****

**Coding week : 27-31 March 2023**

**Project 1: Movie recommender systems using model-based collaborative filtering (Matrix factorization)**

**Objective**

The project’s goal is to train a machine learning algorithm using movieLens data for movie recommendation by implementing a matrix factorization approach.

**Introduction**

The purpose of a recommender system is to search for content that a user may be interested in. These systems are extensively used by companies to advertise for movies, books, music tracks, restaurants and many applications to users who had shown similar taste in the items they trade in. In particular, a movie recommender system in general is a system that is used to predict the chance that users who had watched a particular movie will likely be interested to watch a different movie depending on the choices of movies they make and their choices matches other users.

**Process**

# Project contextualisation

Let’s start first by defining recommender systems and understand their purpose and the most used approaches to implement it.

Indications:

* What is recommender systems? How can we classify its approaches?
* What is matrix factorization?
* What is machine learning? How can we classify its approaches?

# Theory aspects of matrix factorization

* Understand the mathematical aspects of matrix factorization
* Study of ALS algorithm/NMF algorithm

# Loading and preprocessing of data

The data is available on <https://grouplens.org/datasets/movielens/100k/>

****

# Exploratory Data Analysis (EDA) and Data visualization

Before jumping in model building, it is a good practice to get a feel for the data. Exploratory data analysis will be used to understand the dataset using different visualization techniques. This is important as it helps understand our data format and its distribution.

Indications:

* What is EDA? What are its most used techniques?
* Take a look to the first records of the data
* Provide useful statics regarding the dataset
* Heatmaps/ Histograms

# Model building and Evaluation

* Specify the experimentation process (K-cross validation)
* Train and test the model using ALS algorithm/NMF algorithm

***Let’s have fun and work hard***