Digital Skill Fair 35.0 Dibimbing

IRIS CLASSIFICATION USING DECISION TREE

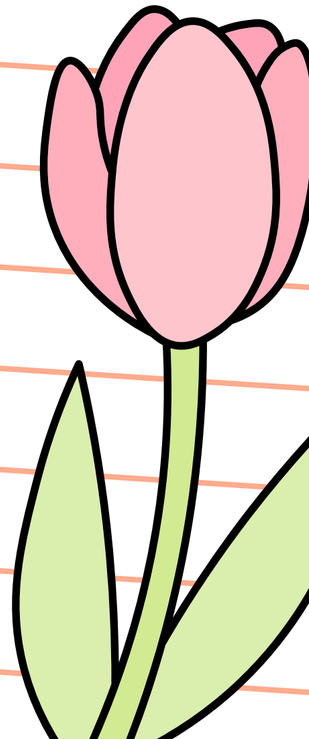
by Septiana Aulia Nur Fadlina

ABOUT IRIS DATASET

The Iris dataset, a collection of 150 samples within the Scikit-learn library representing three species of Iris flowers (Iris setosa, Iris versicolor, and Iris virginica). Each species consist of 50 samples. Each sample includes four numerical measurements:

- Sepal length (cm)
- Sepal width (cm)
- Petal length (cm)
- Petal width (cm)

These measurements capture the physical dimensions of the Iris flowers, allowing for the identification of morphological differences between the species.



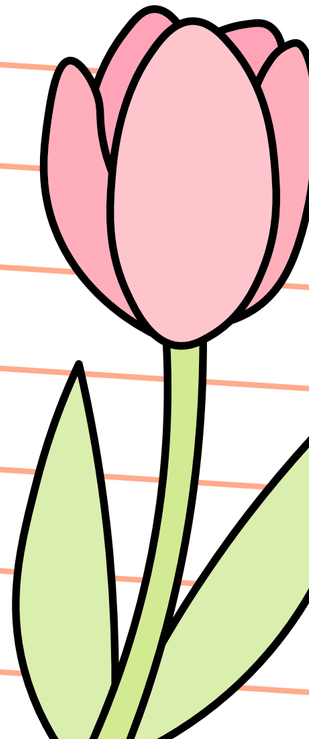
PROJECT GOALS

Build a machine learning model to classify Iris flowers into three species (Iris setosa, Iris versicolor, and Iris virginica) using Decision Tree.

TOOLS USED

pandas

matplotlib



IRIS DATASET

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

150 samples & 5 features

species	count
0	50
1	50
2	50

50 samples/ species

IRIS DATASET

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   sepal length (cm)  150 non-null    float64
 1   sepal width (cm)   150 non-null    float64
 2   petal length (cm)  150 non-null    float64
 3   petal width (cm)   150 non-null    float64
 4   species            150 non-null    int64  
dtypes: float64(4), int64(1)
memory usage: 6.0 KB
```

No missing value detected

EXPLORATORY DATA ANALYSIS

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333	1.000000
std	0.828066	0.435866	1.765298	0.762238	0.819232
min	4.300000	2.000000	1.000000	0.100000	0.000000
25%	5.100000	2.800000	1.600000	0.300000	0.000000
50%	5.800000	3.000000	4.350000	1.300000	1.000000
75%	6.400000	3.300000	5.100000	1.800000	2.000000
max	7.900000	4.400000	6.900000	2.500000	2.000000

SPLIT DATA & TRAIN MODEL

SPLIT DATA

Test size = 20%

```
Number of train data: 120
Number of testing data: 30
```

TRAIN MODEL

Random state = 100

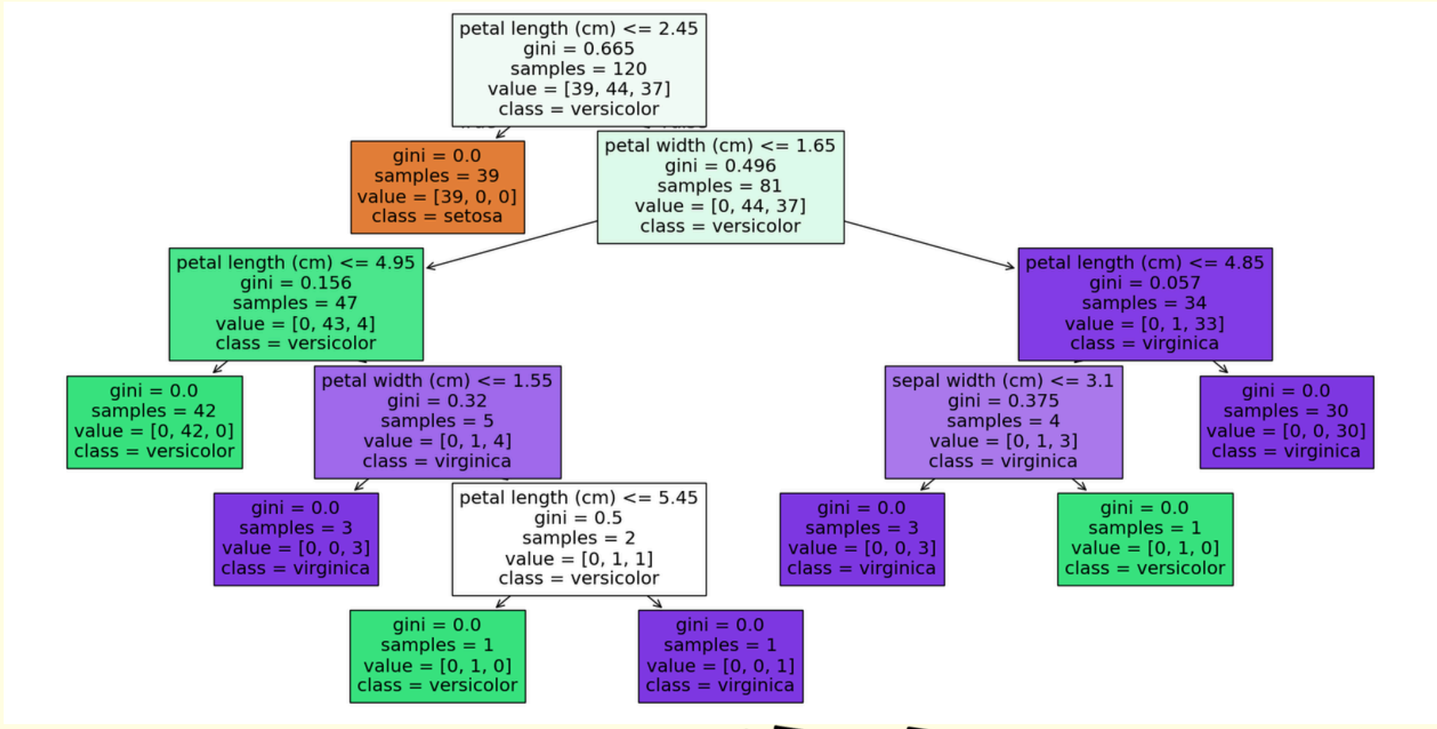
```
DecisionTreeClassifier
DecisionTreeClassifier(random_state=100)
```

MODEL ACCURACY

Accuracy: 96.67%

The model accuracy is good enough

VISUALIZATION



CONCLUSION

IRIS SETOSA

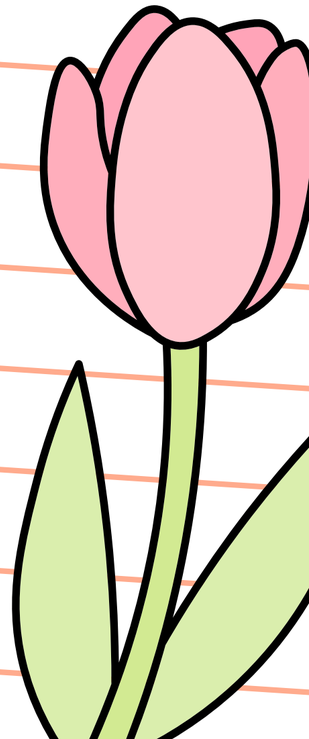
- Generally have the shortest petal length.
- Have unique characteristics that make it easy to distinguish.

IRIS VERSICOLOR

- The petals length are longer than Iris setosa, but shorter than Iris virginica.
- The width of petals and sepals has a typical value range.

IRIS VIRGINICA

- Generally have the biggest petal length and width.
- Often have the widest sepals.



THANK YOU

LET'S WORK TOGETHER!

Bandung, Indonesia

septianaaulianf@gmail.com

septianaanf