For this project we have to predict the difficulty of some sentences. These sentences are in a dataset named “unlabelled test data “.The difficulty labels are like the languages level. It goes from A1 to C2, from the beginning to the expert level. For example, “j’aime la pizza” is probably a A1 sentence, and “Avec mon nouveau collègue on s’entend comme larron en foire” is more likely to be a difficult sentences so between B2 and C2.

We are here in a supervised machine learning problem. For this work we decides to work on the problem individually to have the maximum of models and parameters to see what are the best models and parameters.

For the first weeks we work on these models and then we put together our models and our research to see what are the bests models. In the next section we’ill provide our research and our analysis to find individually the bests models.

**Alessandro part**

**Taylan part**