(Retrieved from the competition page)

Data Description

In this Getting Started Competition, we’re classifying pairs of sentences (consisting of a premise and a hypothesis) into three categories - entailment, contradiction, or neutral. Let’s take a look at an example of each of these cases for the following premise:

He came, he opened the door and I remember looking back and seeing the expression on his face, and I could tell that he was disappointed.

**Hypothesis 1:**

Just by the look on his face when he came through the door I just knew that he was let down.

We know that this is true based on the information in the premise. So, this pair is related by entailment.

**Hypothesis 2:**

He was trying not to make us feel guilty but we knew we had caused him trouble.

This very well might be true, but we can’t conclude this based on the information in the premise. So, this relationship is neutral.

**Hypothesis 3:**

He was so excited and bursting with joy that he practically knocked the door off it's frame.

We know this isn’t true, because it is the complete opposite of what the premise says. So, this pair is related by contradiction.

This dataset contains premise-hypothesis pairs in fifteen different languages, including: Arabic, Bulgarian, Chinese, German, Greek, English, Spanish, French, Hindi, Russian, Swahili, Thai, Turkish, Urdu, and Vietnamese.

**Files:**

* **train.csv**: This file contains the ID, premise, hypothesis, and label, as well as the language of the text and its two-letter abbreviation
* **test.csv**: This file contains the ID, premise, hypothesis, language, and language abbreviation, without labels.
* **sample\_submission.csv**: This is a sample submission file in the correct format: id: a unique identifier for each sample label: the classification of the relationship between the premise and hypothesis (0 for entailment, 1 for neutral, 2 for contradiction)

Special thanks to Tensorflow Datasets (TFDS) for providing this and many other useful datasets! For more information, visit: <https://www.tensorflow.org/datasets>

*Disclaimer: The dataset for this competition contains text that may be considered profane, vulgar, or offensive.*