

**MA477: Data Science**  
**Lesson 19 Board Sheet — 27 February 2026**  
United States Military Academy, West Point  
Instructor: MAJ Patrick Kuiper

---

## 1 Decision Tree Classification Lesson Objectives

- Understand how the algorithm for classification trees differs from regression trees.
- Understand and know how to interpret the Gini index.
- Understand and know how to interpret entropy.
- Use functions in Python's sklearn module to fit classification trees.

## 2 Practice Problem: Entropy with Two Categorical Predictors

### Question

Consider the following small dataset for predicting Heart Disease (HD), where  $Y \in \{\text{Yes}, \text{No}\}$ .

There are two categorical predictors:

- Smoking  $\in \{\text{Yes}, \text{No}\}$
- Exercise  $\in \{\text{Low}, \text{High}\}$

ID	Smoking	Exercise	HD
1	Yes	Low	Yes
2	Yes	Low	Yes
3	Yes	High	No
4	Yes	High	No
5	No	Low	Yes
6	No	Low	No
7	No	High	No
8	No	High	No

### Tasks:

1. Compute the entropy of the parent node.
2. Compute the entropy after splitting on Smoking.
3. Compute the entropy after splitting on Exercise.
4. Which split would a decision tree choose using information gain?