Program: SIM
Term and Year: 2024/2025
Course Title: **Computational Tools for Data Science**

Final Project

Description:

This is a requirement based on what you have learned during the practical part of the sections in the computational tools for data science course. The output is a python project applied on a dataset accomplished with the data science techniques.

1. Data set:

You can choose between the 2 attached datasets.

- OnlineRetail data set
- Loan Prediction data set

2. The user interfaces to enter the required data.

3. Project requirements

● Explain the problem to be solved and clarify what will be the input and the output of the project.

● Full Description for the dataset.

● Data preprocessing:

  ○ Data cleaning (ex: null values, repeated data, ..)

  ○ Data preparation (ex: unifying data types, split columbus, padding, …)

  ○ Apply this stage using Pandas, NumPy, sklearn libraries.

● Data visualization:

  ○ Use a different type of Data Visualization tools in the project (at least 3 plots).

  ○ State whether there are outliers by visualization or not.

● **Based on the nature of the dataset**, apply one of the following data science techniques and explain why:

  ○ Data clustering via k-means:

    ■ Explain the clustering with N. centroid and what are the endpoints that clustered the data into.

- Apriori algorithm:
  - State the association rules from the frequent itemset.
- Decision Trees:
  - Tree structure to reach conclusions or results. At each level, there is a choice to follow either of the branches.
  - State the rules that will make the decisions from the tree.

- Data insights
  - State clear points from the output, what are the insights that can be taken in the consideration from the data.

4. Submission form:

Each team needs to provide the python code of the project and report for the project steps and the insights which are written in markups cells inside the Jupyter notebook. So, the delivery will be as the following:

- Project File in .ipynb "Renamed with DS Python Project + team ID"
- Project File in .pdf "Renamed with DS Python Project+ team ID"
- pdf file including your full names, Ids, role of each member in the team and the group.

5. Groups:

- Team members must be from 2 to 5 maximum.
- There is a leader who represents the team.
- Each member in the team is responsible for an implementation part in the project and will discuss his part by himself in the discussion.
- Each member in the team could be asked for the other members' parts, so all of you are responsible for the project output.
- All team members should be attended to the discussion, no excuse to not be there or let another member represent you.

6. Delivery Dates: (In week 14)

- There will be a discussion day and the deadline, held exactly on the practical exam day.
- The delivery after the deadline is not acceptable.

Important Notes:

- No cheating from other groups, and if happened, all the teams that are involved will be penalized.
- It is not acceptable to take online projects and imitate them by any means.
- NO MARKS FOR THE TEAM IN THESE CASES.