

Data Mining – Project Instructions

January 2021

For the evaluation, you will use the French census data from 2016, for the Grand-Est region. It is available on Arche. Together with the data you will find a PDF file in French describing the various attributes in the data and their values.

You will have to form groups of 2 or 3, and write a report about your project process. Explain what you did and why you did it, what were your intuitions and what were the results. You will apply the methods mentioned in class and in the practical sessions to extract concepts, implication rules and other information and draw conclusion on them. A typical Project would follow the steps below. You might need to implement some algorithms or components in addition to those implemented during practical sessions.

1. Read the PDF with the description of the attributes present in the dataset.
2. Ask yourself questions like the following (replace ... by one or multiple attributes) Think of the interesting things we found on the banking dataset:
 - What is the impact of ... on ...?
 - What are the effects of ...?
 - What causes ...? Is it possible to predict ...?
 - Are ... and ... related?
 - What would be categories of people based on ..., ... and ...?
 - *etc.*
3. Select the relevant attributes (a dozen of them) from the dataset.
4. Apply scaling on the attributes, based on their nature (ordinal, nominal, numeric) and what to do with them. You are not limited to the scaling methods seen in practical session, you can use those discussed in the theoretical class too.
5. Select the FCA tools you want to use: lattices; interest measures, and which ones (see article referenced in class); implications and association rules; *etc.*
6. Apply them.
7. Interpret the results.
 - Did you learn anything conclusive from the data? Why?
 - Do the conclusions match what you expected?
 - *etc.*