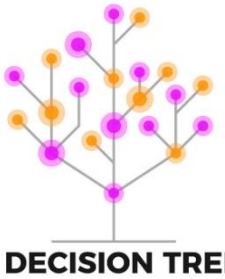
Lab sheet N°9: Classification (part1)



This lab sheet focuses on understanding and implementing decision tree algorithms in the context of data mining. You will explore the theoretical aspects of decision trees, implement binary splits for continuous attributes, and finally, train and visualize decision trees.

Resources

- Decision tree in scikit-learn: **1.10. Decision Trees**

Requirements

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

Used dataset

1. Diabets.csv

Lab parts

- ### Part 1: Theoretical Exercises (90 minutes)

This part focuses on theoretical exercises to understand how decision tree algorithms choose the best split based on training data.

File: Lab 9 (part 1) - Classification using Decision tree exercices

- ### Part 2: Continuous Attribute Split (45 minutes)

Implementation of how to choose the best binary split in the case of continuous attributes.

Notebook: Lab 9 (part 2) - Continuous attribute split

- ### Part 3: Training and Visualizing Decision Trees (45 minutes)

Training of decision trees using scikit-learn and visualization of the produced decision tree.

Notebook: Lab 9 (part 3) - Decision Trees