

## **MAIS 202 – PROJECT DELIVERABLE 2**

Anthony Dumais & Alberto Lopez

### **1 – Problem statement**

---

The goal is to create a natural language inferencing (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.

### **2 – Data preprocessing**

---

The dataset we will use is the one provided in the ‘Contradictory, My Dear Watson’ Kaggle competition. The dataset is composed of 5195 unique samples and each one of them includes an ID, premise, hypothesis, label, as well as the language of the text and its two-letter abbreviation. The main preprocessing method is done using a tokenizer to transform the premises and hypothesis (or any sequences of words) into arrays of numbers.

### **3 – Machine learning model**

---

- A. We implemented a preliminary model using Tensorflow, Keras and Bert, since it was suggested in the tutorial notebook of the competition. The dataset was already split into training data and test data.
- B. We don’t have this information yet, see section 4.
- C. A challenge we face was runtime errors during the implementation of the model. In order to solve this problem, we carried out online research and read appropriate documentation to understand what exactly was going wrong.

### **4 – Preliminary results**

---

The percentage of relationships correctly predicted will be used as a score to measure the accuracy of our model. The test data provided by Kaggle does not contain the labels needed to evaluate the accuracy of our model. To obtain this metric, participants of the competition need to submit their predictions to the website. We worked on this project using Google Colab notebooks and we are

now encountering difficulties in doing a Kaggle submission. Thus, we don't have any concrete results yet, but we are expecting an accuracy of about 60% for this first try.

## **5 – Next steps**

---

We think that BERT was a good choice to try out since there were good resources available online, like the tutorial notebook provided in the competition, to guide us in implementing the model. We plan on trying to make a successful submission to Kaggle to obtain the accuracy. We might look for ways to modify the model to obtain better performances or even try a different model and compare the difference in accuracy.