

Decision Tree

Importation des libs :

```
In [1]: from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn import tree
import matplotlib.pyplot as plt
```

Chargement de la base de données iris et préparation:

```
In [2]: # Step 1: Load the Iris dataset
iris = datasets.load_iris()

# Step 2: Prepare the data
X = iris.data # Features (sepal length, sepal width, petal length, petal width)
y = iris.target # Target Labels (species)
```

Séparation en train/test

```
In [3]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0) # 80% entraînement, 20% test
```

Entraînement de l'arbre de décision

```
In [4]: clf = DecisionTreeClassifier(random_state=42)
clf.fit(X_train, y_train)
```

```
Out[4]: DecisionTreeClassifier
DecisionTreeClassifier(random_state=42)
```

Evaluation du modèle à l'aide de l'accuracy

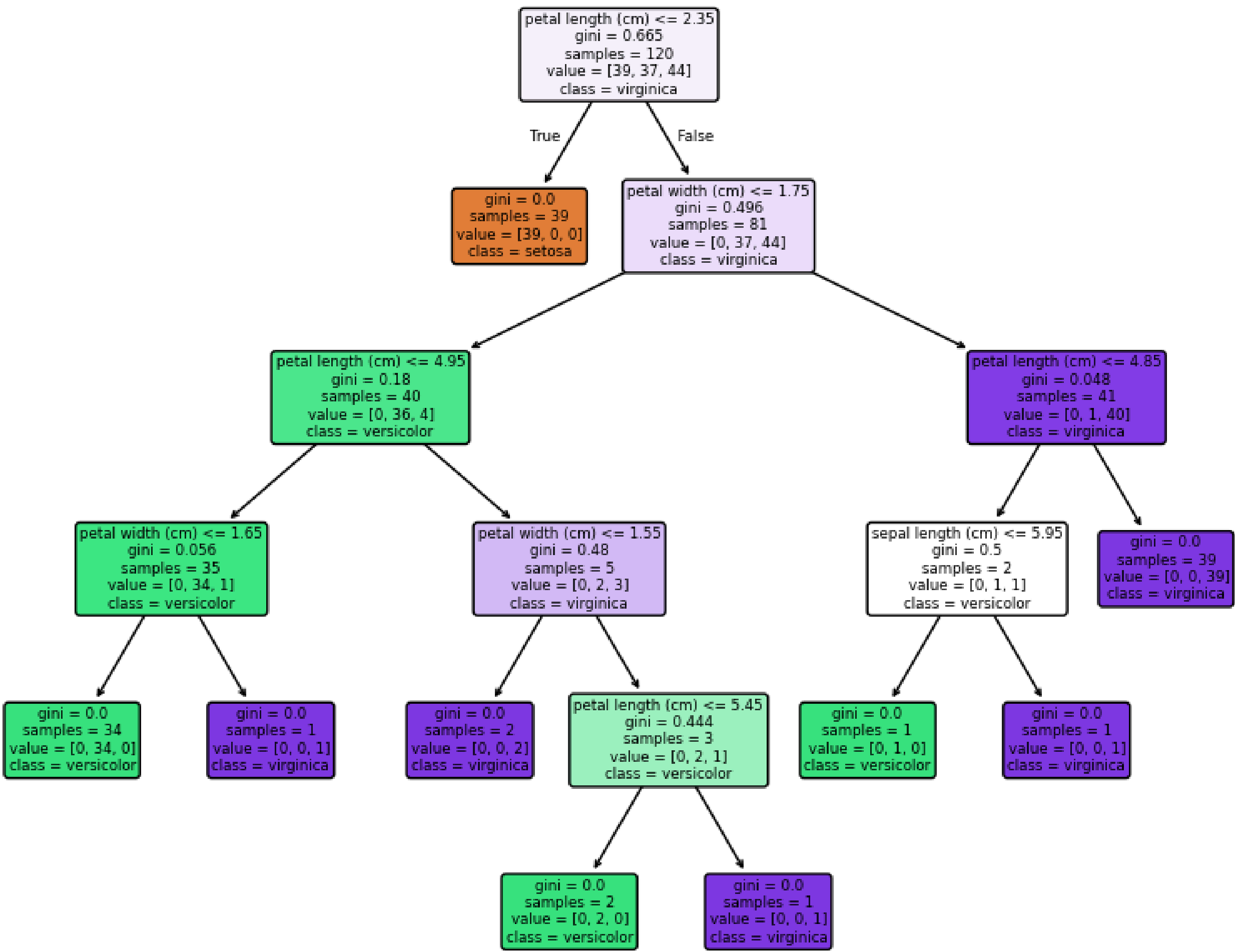
```
In [5]: y_pred = clf.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy of the Decision Tree Classifier: {accuracy * 100:.2f}%")
```

Accuracy of the Decision Tree Classifier: 100.00%

Affichage de l'arbre de décision

```
In [6]: plt.figure(figsize=(10, 8))
tree.plot_tree(clf, filled=True, feature_names=iris.feature_names, class_names=iris.target_names, rounded=True)
plt.show()
```



Essai d'une classification

```
In [7]: new_sample = [[5.5, 2.4, 3.8, 1.1]] # Example: sepal Length = 5.5, sepal width = 2.4, petal Length = 3.8, petal width = 1.1

# Predict the class for the new sample
predicted_class = clf.predict(new_sample)

# Print the predicted class and corresponding species name
print(f"The predicted class for the new sample is: {iris.target_names[predicted_class][0]}")
```

The predicted class for the new sample is: versicolor

Conclusion

Dans le cas de notre arbre de décision, il n'y a aucun paramètre que l'on peut modifier car nous avons déjà une accuracy de 100% dès le départ. Cependant, cela veut aussi dire que l'on pourra classifier de façon sûr n'importe quelle donnée.