|  |  |
| --- | --- |
| Name: | Prerna Sunil Jadhav |
| Sap  Id: | 60004220127 |
| Class: | T. Y. B.Tech (Computer Engineering) |
| Course: | Data Mining and Warehouse Laboratory |
| Course Code: | DJ19CEL501 |
| Experiment  No.: | 03 |

**AIM:** Implementation of Classification algorithm Using

1. Decision Tree ID3
2. Naïve Bayes algorithm

**THEORY**:

IMPLEMENTATION OF CLASSIFICATION ALGORITHM USING DECISION TREE ID3:

* Decision Tree ID3 is a machine learning algorithm that builds a decision tree for classification by selecting the best attributes to split the data based on information gain.
* It recursively divides the data into subsets until a stopping criterion is met.
* The resulting tree can be used to classify new instances based on their feature values.

IMPLEMENTATION OF CLASSIFICATION ALGORITHM USING NAÏVE BAYES:

* The Naïve Bayes algorithm is a probabilistic classifier that applies Bayes' theorem with the "naïve" assumption that features are conditionally independent.
* It calculates the probability of an instance belonging to a particular class based on the probabilities of its features.
* Naïve Bayes is simple, efficient, and works well with text classification and other applications where independence assumptions hold.

**PROCEDURE:**

IMPORT LIBRARIES:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

from sklearn.datasets import load\_iris

import seaborn as sns

from sklearn.model\_selection import train\_test\_split, GridSearchCV, cross\_val\_score

from sklearn.naive\_bayes import GaussianNB

from sklearn.tree import DecisionTreeClassifier

from sklearn import tree

from sklearn.metrics import accuracy\_score, classification\_report, confusion\_matrix

from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier

from ucimlrepo import fetch\_ucirepo  # For fetching UCI datasets

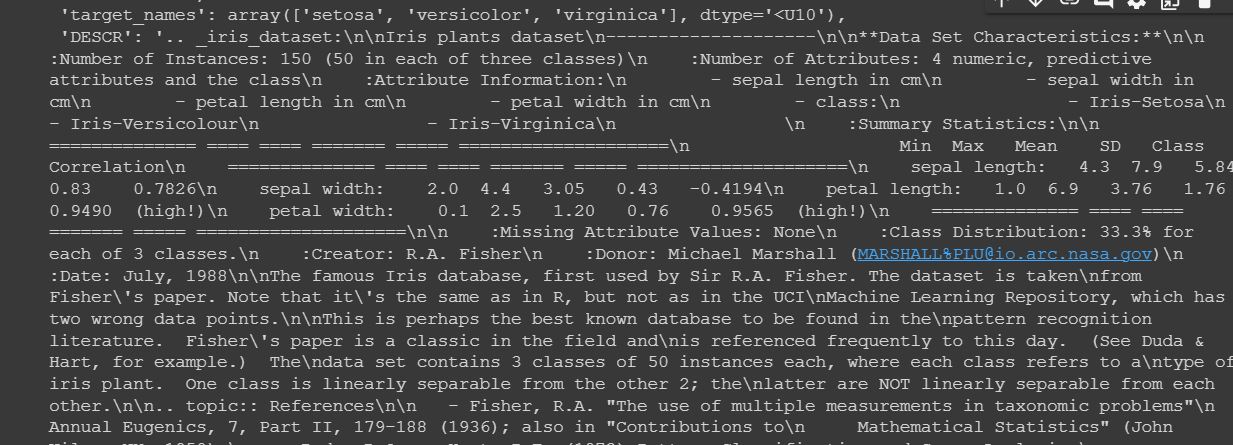
Part A: Applying Gaussian Naive Bayes and Decision Tree Classifier on Dataset

Dataset 1 - Iris Floris Dataset

df1 = load\_iris()

df1

OUTPUT:

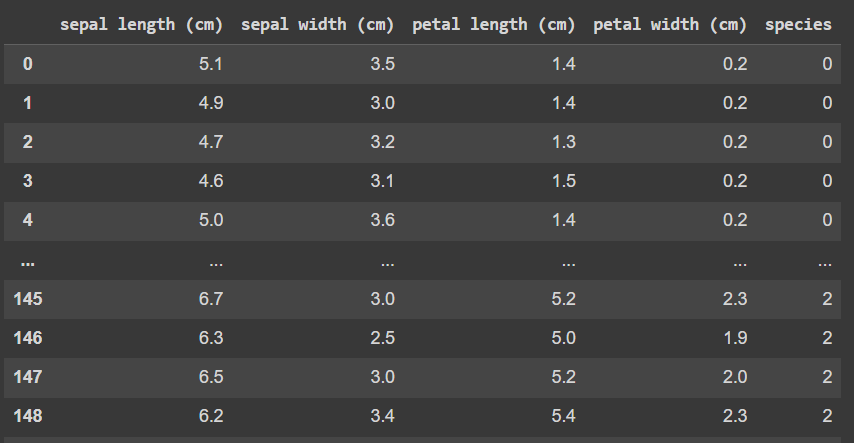


dataset1 = pd.DataFrame(df1.data)

dataset1.columns = df1.feature\_names

dataset1['species'] = df1.target

dataset1

OUTPUT:

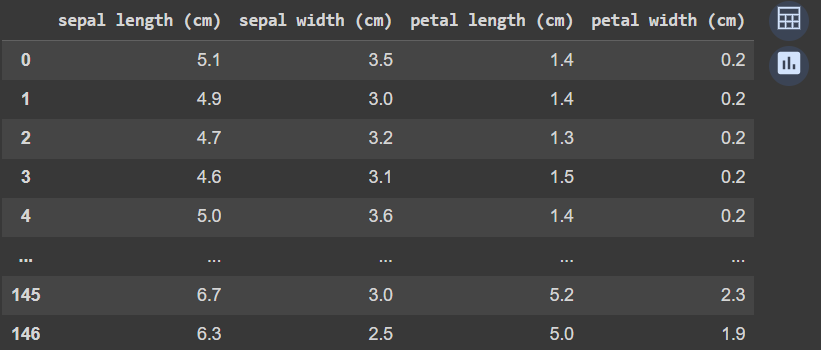
# independent and dependent features

x = dataset1.iloc[:,:-1]

y = dataset1.iloc[:,-1]

x

OUTPUT:



y

OUTPUT:

A screen shot of a computer

Description automatically generated

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.33, random\_state=42)

# Gaussian Naive Bayes Classifier

treemodel = GaussianNB()

treemodel

OUTPUT:



A screen shot of a computer

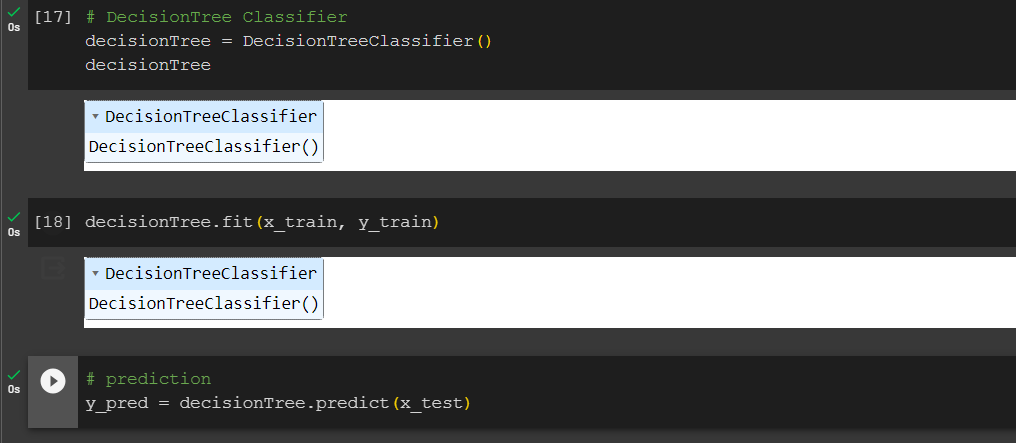
Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

plt.figure(figsize=(5,5))

labels = ['Decision Tree', 'Naive Bayes']

accuracy = [acc\_dt1, acc\_nb1]

plt.bar(labels, accuracy)

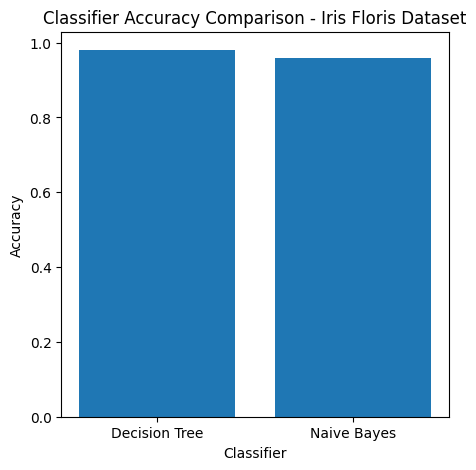
plt.xlabel('Classifier')

plt.ylabel('Accuracy')

plt.title('Classifier Accuracy Comparison - Iris Floris Dataset')

plt.show()

OUTPUT:



Part B: Applying Gaussian Naive Bayes and Decision Tree Classifier on 4 Datasets

Dataset 2 - Wine Dataset

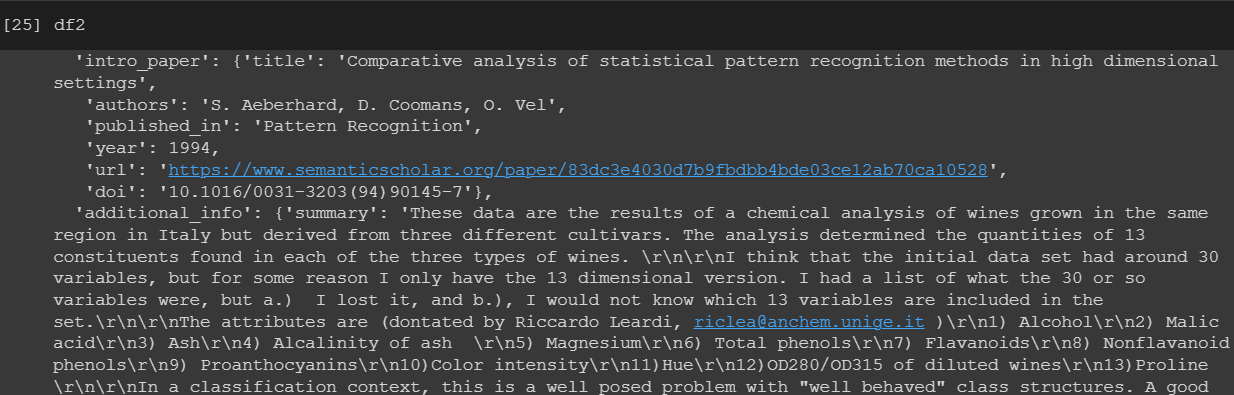
# fetch dataset

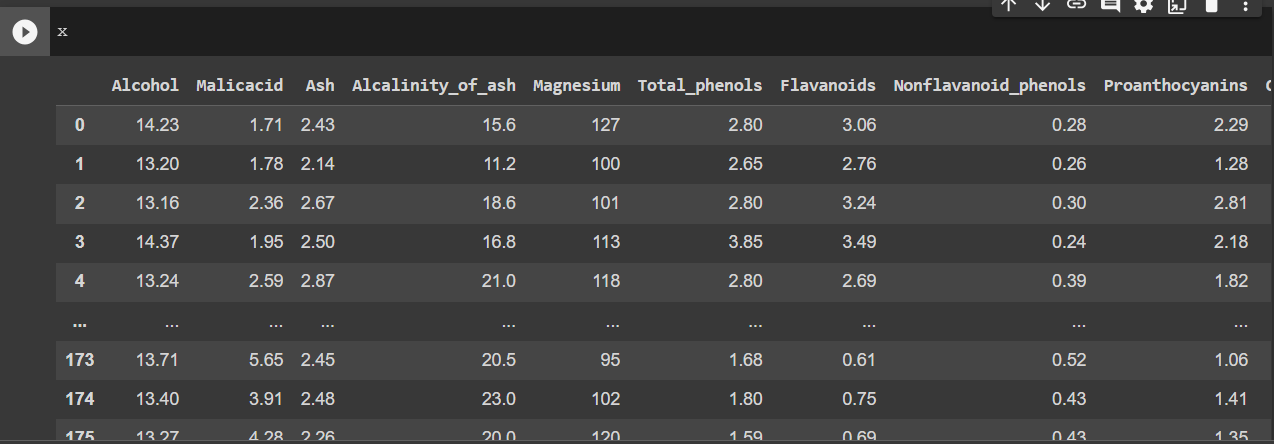
df2 = fetch\_ucirepo(id=109)

# data (as pandas dataframes)

x = df2.data.features

y = df2.data.targets



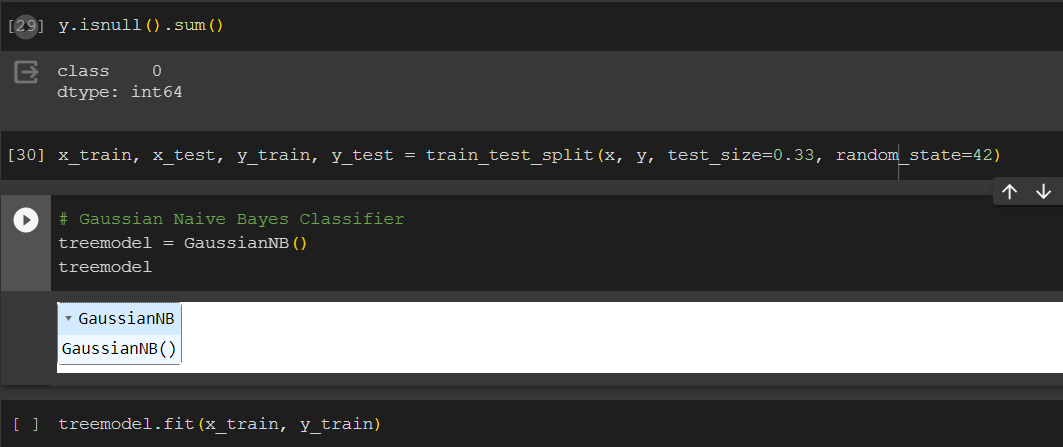


A screenshot of a computer

Description automatically generated

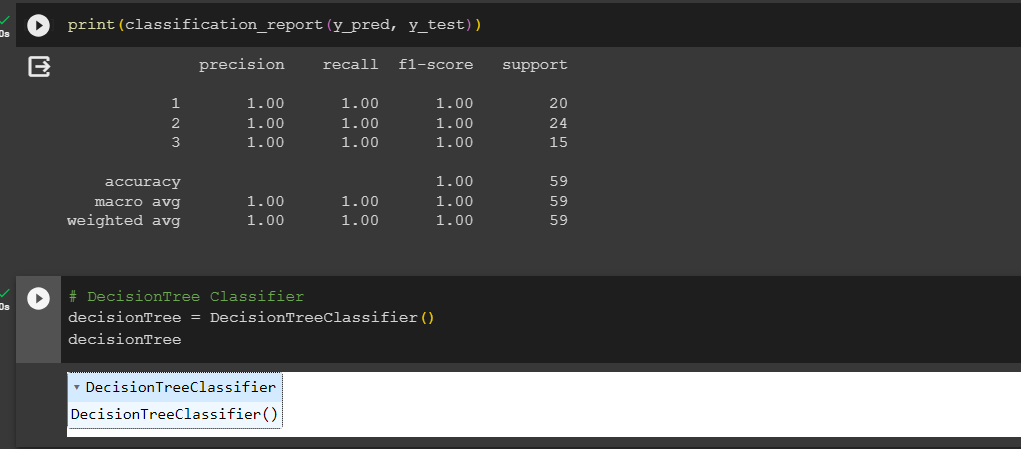
A screen shot of a computer

Description automatically generated



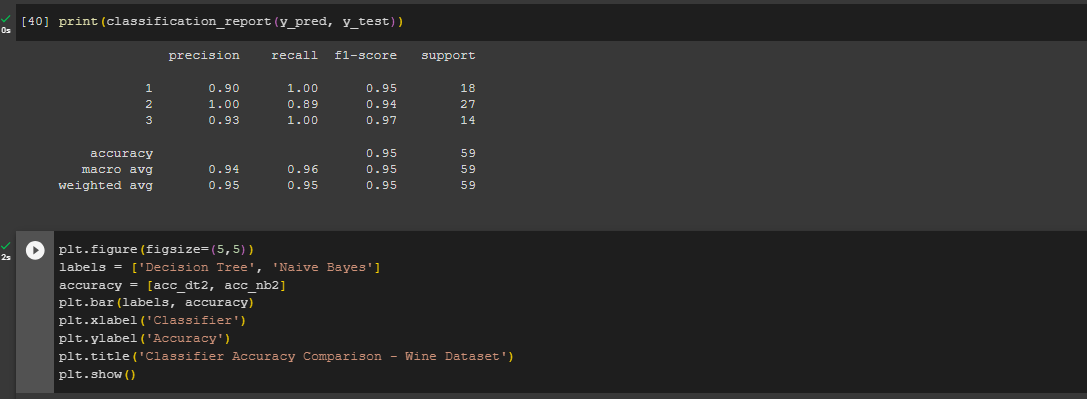
A screenshot of a computer program

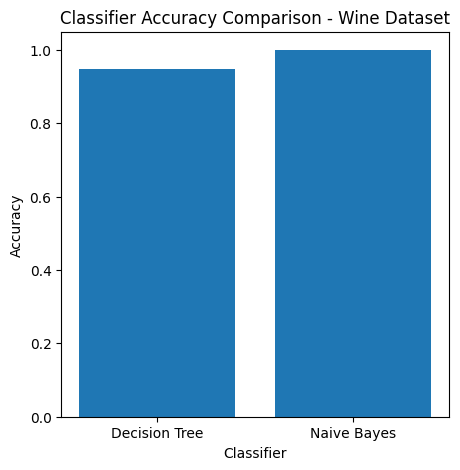
Description automatically generated

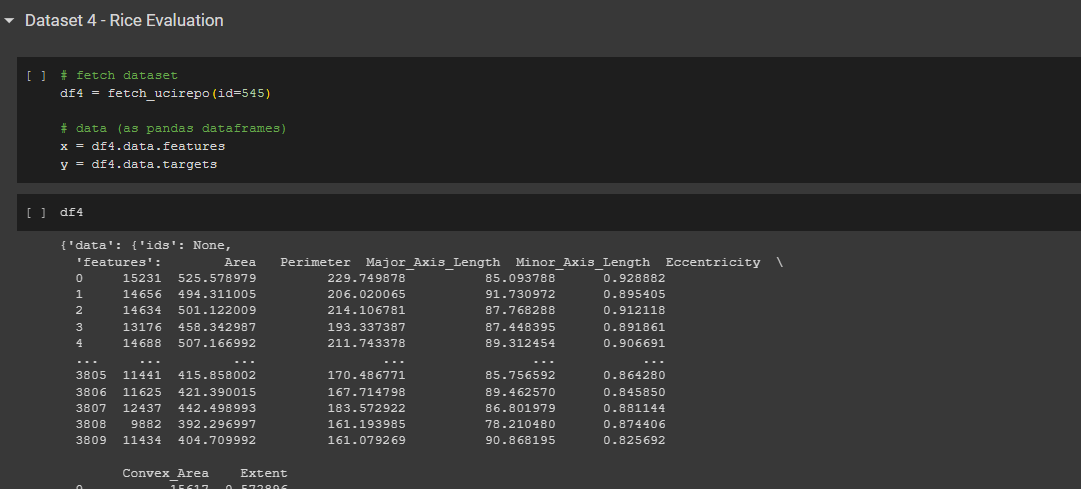


A screenshot of a computer

Description automatically generated







A screenshot of a computer

Description automatically generated

A black screen with a black border

Description automatically generated

A black screen with text

Description automatically generated

A screenshot of a computer

Description automatically generated

A black and white striped object

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph of different colored bars

Description automatically generatedA blue bar graph with white text

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

**CONCLUSION**:

* In conclusion, the implementation of classification algorithms using Decision Tree ID3, and Naïve Bayes has provided effective tools for making predictions and categorizing data.
* Decision Tree ID3 creates a tree-like structure to make decisions, while Naïve Bayes leverages probability and independence assumptions.
* These methods offer versatile options for classification tasks in machine learning.