Introduction to Data Science
Supervised learning (Classification)

Suppose you are working for a bank that is considering whether to approve or reject loan applications from customers. Here are the variables in the dataset:

- **Credit score (numeric):** A number representing the customer's creditworthiness, ranging from 300 to 850.
- **Income (numeric):** The customer's monthly income, in dollars.
- **Loan amount (numeric):** The amount of money the customer is requesting to borrow, in dollars.
- **Loan term (categorical):** A categorical variable representing the length of the loan, either "short-term" (less than 3 years) or "long-term" (3 years or more).
- **Employment status (categorical):** A categorical variable representing the customer's employment status, either "employed", "self-employed", "unemployed", or "other".
- **Previous delinquencies (categorical):** A categorical variable representing whether or not the customer has had any previous delinquencies on loans, either "yes" or "no".
- The target variable is a binary variable representing whether or not the loan application was approved, either "yes" or "no".

Tasks

- **(10pt)** Using this dataset, you could build a decision tree to help the bank determine which loan applications to approve and which to reject based on the customer's characteristics.
- **(10pt)** Create dataset for the above structure and draw the decision tree
- **(10pt)** Use DecisionTreeClassifier from Python library sklearn to create the decision tee
- **(10pt)** Evaluate your classifier (calculate the accuracy of the model)
- **(10pt)** (Bonus 10pt) Visualize your decision tree

Suppose you work for a company that sells products online. Your boss has asked you to create a decision tree to help them decide whether or not to launch a new product. Here's the data you have:

- The cost to manufacture the product is $20 per unit.
- The price at which the product can be sold is $40 per unit.
- The fixed cost to launch the product is $50,000.
- The estimated demand for the product is 10,000 units.

Here are the possible outcomes:

- If the product is launched and demand is greater than or equal to 10,000 units, the company will make a profit of $200,000.
- If the product is launched and demand is less than 10,000 units, the company will make a profit of $100,000.
- If the product is not launched, the company will not make any profit or loss.

(10pt) Your task is to create a decision tree to help your boss decide whether or not to launch the new product. Use the data provided above to determine the expected value of each decision.

(10pt) Once you have created the decision tree, calculate the expected value of launching the product and not launching the product. Based on your analysis, what recommendation would you give your boss?