# Class Project 1

Kanbak Can, Kubik Anna, Vostriakov Alexander
*Ecole Polytechnique Federale de Lausanne*

*Abstract*—**Computer Science have revolutionized the society as well as science in general. Nowadays, Machine Learning is revolutionizing Computer Science. This report summarize what we learned so far in Machine Learning and how we used that knowledge to classify the provided test set.**

## I. INTRODUCTION

Quick introduction

## II. THE STRUCTURE OF A PAPER

### A. Data sets

The train data used for this project is composed of the output variables $\mathbf{y}$ and of a set of input variables $\mathbf{tX}$. The input variables $\mathbf{tX}$ have a dimensionality $D = 30$ and a cardinality of $N = 250000$ - same cardinality as the output variables $\mathbf{y}$. The 30 variables of $\mathbf{tX}$ contains 29 real variables and one categorical ones containing 4 categories.

The test data used to generate prediction of possible outputs $\mathbf{y_{pred}}$ consists of a set of input variables $\mathbf{tX_{test}}$ of the same dimensionality as $\mathbf{tX}$ and a cardinality of $N = 568238$.

### B. Data pre-processing

After a visual inspection of the data, it appeared that most the input variables contained the value $-999$. We supposed that those values corresponded to an error or a missing value. To treat this issue, we replaced those values by the average value of the corresponding input if that input had more good values that errors. Otherwise, we did not take in account that input. Furthermore, it seemed that in the last input, that error value was $0$ instead of $-999$. Therefore we applied the same method as previously.

### C. Classification

What we did and how

### D. Results

What results we got, analysis.

## III. SUMMARY

Summary and conclusion