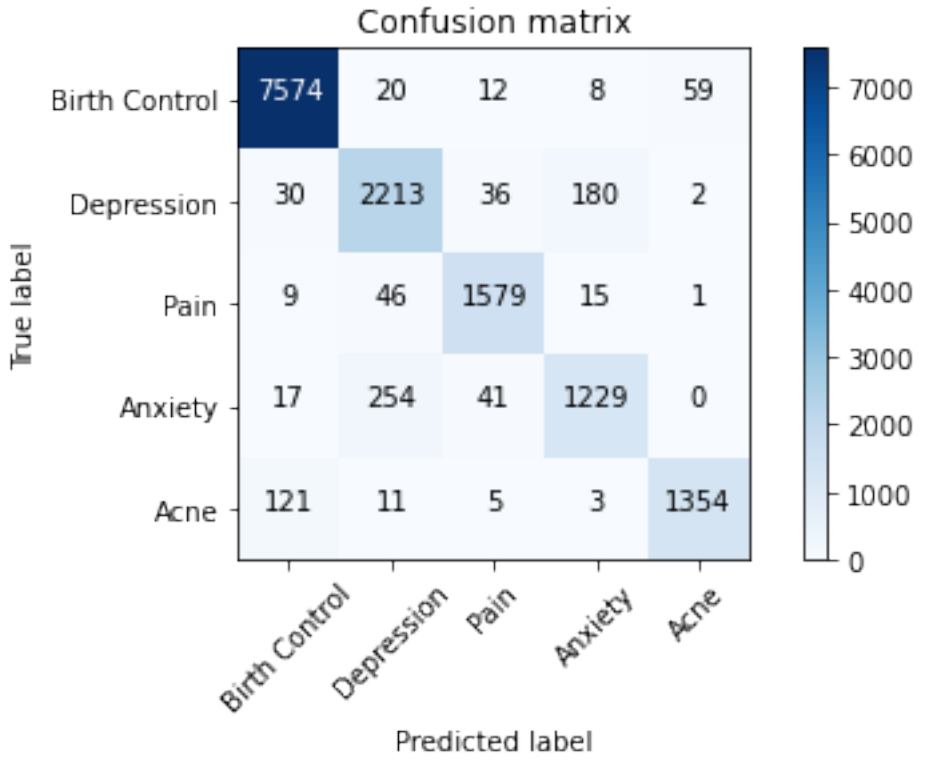


Top 5 drugs with 10/10 rating:
Levonorgestrel, Phentermine, Etonogestrel,Varenicline, Norethindrone

Logistic Regression

Classification Metric:

Accuracy: 0.94
Recall: 0.91
Precision: 0.92
F1- Score: 0.91

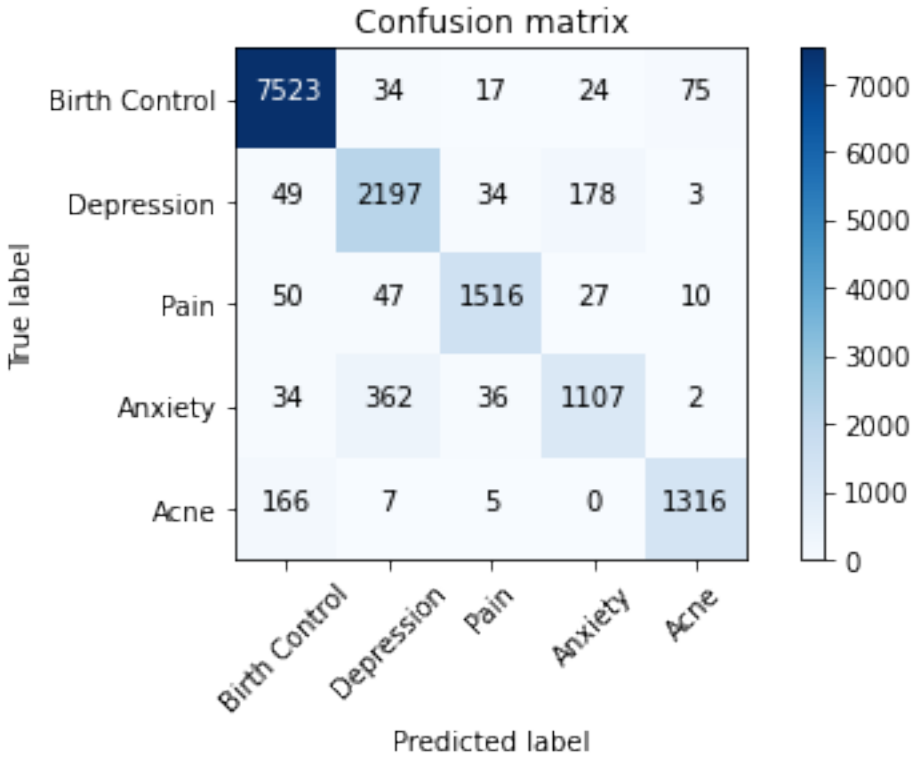


Interpretation of Confusion Matrix:
The diagonal elements represent the number of points for which the **predicted label** is equal to the **true label**, while off-diagonal elements are those that are mislabeled by the classifier.

Naïve Bayes Classifier

Classification Metric:

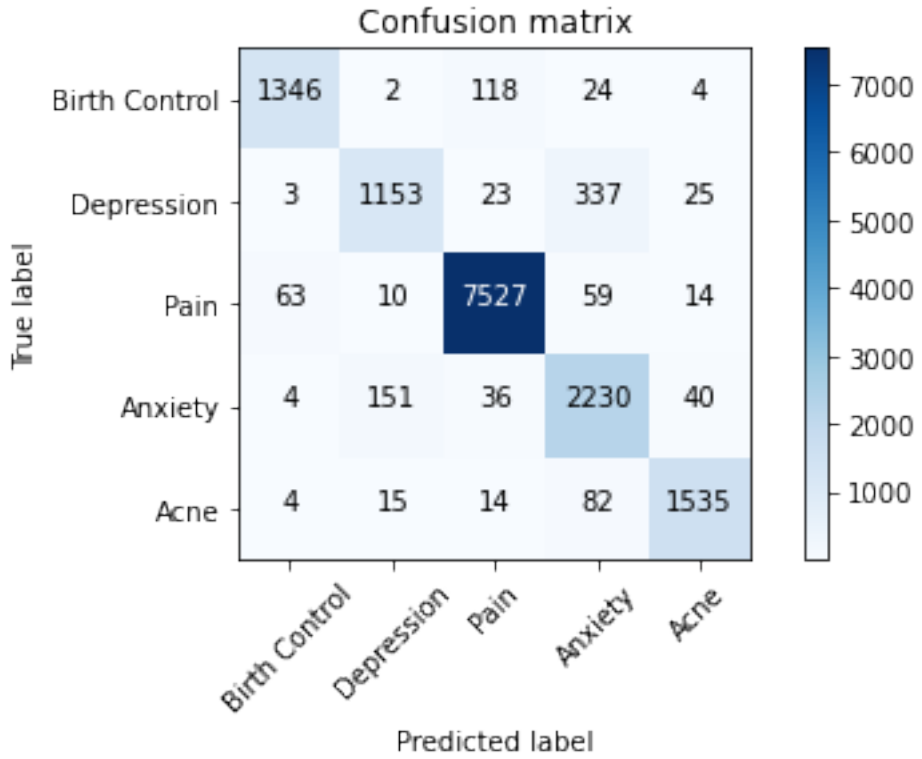
Accuracy: 0.92
Recall: 0.88
Precision: 0.90
F1- Score: 0.89



XGBoost

Classification Metric:

Accuracy: 0.93
Recall: 0.89
Precision: 0.91
F1- Score: 0.90



Model Selection

- **Accuracy** is the chosen metric for this classification problem.

- In terms of Accuracy,

Logistic Regression model

Conclusion

Next Steps

- TF-IDF, Word2vec, GloVe
- Few Shot Learning techniques (FastText, SetFit, etc.)
- Zero Shot Learning techniques