

# Introduction

---

This rapport summarizes the main points and explains the rational behind my choices in the practical section of the first assignment of the Machine Learning course Fall 2022.

Similar to the practical section, this rapport can be divided into sections:

1. Preprocessing
2. Training

Each section is further divided into subsections addressing the corresponding sub-tasks.

## Preprocessing

---

### Addressing the column 'var7'

The column 'var7' represents date-time data. My work with this column can be summarized as follows:

- fixing the invalid dates: 29-th of february in a non-leap year. My solution is simple: reducing such dates by 24 hours.
- decomposing the date-time data into smaller components: year, month, day, time.
- A straight-forward analysis reveals that:
  1. There is only one year in the entire column: 2019
  2. Only the first 7 months were considered
  3. The distributions of months and days per month are quite similar between the positive and negative classes
  - 4.