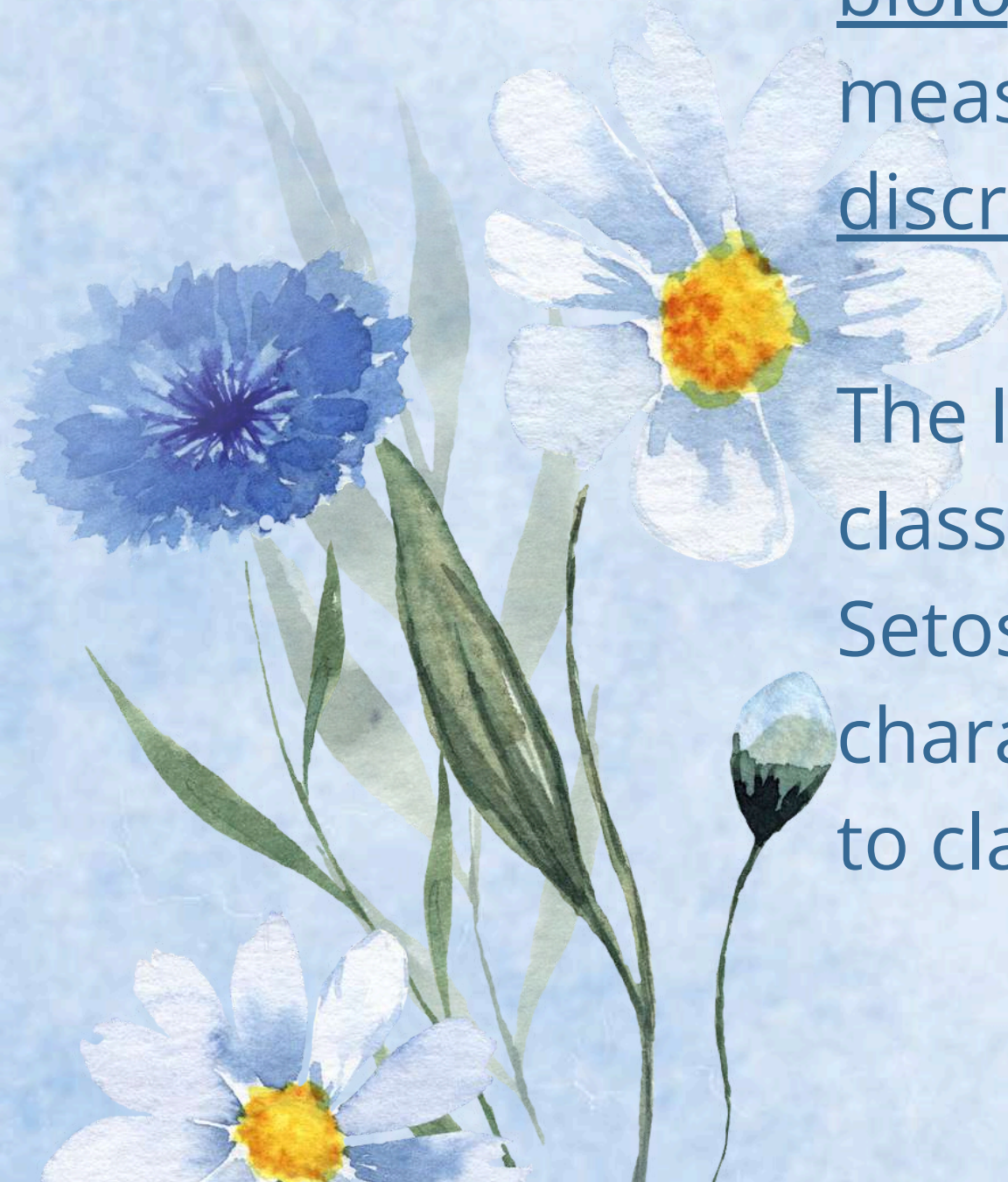# Classification of Iris Using Decision Tree

by Devina Kartika

About

The Iris flower data set or Fisher's Iris data set is a <u>multivariate data set</u> used and made famous by the British <u>statistician</u> and <u>biologist</u> <u>Ronald Fisher</u> in his 1936 paper The use of multiple measurements in taxonomic problems as an example of <u>linear discriminant analysis</u>.

The Iris dataset is a benchmark dataset for machine learning classification. Presented 150 samples from three iris species: Setoso, Versicolor, and Virginica, with four distinguishing characteristics (sepal and petal dimensions). This project aims to classify iris species with Decision Tree
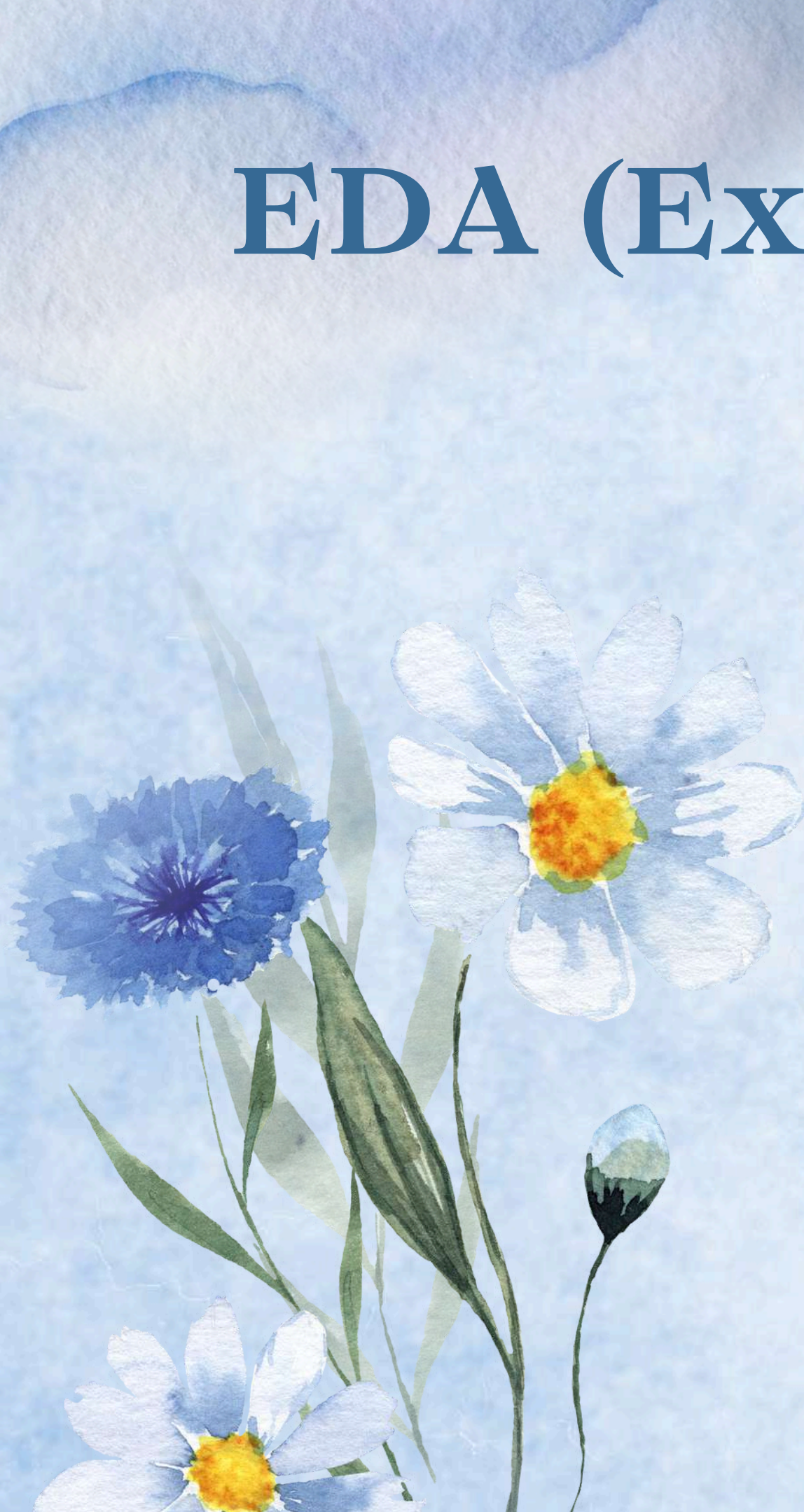
Tools and Library

EDA (Exploratory Data Analysis)

Load Dataset

The data dimensions of the feature, there are 150 rows and 4 columns and the target label classes are 3 classes, namely, 0, 1, and 2

```python
from sklearn.datasets import load_iris

X, y = load_iris(return_X_y=True)

print(f'Dimensi Feature: {X.shape}')
print(f'Clas: {set(y)}')
```

```
Dimensi Feature: (150, 4)
Clas: {0, 1, 2}
```

Training & Testing Set

```python
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,
                                                    y,
                                                    test_size=0.4,
                                                    random_state=0)
```

Classification with Decision Trees

```python
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier(max_depth=4)

model.fit(X_train, y_train)
```

Decision Tree Classifier to perform classification of iris species with parameters into a maximum of 4 layers

DecisionTreeClassifier

DecisionTreeClassifier(max_depth=4)

Model Visualization

```python
import matplotlib.pyplot as plt
from sklearn import tree

plt.rcParams['figure.dpi'] = 90
plt.subplots(figsize=(10, 10))
tree.plot_tree(model, fontsize=8)
plt.show()
```
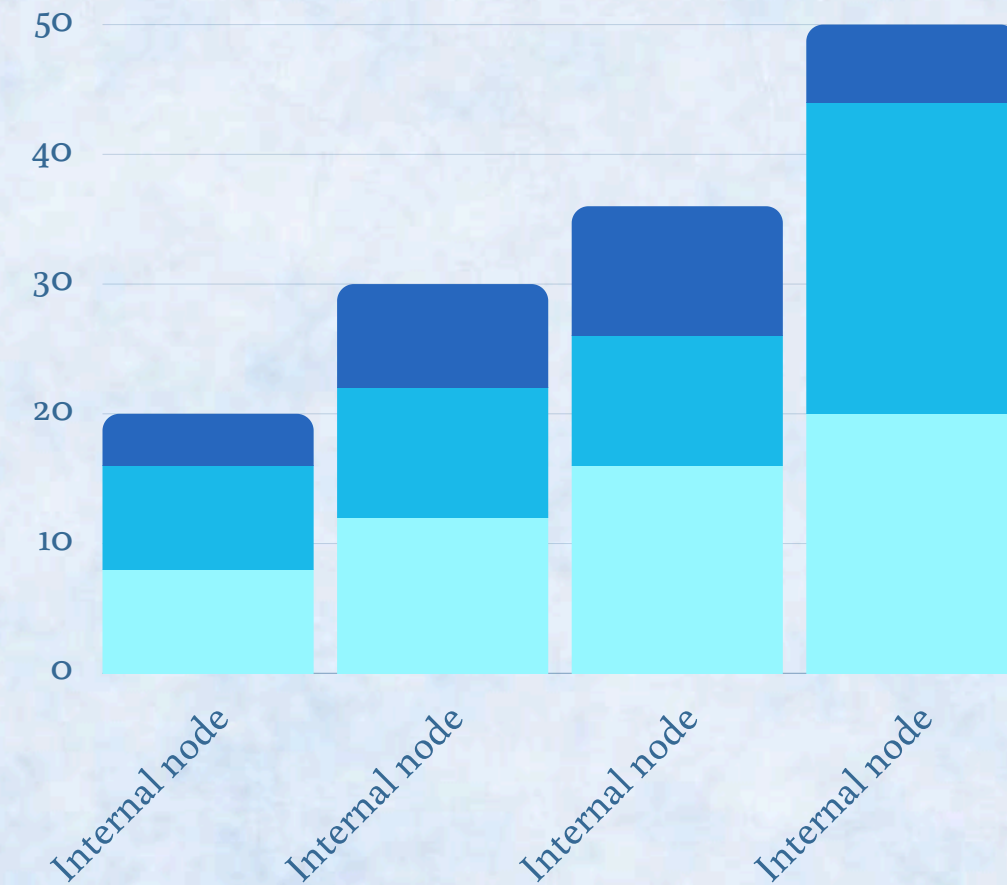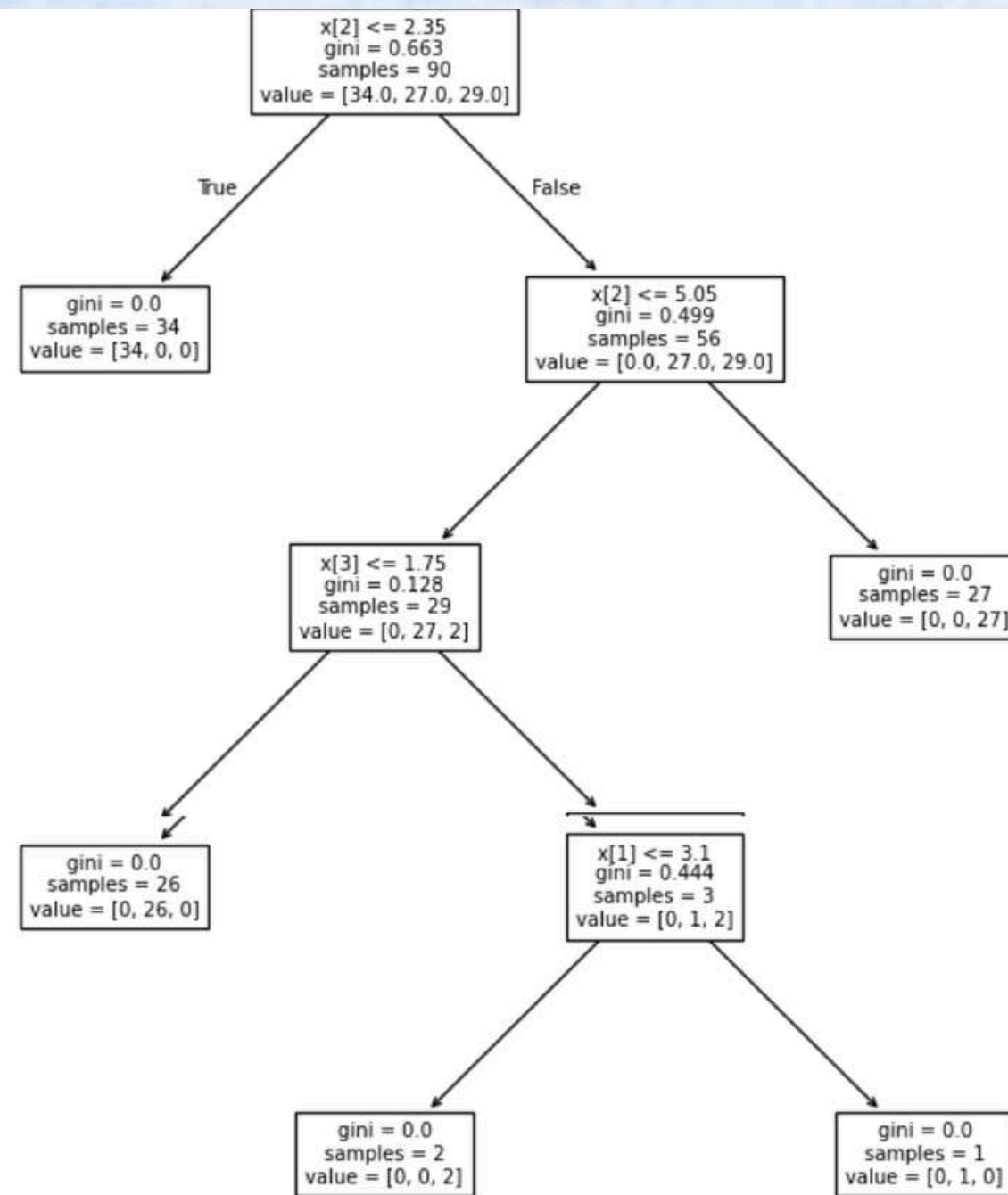
There are two modules for importing, namely
matplotlib and sklearn import tree.
The figure dpi is 90 with a figure size of 10 x 10

Visualization Results From the Decision Tree

x[2] <= 2.35
gini = 0.663
samples = 90
value = [34.0, 27.0, 29.0]

True

False

gini = 0.0
samples = 34
value = [34, 0, 0]

x[2] <= 5.05
gini = 0.499
samples = 56
value = [0.0, 27.0, 29.0]

x[3] <= 1.75
gini = 0.128
samples = 29
value = [0, 27, 2]

gini = 0.0
samples = 27
value = [0, 0, 27]

gini = 0.0
samples = 26
value = [0, 26, 0]

x[1] <= 3.1
gini = 0.444
samples = 3
value = [0, 1, 2]

gini = 0.0
samples = 2
value = [0, 0, 2]

gini = 0.0
samples = 1
value = [0, 1, 0]

Internal node   Internal node   Internal node   Internal node

The route node is at the top, then there are 5 leaf nodes, there are 4 internal nodes and each node contains information from Iris Classification with Decision Tree

Model Evaluation

```python
from sklearn.metrics import classification_report

y_pred = model.predict(X_test)

print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        16
           1       0.92      0.96      0.94        23
           2       0.95      0.90      0.93        21


    accuracy                           0.95        60
   macro avg       0.96      0.95      0.95        60
weighted avg       0.95      0.95      0.95        60
```

The evaluation model uses a classification
report with an accuracy value of 95

Thank You