

# Final assestment NLP1

Prof. Emmanuel Páez

August 23, 2023

## 1 Instructions

This is the practical evaluation assestment of the course, please read carefully each section, remember this must be delivered through canvas and it is required to be completed by class teams. You must upload one notebook file per team (".ipynb"). You must include a header for each section and you must all the code used to complete the assessment.

Before you begin, download the following dataset [Dog and Cats dataset](#). The data set contains sounds labeled by two main classes (Dogs and cats).

## 2 Tasks (total 100 pts)

### 2.1 Data set Visualization (20 pts)

By using the libraries and methods discussed in class, make an inspection and analysis of the dataset, include The visualization of some sound waves that might help to understand the sound characteristics of the data set.

Include plots and comments to understand time domain and frequency features like amplitude, frequency, amplitude envelope or RMS (ADSR model), zero cross rate.

### 2.2 Feature extraction data set (20 pts)

In this task you must prepare your data by extracting features of each sound file, include as many features as you like but you must include at least the ones reviewed in class.

Create a quick visualization of the data set and include some graphs and statistics to give a general review of the features and characteristics founded for each label (dogs and cats labels).

Include you comments and conclusions at the end of the section.

### 2.3 Dimension analysis (40 pts)

By using PCA in the data set you prepared in section 2.2, perform a dimension analysis in order to found the number of dimensions by which we might cover at least 85% of cumulative variance, include at least two 2D plots to visualize how it behaves the dimensions after being transformed.

Then by using LDA repeat the same analysis and transformation over the original data set, repeat the visualization process to see how it performs now the correlation.

Include comments and conclusions at the end of the section.

### 2.4 ML model (20 pts)

Choose at least two ML models, and train your ML models to be able to classify your validation set.

Create a confusion matrix and report the accuracy of each model.

Include conclusions and comments of which one its better and why you think this happened.