

Data Mining

ENSIA 2025-2026

Lab sheet N°7: Classification



Objectives: Understand and implement decision tree algorithms. You will explore the theoretical aspects of decision trees, train and visualize decision trees using Scikit-learn, and fine-tune hyperparameters to improve model performance.

Resources

- **Decision trees in scikit-learn:** [1.10. Decision Trees — scikit-learn 1.7.2 documentation](#)
- **Decision tree classifiers:** [DecisionTreeClassifier — scikit-learn 1.7.2 documentation](#)
- **Model selection:** [sklearn.model_selection](#)
- **Model evaluation:** [3.4. Metrics and scoring: quantifying the quality of predictions — scikit-learn 1.7.2 documentation](#)

Requirements

Run the following commands to install required libraries before starting:

1. pip install six
2. pip install pydotplus
3. conda install python-graphviz


Used dataset

1. [diabetes.csv](#)
2. [load_breast_cancer — scikit-learn 1.7.2 documentation](#)

Lab parts

Part 1: Theoretical Exercises (45 minutes)

Exercises designed to enhance understanding of evaluating classification models.

File:  Lab 7 (part 1) - Classification Evaluation.pdf

Part 2: Training and Visualizing Decision Trees (45 minutes)

Train decision trees using Scikit-learn and visualize the produced tree.

Notebook: [Lab 7 \(part 2\) - Decision Trees.ipynb](#)

Part 3: Hyperparameter Tuning in Decision Trees (90 minutes)

Hyperparameter tuning, model selection, and evaluation of decision tree models.

Notebook: [Lab 7 \(part 3\)- Hyperparameter Tuning in Decision Trees.ipynb](#)