Assignment 2: Implementation and Analysis of Decision Tree and Naïve Bayes algorithms

Submitted by:
Mohd. Sayemul Haque
Roll: 53

Submitted To:

Dr. Chowdhury Farhan Ahmed

Professor

Department of CSE,
University of Dhaka

Implementation:

- Method: Decision tree using gain ratio, Naïve Bayes(Gaussian for Continuos attributes)
- Language: Python
- Libraries Used: *copy,sklearn.metrics,math*
- Processor: Intel Core i5-7300HQ, 8GB Ram, 64bit Operating System(Windows 10)

Datasets:

- Large : Adult,KRK
- Strongly correlated: Cmc
- Moderately correlated: KRK,German
- Hardly correlated: Adult, Iris, Wine, Zoo, Crx

This report contains the performance analysis of Naive_Bayesian and Decision_Tree classification. For that purpose, the algorithms were tested on 8 different datasets. The datasets were divided into a training set(80%) and a testing set(20%) with randomization. The performance analysis was done based on accuracy, precision, recall and F measure. For each dataset a true_class was taken and the measures were taken based on that class.

Both algorithm tends to perform well for datasets that have a very small correlation among features. For example: Adult, Iris, Wine
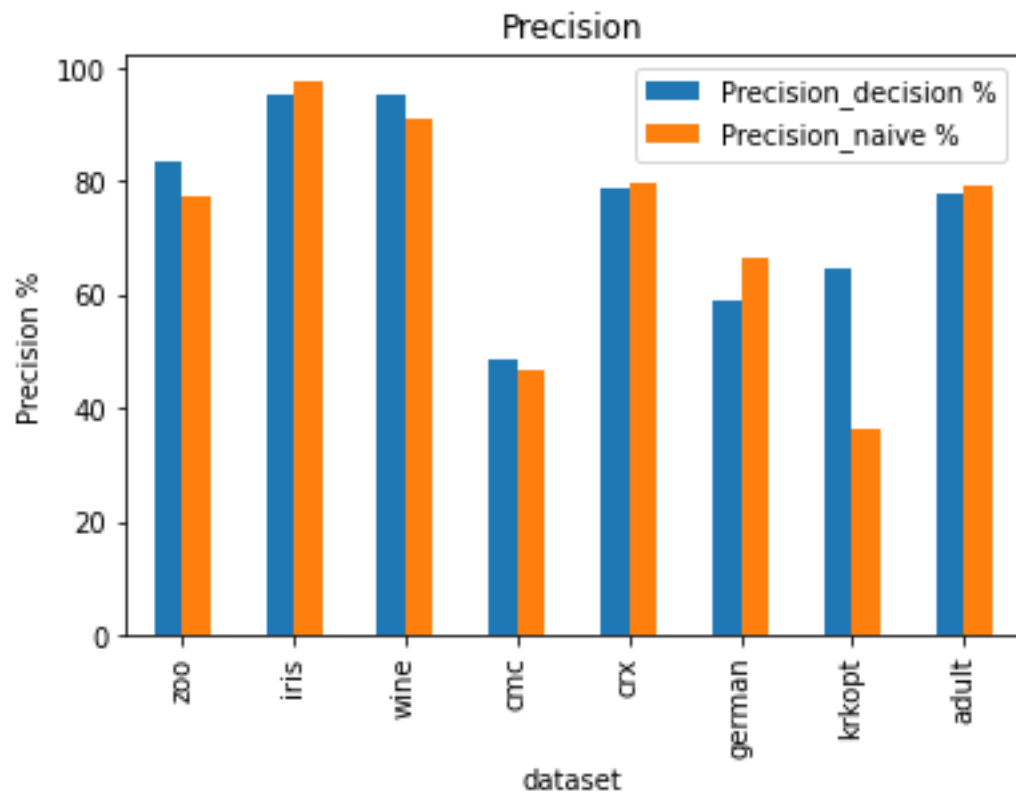
Whereas, for datasets that have a high correlation both algorithm performs poorly.

The comparison of the performences of Decision Tree and Naïve Bayes algorithms with respect to Accuracy, Precision, Recall and F1 score is given below.
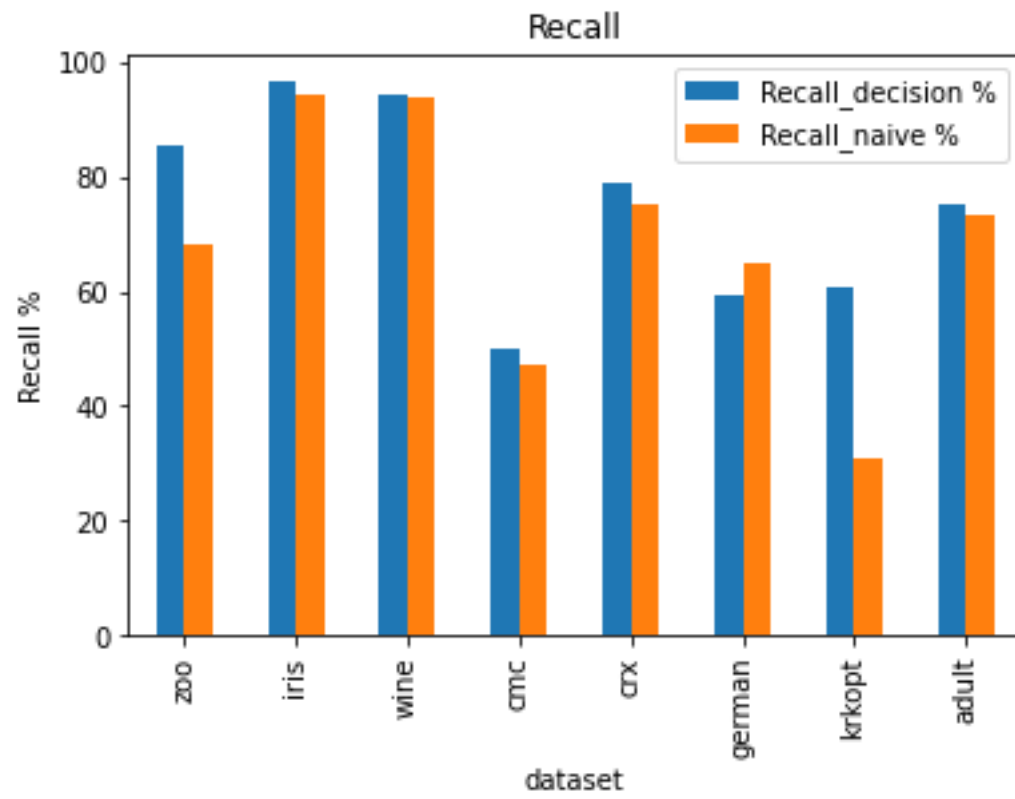
1. **Accuracy:**
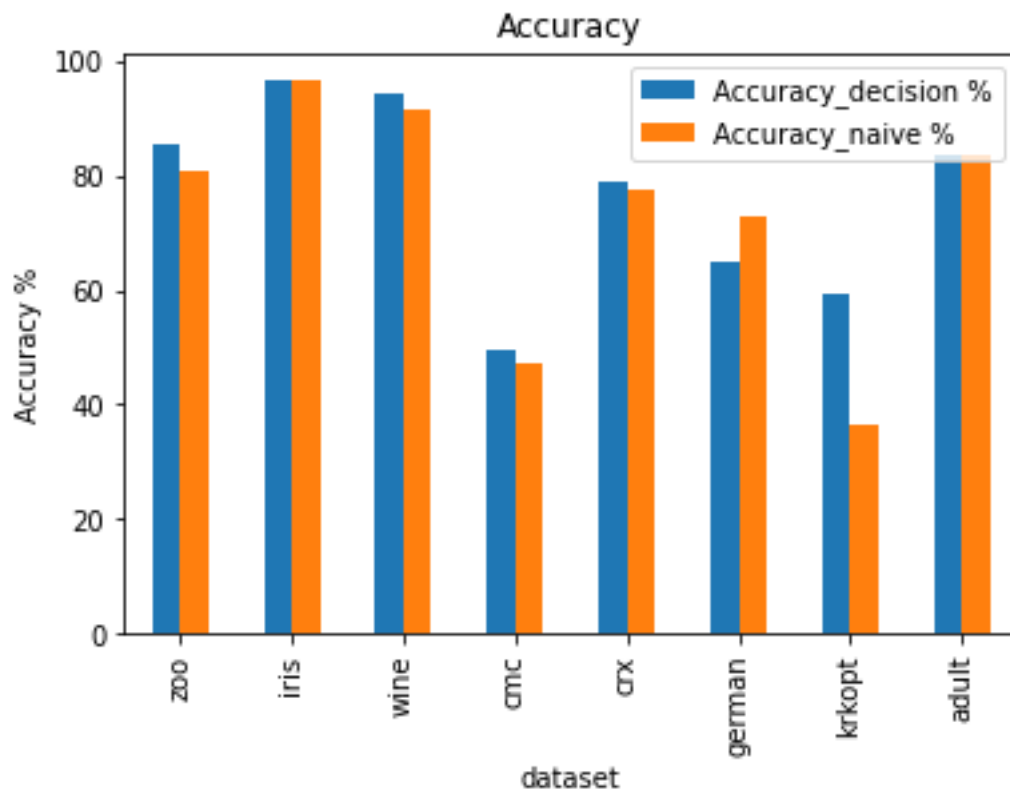
2. **Precision:**

3. Recall:

4. F1 score:

f1