**F21DL:**

**Data Mining & Machine Learning**

*Coursework 1*

*Students:*

Mohammad ALKHALDI

Anis BENAMER

Quentin DUCASSE

Sylvain TOUANEN

*Lecturers:*

Diana BENTAL

&

Ekaterina KOMENDANTSKAYA



Table of Contents

[Introduction: 4](#_Toc23860078)

[File Management, Data Pre-Processing, Transformation and Selection 5](#_Toc23860079)

[Naïve Bayes Nets 6](#_Toc23860080)

[Complex Bayes Nets 7](#_Toc23860081)

[Clustering 8](#_Toc23860082)

[Research Question 9](#_Toc23860083)

# 

# Introduction

In the scope of the F21DL-Data Science and Machine Learning course, the first coursework uses a bank of images of street signs and wants to predict the type of a given image. In order to respond to this objective, several tools are used.

The first part of the coursework, and therefore the first one of this report is centered around data *pre-processing*, how we managed files, how we processed, transformed and selected the data available. The second part is focused on *Naïve Bayes* and both its accuracy and issues. Along with this part comes a reflection on the information we learned about the dataset. The third part revolves around *Complex Bayesian Architectures*, how to build them and how they perform compared to the previous model. Along with this part come reflections on the new properties found about the data as well as the help *Bayesian Nets* provide over *Naïve Bayes*. The next part is centered on *K-Means and Clustering* and how it can be applied to our problem. This part comes with a discussion on the results obtained along with the clustering part. Finally, a research question is asked and a solution to it is provided in the final part.

This coursework has been the occasion to train the machine learning skills we obtained throughout the course, either during lectures or labs. As the subject was completely different from the ones in the labs, because it is related to computer vision, we were taking the coursework as a new challenge to prove our understanding of the course.

# Data Pre-Processing

***File Management***

***Pre-processing data***

***Transformation***

***Selection***

# Naïve Bayes Nets

# Complex Bayes Nets

# Clustering

# Research Question

Bayesian optimisation

<https://machinelearningmastery.com/what-is-bayesian-optimization/>