

ML Uses Cases

TD2/3 – project

You work for the Polish train company.

We have collected data every five minutes, on train circulation: predicted arrival delay at each station (compared to official timetable).

Data source : <https://www.kaggle.com/datasets/bartek358/train-delays>

No more information available than what is explained on this link

You work in groups which must be defined before October 10th.

Your first step is to propose uses cases: what can be achieved with such a data set ? Try to think as your customer do.

My initial request, as your customer for this project, is "to predict for each train its real arrival time"

You have to prepare 2 sprint review (30 min each).

- The first one on October 12/13.
 - o Share suggestion on what can be done with the dataset
 - o Prepare questions to transform my initial request in a real project.

Between the two sprints review you may send a bunch of questions anytime, I'll do my best to answer

- The second and last sprint review (October 26 27) will be public (in front of other groups) and you will present your result. It may be not definitive but should give a real and concrete insight of where the project might land

You have to prepare a document for each sprint review that I have to receive the day before.

I can steer different groups in different directions, depending on what I have in mind or ideas you shared at the start.

The evaluation will be based less on precision of your models than on the questions that you asked and secured at the beginning and your project management