







# SIMPLE CLASSIFCATION OF IRIS DATASET

### USING MACHINE LEARNING



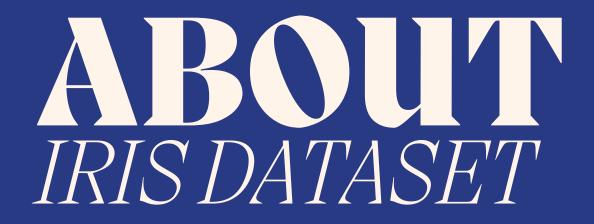


BY

NI PUTU ESSA KIRANA PRAYASCITA







The Iris dataset is a widely used dataset from the Scikit-learn library, consisting of 150 samples representing three species of Iris flowers: Iris setosa, Iris versicolor, and Iris virginica. Each sample includes four numerical measurement features:

- Sepal length (in centimeters)
- Sepal width (in centimeters)
- Petal length (in centimeters)
- Petal width (in centimeters)

These features capture the physical dimensions of the Iris flowers, aiming to highlight the morphological differences between the species for exploring patterns and predict the relationships in data.





### ATTRIBUTES

#### Number of attributes:

4 numerical attributes used as features to predict the target variable

### INFORMATIONS

#### **Attributes Characteristic:**

Sepal Length (cm), Sepal Width (cm), Petal Length (cm), Petal Width (cm), and the target variable, which is the species of the flower (Iris setosa, Iris versicolor, and Iris virginica). Each species consist of 50 samples.





# EXPLORATORY DATA ANALYSIS (EDA)

<del></del>	Se	epal 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	

**IRIS DATAFRAME** 

[ <del>*</del> ]		sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species	
	count	150.000000	150.000000	150.000000	150.000000	150.000000	11.
	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	





# EXPLORATORY DATA ANALYSIS (EDA)

#### DATAFRAME INFORMATIONS

```
Target unique info: array([0, 1, 2])
species
0 50
1 50
2 50
Name: count, dtype: int64
```

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





# DATA PRE-PROCESSING

```
Sepal length (cm) 0
sepal width (cm) 0
petal length (cm) 0
petal width (cm) 0
species 0
```

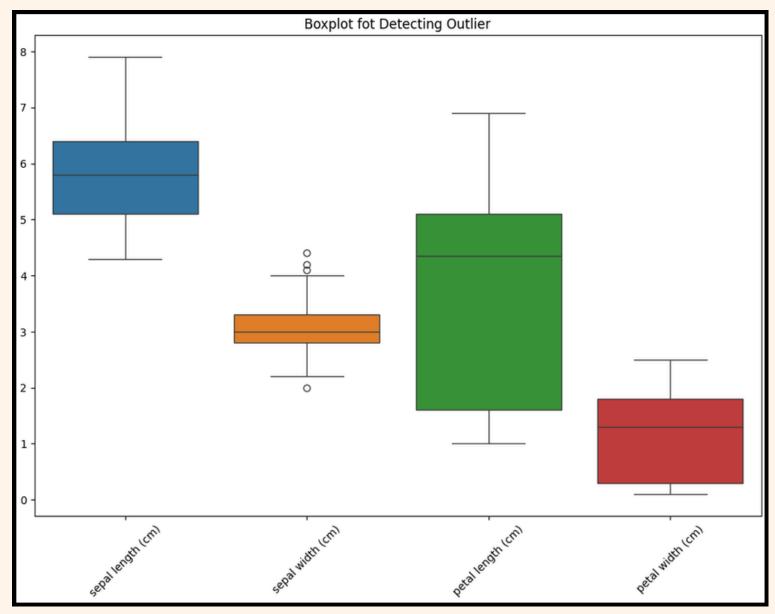
```
Outlier in feature 'sepal length (cm)':
No outlier detected.
Outlier in feature 'sepal width (cm)':
    sepal length (cm) sepal width (cm) petal length (cm) petal width (cm)
                  5.7
                                                       1.5
                                                                         0.4
                  5.2
                                    4.1
                                                       1.5
                                                                         0.1
33
                  5.5
                                    4.2
                                                       1.4
                                                                         0.2
                  5.0
                                    2.0
                                                       3.5
                                                                         1.0
    species
33
Outlier in feature 'petal length (cm)':
No outlier detected.
Outlier in feature 'petal width (cm)':
No outlier detected.
```





#### **OUTLIER VISUALIZATION**

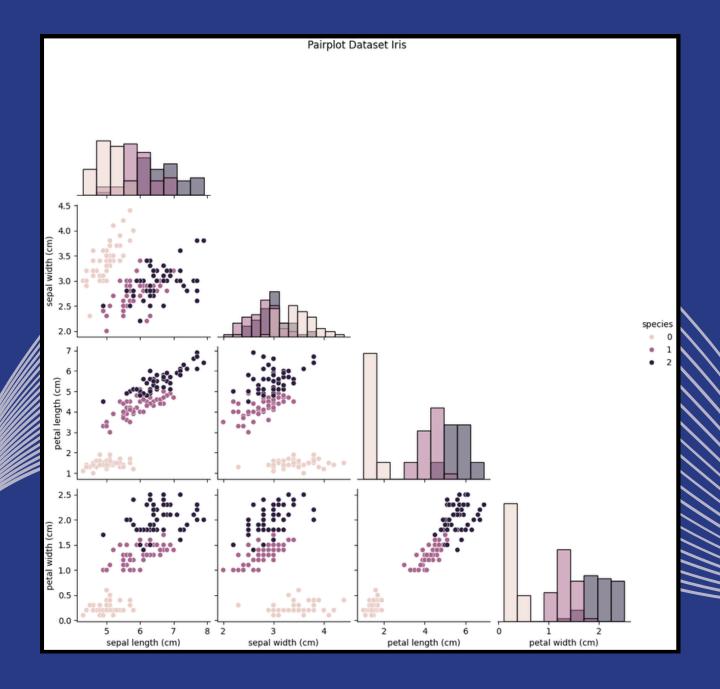


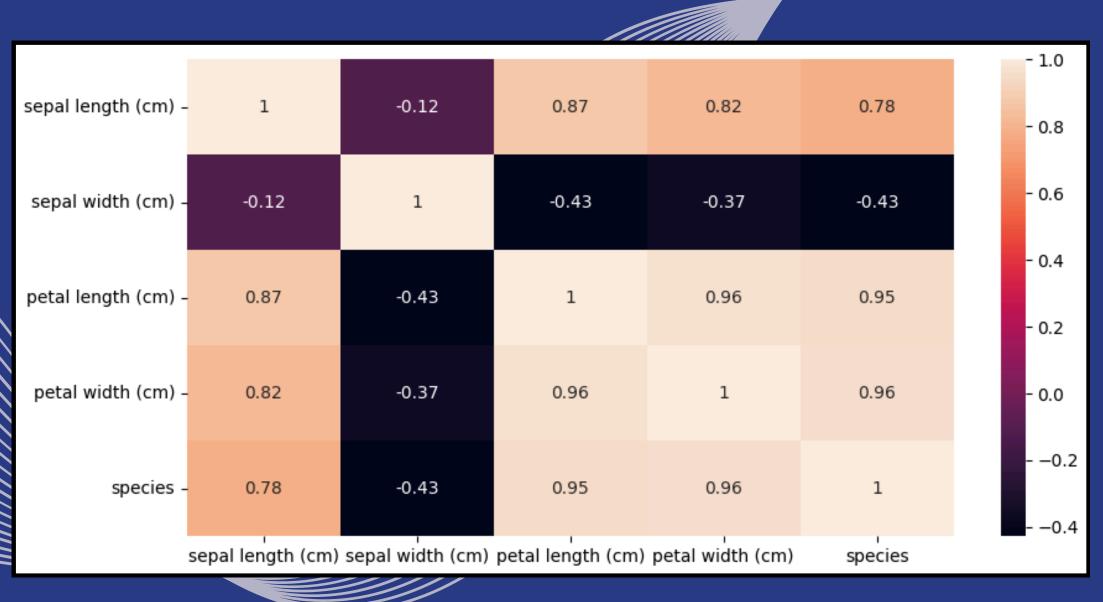






# DATA VISUALIZATION











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

```
Train data after scaling:
   sepal length (cm) sepal width (cm) petal length (cm) petal width (cm)
           -1.473937
                                                 -1.562535
                                                                   -1.312603
                              1.203658
           -0.133071
                              2,992376
                                                 -1.276006
                                                                   -1.045633
            1.085898
                              0.085709
                                                  0.385858
                                                                    0.289218
           -1.230143
                              0.756479
                                                 -1.218701
                                                                   -1.312603
           -1.717731
                                                 -1.390618
                               0.309299
                                                                   -1.312603
Testing data after scaling:
   sepal length (cm) sepal width (cm)
                                         petal length (cm)
                                                            petal width (cm)
                                                                    0.022248
            0.354517
                              -0.585060
                                                  0.557775
                              1.650837
                                                                   -1.179118
           -0.133071
                                                 -1.161395
            2.304867
                              -1.032239
                                                  1.818500
                                                                    1.490583
            0.232620
                              -0.361470
                                                  0.443164
                                                                    0.422703
            1.207795
                              -0.585060
                                                  0.615081
                                                                    0.289218
```





# MODEL ACCURACY



 $\Diamond$ 

Classification report SVM Model: Accuracy: 96.67%



Calssification Report KNN Model: Accuracy: 100.00%



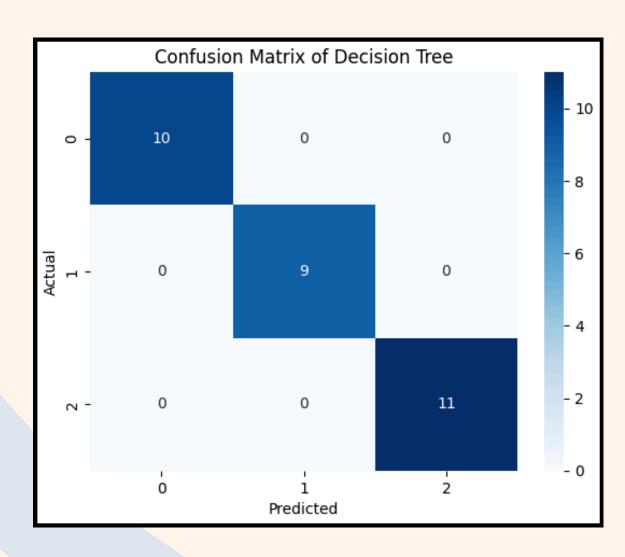
Classification Report Decision Tree Model: Accuracy: 100.00%

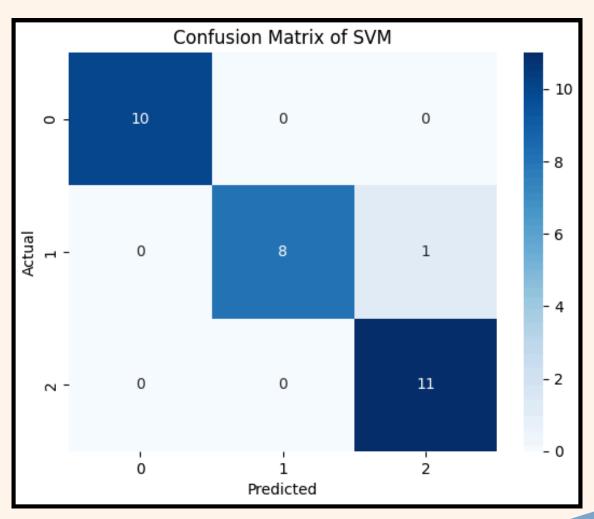


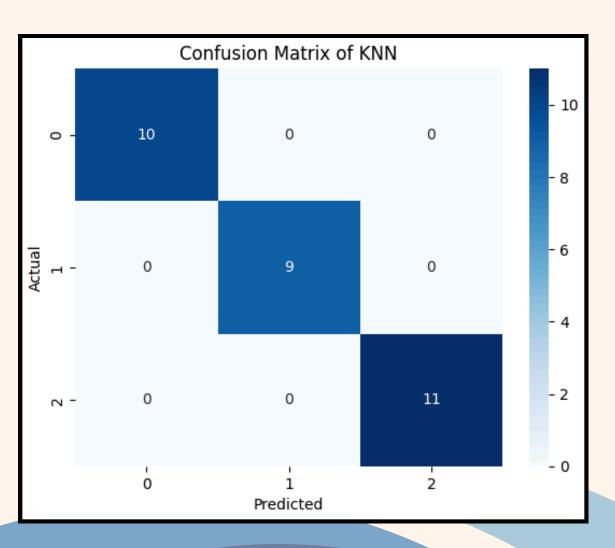




## MODEL VISUALIZATION COMPARISON









# CONCLUSION

Although KNN and Decision Tree achieved perfect accuracy (100%), the SVM model (96.67%) is chosen as the best model. SVM model has better generalization ability and resistance to overfitting, making it more stable and reliable for new data.

# GETIN TOUCH

EMAIL essakiranaprayascita@gmail.com

http://linkedin.com/in/ni-putu-essa-kiranaprayascita-3930a2287

GITHUB https://github.com/essakirana



LINKEDIN