

Comp 472 - Project 1

Team Vipers

Tommy Andrews - 40094870

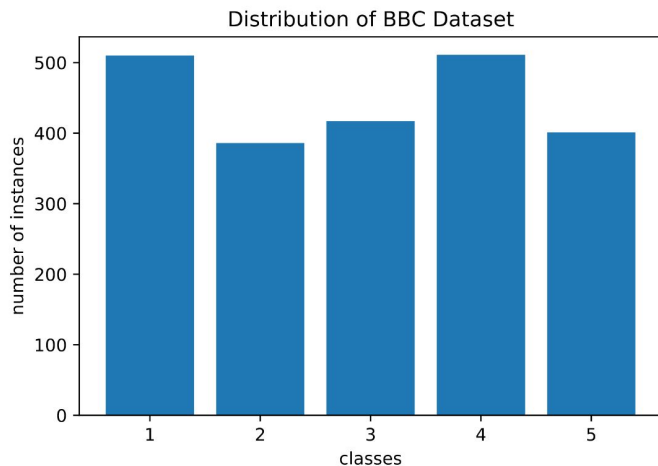
Noah Freger - 40097005

Ashwin Philip - 40096505

Dataset analysis

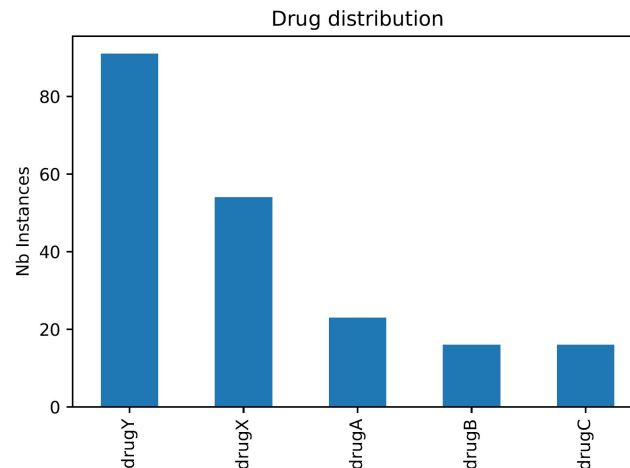
Task 1: BBC

- Balanced data set
 - Evenly distributed data set
- Large data set



Task 2: Drugs

- Unbalanced and small data set
 - Can cause a bias towards the dominant class



Task 1 - Result Analysis

MultinomialNB

- High precision & recall
- High accuracy
- Meaning:
 - Strong Model
 - Well balanced and varied

Task 2 - Result Analysis

Gaussian Naive Bayes

- Low precision for under-represented classes (A,B,C)
- Low recall for over-represented class (Y)
- Meaning:
 - Bias towards the drugY class
 - Dataset's size a factor

Perceptron

- 0 precision and recall for under-represented classes (A,B,C)
 - Even stronger bias towards Y and even X
- Still relatively high accuracy of 72%
 - Hinting that accuracy is not a good metric to evaluate a model when classes are not equally represented

Base Decision Tree

- Perfect precision and recall
- Meaning:
 - The dataset's size and imbalance is not a factor for this model

Top Decision Tree

- Perfect precision and recall
- Chosen hyperparameters almost identical to default ones:

Default (Base)	Best (Top)
Criterion: gini Max depth: none Min samples split: 2	Criterion: gini (from {gini,entropy}) Max depth: 20 (from {20,90}) Min samples split: 2 (from {2,3,4})

Task 2 - Result Analysis (cont.)

Multi-Layered Perceptron

- Similar results as regular Perceptron

Top Multi-Layered Perceptron

- Performance slightly increased from MLP
 - Non zero precision and recall for under-represented (A,B)
 - Still low and drugC has 0 precision with only 1 guess for it

Task Distribution

- Tommy
 - Task 2's data loading and numerical conversion
 - Task 2's distribution figure
 - Task 2's Gaussian NB and Base Decision Tree training
 - Task 2's metrics computation and performance printing
 - Powerpoint
- Noah
 - Task 2's Top-Dt
 - Task 2's Perceptron
 - Task 2's Base-MLP
 - Task 2's Top-MLP
 - Powerpoint
- Ashwin
 - All Task 1
 - Powerpoint