**Shared Task SemEval2023**

**Group 2**

**How to run:**

Step 1:

The first step is to upload the **training file**.

Step 2:

As the code has been written in order, you can run the cells one by one. The main parts of the code are **preprocessing** and **evaluate** functions that provide appropriate data and run all the four models (LR, SVM, RF and GuassianNB) in a loop.

 \* *Note that for each combination of cleaning approach, there is a code named step.*

The signature of the main functions:

1) **X , y = preprocessing(df,step)**

**Parameters**:

**df** : uncleaned input dataframe

**step** :

**0** : means just doing lowercasing

**1** : means doing lowercasing, removing punctuations and  stopwords

**2** : means doing lowercasing, removing punctuations and  stopwords,lemmatization,removing url/html tags

**3** : means doing lowercasing, removing punctuations and  stopwords,lemmatization,removing url/html tags,emojis,and chatwords

2) **evaluate(step)**