Please follow the steps to answer the corresponding questions.
(Set all random_state to 2023 if models and functions have this parameter)

1. Use breast cancer dataset
Hint:
from sklearn.datasets import load_breast_cancer
cancer = load_breast_cancer()

Questions:

a. How many instances, features, and targets (results) are there?
b. Is this dataset an imbalanced dataset? Why?
c. If this dataset is an imbalanced dataset, please let it be a balanced dataset.

2. Split this dataset to training data and testing data.

Questions:

a. Which method will you use? Sequential or random? Why?

3. Decrease the ratio of training data to the whole dataset from 95% to 5% gradually and every change is 5% (i.e., 95, 90, 85, …, 10, 5%, and testing data is 5~95%).

Questions:

a. Try to discuss the impact of such changes on the six classification models, Decision Tree, Random Forest, XGBoost, SVC, KNN, and Logistic Regression. The impact is like accuracy, recall, f1-score, or others you image.
b. In your experiments, which ratio of these six classification models will perform best, respectively?

4. According to the result of 3-b, please find the important features for these six classification models and draw the corresponding figure by descending order.

Questions:

a. What are the top 3 important features in each classification models?
b. Is there same important feature in these six classification models?