## **Description of the Model:**

For the question classification problem, the method we used was **Support Vector Machines** from *scikit-learn* for the classification and **TF-IDF Vectorizer**, also from *scikit-learn*, to calculate the frequency of words in the document. Our pre-processing consisted of ignoring the terms that had a document frequency below the threshold of 3 (min\_df = 3) , normalizing the words by stripping accents (strip\_accents = ‘unicode’), removal of stop words (coarse and fine had different stop words, as through experimentation we realized this choice would lead to less errors) and word manipulation by using **WordNetLemmatizer** from *nltk* to convert each word to its lemma, followed by **PorterStemmer** from *nltk* to convert the resulting lemma into its stem form.

Regarding the difference between coarse and fine models, for coarse we simply vectorized the document and used a single SVM to classify all questions, while for fine we went with a hierarchical strategy and first classified questions with our coarse model approach, then created a different SVM for each of the coarse labels, each classifying every question in the subset of questions within the same coarse label classification with a fine label. This way, the choices were narrower and led to less classification errors.

## **ACCURACY:**

In the “coarse” model, we observed an accuracy of **83,91%**. The remaining **16,09%** are distributed, among the classification categories, like this:

**{'HUM': 33, 'ENTY': 30, 'ABBR': 10, 'DESC': 33, 'NUM': 23, 'LOC': 24}**

In the “fine” model, we observed an accuracy of **76,13%.** The remaining **23.87%** are distributed, among the classification categories, like this:

**{'HUM': 45, 'ENTY': 82, 'ABBR': 10, 'DESC': 39, 'NUM': 26, 'LOC': 25}**

The numbers after the tag represent how many times our model failed to correctly identify each category (for example, 'HUM': 45, means the model classified a “Human” question as another category 45 times).

## **Short Error Analysis:**

We conclude that both our models fail at classifying questions in the “Human”, “Entity” and “Description” categories, specifically, the fine model fails the most in classifying the subcategories of “Entity”. It is also important to mention that when the models fail, usually they incorrectly classify questions as “Entity” (for example, they classify “Human” or “Description” as “Entity”).

## **bIBLIOGRAPHY:**

<https://scikit-learn.org/stable/>

<https://www.nltk.org/>