Programming 1 - M1S1 - 2020

Project 02: **The way to NLP processing**

Detecting contradiction and entailment in multilingual text

> *"…when you have eliminated the impossible, whatever remains, however improbable, must be the truth"* -Sir Arthur Conan Doyle

Our brains process the meaning of a sentence like this rather quickly.

We're able to surmise:

- Some things to be true: "You can find the right answer through the process of elimination."
- Others that may have truth: "Ideas that are improbable are not impossible!"
- And some claims are clearly contradictory: "Things that you have ruled out as impossible are where the truth lies."

Natural language processing (NLP) has grown increasingly elaborate over the past few years. Machine learning models tackle question answering, text extraction, sentence generation, and many other complex tasks. But, can machines determine the relationships between sentences, or is that still left to humans? If NLP can be applied between sentences, this could have profound implications for fact-checking, identifying fake news, analyzing text, and much more.

The Challenge:

If you have two sentences, there are three ways they could be related: one could entail the other, one could contradict the other, or they could be unrelated. Natural Language Inferencing (NLI) is a popular NLP problem that involves determining how pairs of sentences (consisting of a premise and a hypothesis) are related.

Your task is to create an NLI model that assigns labels of 0, 1, or 2 (corresponding to entailment, neutral, and contradiction) to pairs of premises and hypotheses. To make things more interesting, the train and test set include text in fifteen different languages! You can find more details on the dataset by reviewing the Data.

Today, the most common approaches to NLI problems include using embeddings and transformers like BERT.

BUT BEFORE BEFORE THIS challenge, we are asking you a small challenge:

Step 1: To prepare a simple notebook that reads the Data files in csv format and describe them.

Step 2: To save the data in a JSON file, for sure the JSON will be more explicit becaus it will hold each time key:value, in the format that you can propose.