**ProFan**

ProFan is a project which takes part in the action « Innovation numérique pour l’excellence éducative ». The core of this project is an experiment conducted over 20,000 professional high-school students with the help of more than 1000 teachers. Those students are undertaking one of those high-school majors related to technical fields:
- «Accompagnement, Soin et Services à la Personne» (ASSP),
- «Métiers de l’Electricité et de ses Environnements Connectés» (MELEC),
- «Commerce» (COM)

It aims to evaluate the effectiveness of collaborative learning through Jigsaw method. Along with scholar results, progression is also measured with activity like Text coping, Brainstorming…

**Brainstorming data overview**

Brainstorming activity produced chat conversation processed to determine the idea of each line. Based on that, three criteria evaluate creativity: flexibility, fluidity and originality.

- Large number of data available (> 50k annotated lines)
- Annotations provided by experts
- Features:
  - Poor context
  - Unusual spelling (mistakes, abbreviations, …)
  - Reality representative

**Objectives**

Computerize annotating process

The following sub-goals have been identified:
- Separate lines with idea from those without
- Find a way to give annotations to lines with idea
- Compute criteria according to expert’s formula

**A first approach using machine learning**

Google tool allowing us to transform text data into vectors of floats. To build the related space, the whole chat conversations have been used.

It can be modeled by a two-class classification problem where lines with an idea are positive and those without negative.

We implemented a neural network that achieves 86% of good-classification rate.

**Results**

In order to avoid, for now, noise from the classification mistake of the Filtering step, we extracted only the true positive data. They have been used to build a first classifier that aims to differentiate the categories. Among the 23,133 lines available, 4,627 composed the testing set.

The global accuracy reaches 58% which is low. However, the confusion matrix has an interesting feature: outside the first class, the mistake are very few as shown in Fig. 3.

After investigating, it was revealed that some of the first class predicted were in fact undetermined. They are easily identifiable so we can retrieve them which produces the Fig. 4.

Those results are encouraging because ~66% of the initial data are determined with high accuracy. Nonetheless, there are still ~34% to work on and other needs for the criteria that makes it unsatisfying.