ML Spark Exercise

Now we will use the *MLlib* library to train an ML model for predicting sentiment.  
That is, whether the sentence is positive or negative.

The dataset is composed of 3 datasets that you should work on all of them.

They all have the same schema but the topics are different.

Follow the examples from the lectures and the knowledge you acquired from the previous exercise and train a model.

The model should be imported as follows:

from pyspark.ml.classification import LogisticRegression

The sentences should be transformed using:

1. Tokenizer
2. HashingTF
3. IDF

All from pyspark.ml.feature.

Train the model and print the metrics.

You **do not** have a testset, so use a subset of the train as a dev.

The test set will be on Sunday within the final exercise.

If you finished and don’t want to play games (achtung, snake.io, etc.), try to tune the model on the dev (same dev) for better results.