DECISION TREE

-----A MACHINE LEARNING ALGORITHM
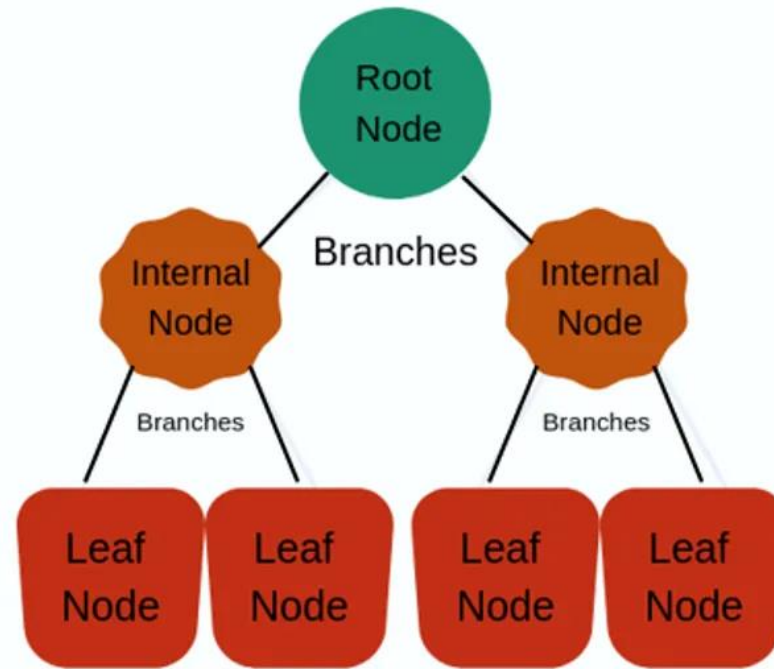
INTRODUCTION

What is Decision Tree?????

➢ Supervised learning algorithm used for classification and regression problem.

➢ Tree like model.

➢ Structure:- 4 key components.

- Root Node

- Internal Node

- Branches

- Leaf Node

STRUCTURE OF DECISION TREE

HOW A DECISION TREE WORK?????

- Decision tree operate by breaking the data into subsets on behalf of the best feature at each step.
- Splitting Criteria:-

    - Gini Impurity.

    - Entropy.

    - Variance Reduction.
- The decision Tree splits the data till the data is pure.

ENTROPY VS GINI

➢ Both measure impurity and helps decision making.

➢ Entropy is calculated on behalf of theoretic information.

➢ Entropy is more sensitive to splits.

➢ Gini less sensitive.

➢ Entropy takes more time .

➢ Gini less time.

DECISION TREE DATA PROCESSING

- ➤ Loading and processing of libraries.

```python
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn import tree
```

➢ Selection and loading of dataset.

- Iris Data

```
iris = load_iris()

iris.data

array([[5.1, 3.5, 1.4, 0.2],
       [4.9, 3. , 1.4, 0.2],
       [4.7, 3.2, 1.3, 0.2],
       [4.6, 3.1, 1.5, 0.2],
       [5. , 3.6, 1.4, 0.2],
       [5.4, 3.9, 1.7, 0.4],
       [4.6, 3.4, 1.4, 0.3],
       [5. , 3.4, 1.5, 0.2],
       [4.4, 2.9, 1.4, 0.2],
       [4.9, 3.1, 1.5, 0.1],
       [5.4, 3.7, 1.5, 0.2],
       [4.8, 3.4, 1.6, 0.2],
       [4.8, 3. , 1.4, 0.1],
       [4.3, 3. , 1.1, 0.1],
       [5.8, 4. , 1.2, 0.2],
       [5.7, 4.4, 1.5, 0.4],
       [5.4, 3.9, 1.3, 0.4],
       [5.1, 3.5, 1.4, 0.3],
```
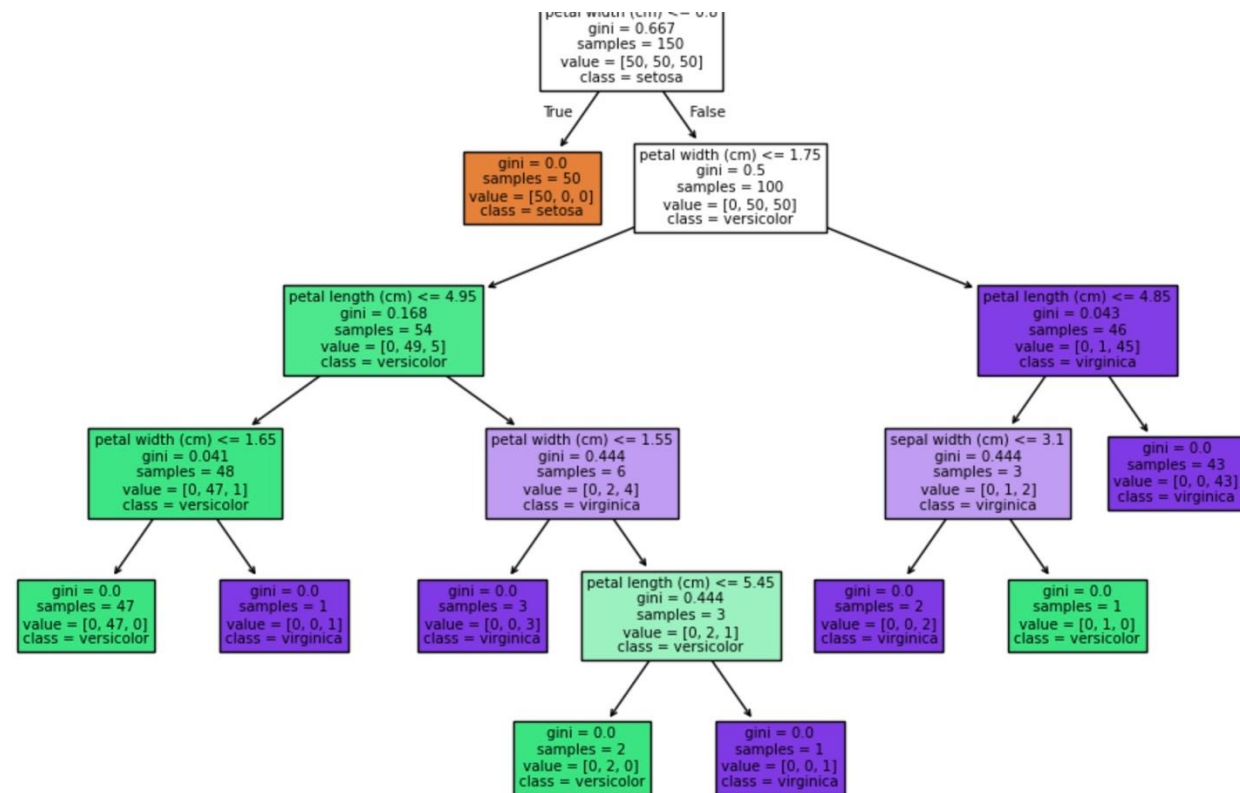
➢ Decision Tree for Classifier.

```
classifier = DecisionTreeClassifier()
classifier.fit(iris.data, iris.target)
```

```
▼   DecisionTreeClassifier ⓘ ❓

DecisionTreeClassifier()
```

```
plt.figure(figsize=(12, 8))
plot_tree(classifier, filled=True, feature_names=iris.feature_names, class_names=iris.target_names)
plt.show()
```
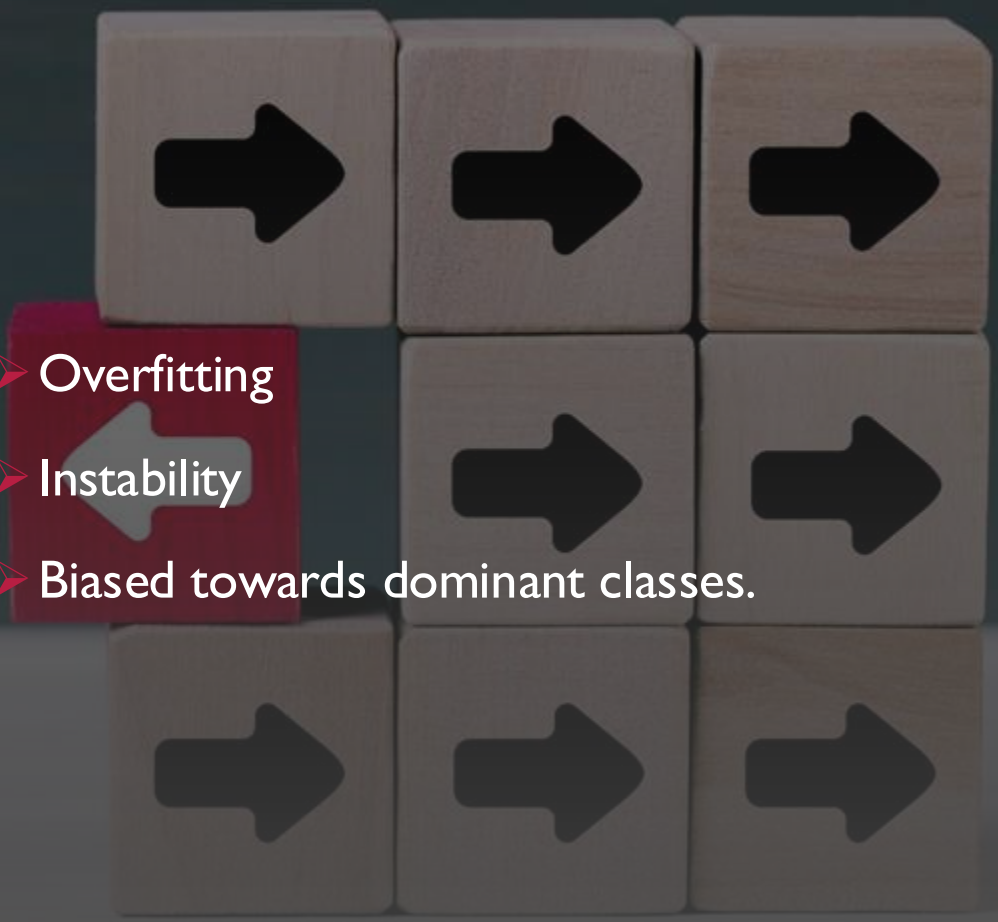
RESULT

ADVANTAGES

➢ Easy to understand and interpret.

➢ No need of feature scaling.

➢ Handles both numerical and categorical data.

➢ Robust to outliers.

DIS-ADVANTAGES

- Overfitting
- Instability
- Biased towards dominant classes.

CONCLUSION

- One of the most widely used and user-friendly machine learning algorithm, decision tree are generally utilized for problems involving regression and classification. Here iris dataset is used for plotting decision tree.