Decision Tree

General Assembly Data Science Immersive - Sample Lesson
Shilpa

Lesson Objective

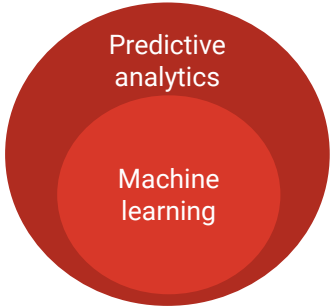
What we will accomplish:

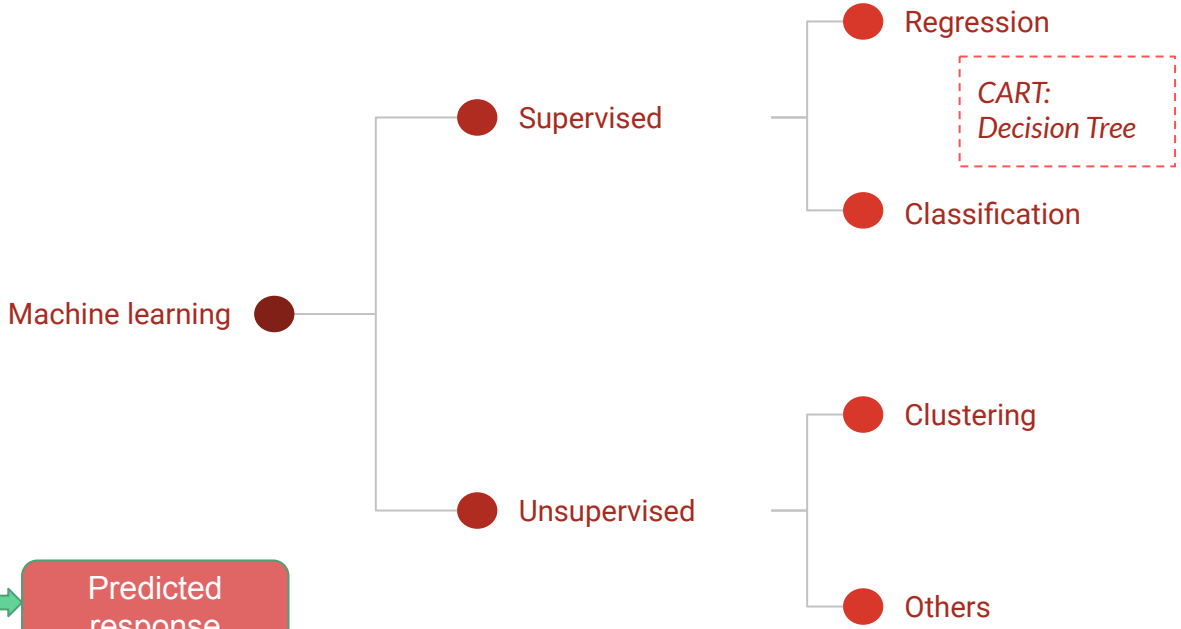
- Gain understanding of Decision Tree and application in machine learning

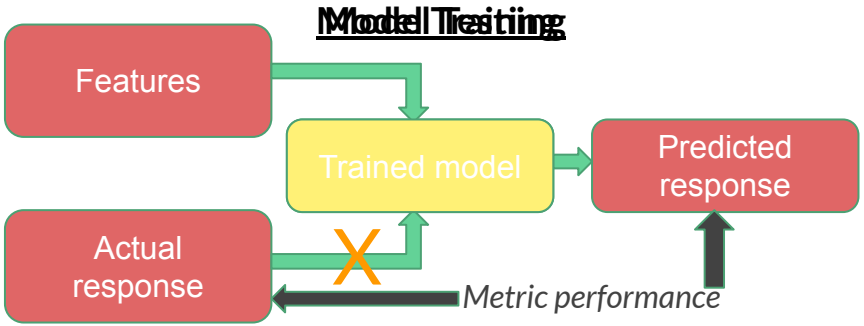
Prerequisites:

- Basic coding knowledge

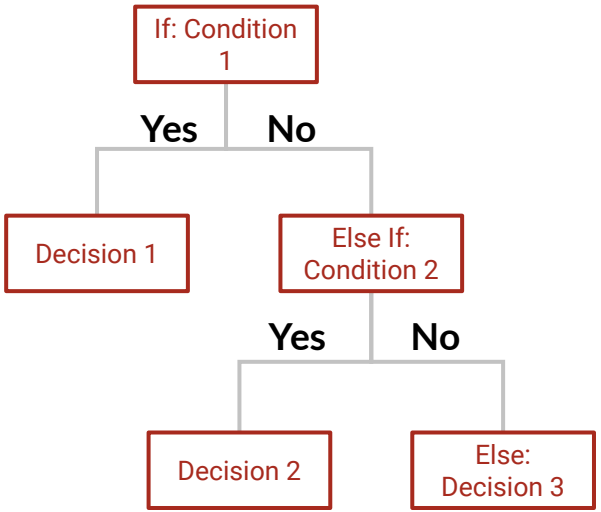
Machine learning fundamentals







Understanding Decision tree *(1 of 2)*

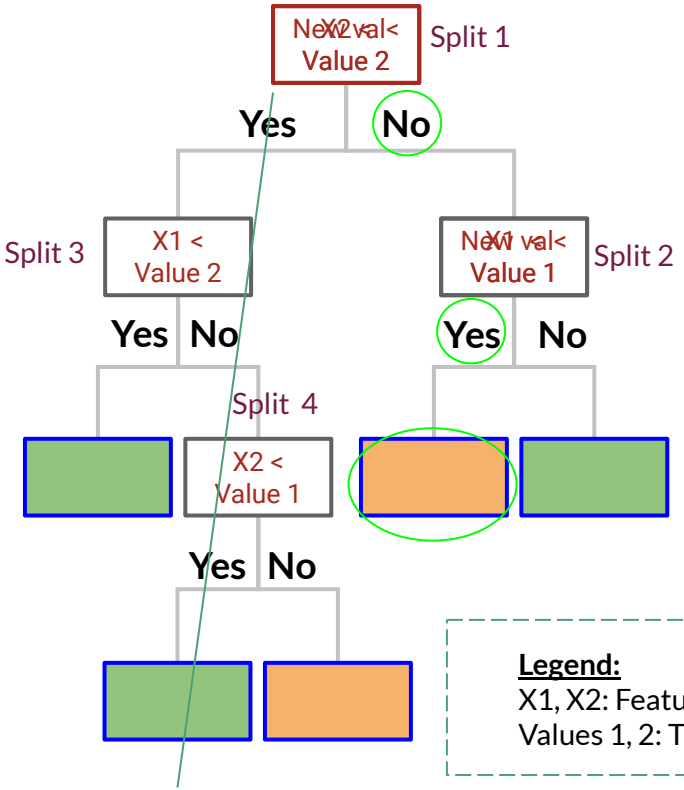


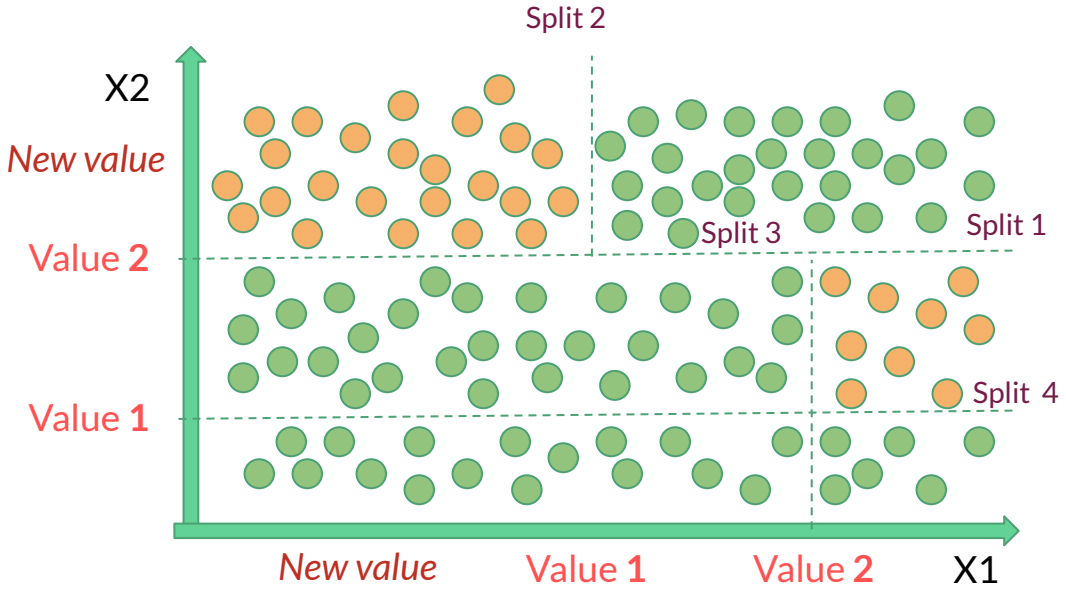
Refresher: Conditional If-Else Flowchart

Features		Response
Years in grade X1	Age X2	Status y
Value 1	Value 1	Active
Value 2	Value 2	Resign
Value 3	Value 3	Resign
Value 4	Value 4	Active
Value 5	Value 5	Active
. .	. .	. .

Example Business Scenario

Understanding Decision tree *(2 of 2)*





Legend:
X1, X2: Features
Values 1, 2: Thresholds

Active Prediction
Resign Prediction

Nodes:
Root
Leaf
Intermediate

Branches:

Building a Decision tree ML model

Jupyter notebook: Link

Task: To classify a breast cancer cell between benign and malignant (response) based on several cell characteristics (features)

1. Import

Importing necessary libraries to work with the data

2. Load

Dataset loading

3. Process

Exploratory data analysis

- Feature engineering
- Train-test split
- Statistical tests
- Feature selection
- Feature importance

4. Model

Model building steps

- Training
- Diagnostics
- Test evaluation
- Feature importance

Knowledge check

Decision Tree Recap:

Pros:
- Intuitive, simple to understand

Cons:
- Weak learner
- Prone to overfitting

Further explorations -

- Regression decision tree
- Hyperparameter tuning

Next lesson - Random Forest

Thank you! See you next time