**EEG Data Analysis and Machine Learning (SVM)**

***Alcoholic vs Control Groups***

Students: Iván García Álvarez, Jorge Iñiguez Avalos y Ulises Aguirre Jauregui

GitHub Repository: <https://github.com/navi931/EEG-alcohol>

#### Table of Contents

[1. Project Overview](http://localhost:8889/notebooks/Downloads/SVM_Analysis.ipynb#intro)

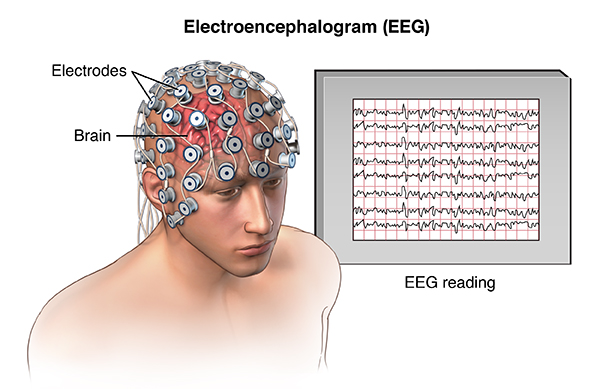
[2. Setting up the Environment](http://localhost:8889/notebooks/Downloads/SVM_Analysis.ipynb#env)

[3. Preprocessing](http://localhost:8889/notebooks/Downloads/SVM_Analysis.ipynb#Preprocessing)

[4. Training](http://localhost:8889/notebooks/Downloads/SVM_Analysis.ipynb" \l "Training) & Testing

[5. Conclusions](http://localhost:8889/notebooks/Downloads/SVM_Analysis.ipynb" \l "Conclusiones)

# **1.Project Overview**

This project has been published in Kaggle EEG-Alcohol as a challenge to implement Machine Learning, so we have taken the challenge and in this project we will implement Support Vector Machines. Like the study of Kaggle I have chosen the EEG-Alcohol dataset, which EEG (Electroencephalography) which are the data of two groups - Alcoholic and Control Group.

The number of subjects in each group is 8. The 64 electrodes were placed on the scalp of the subject to measure the electrical activity of the brain. The response values were sampled at 256 Hz (3.9 ms time) for 1 second. Each subject was exposed to a single stimulus (S1) or to two stimuli (S1 and S2) that were images of objects chosen from 1980 Snodgrass and Vanderwart picture set. When two stimuli were shown, they occurred in a matched condition where S1 was identical to S2 or in an unpaired condition where S1 differed from S2.

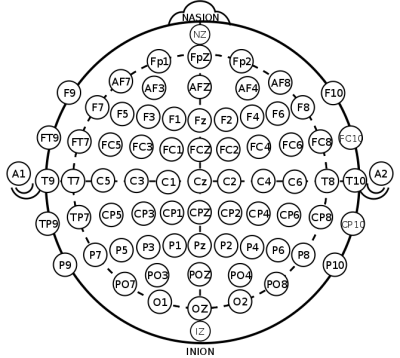
The purpose of the algorithm will be to detect differences between the response values for different stimuli between the control and the alcoholic group.

***Electroencephalography:***

**Electroencephalography** (**EEG**) is an [electrophysiological](https://en.wikipedia.org/wiki/Electrophysiology) monitoring method to record electrical activity of the [brain](https://en.wikipedia.org/wiki/Brain). It is typically noninvasive, with the [electrodes](https://en.wikipedia.org/wiki/Electrode) placed along the [scalp](https://en.wikipedia.org/wiki/Scalp), although invasive electrodes are sometimes used, as in [electrocorticography](https://en.wikipedia.org/wiki/Electrocorticography). EEG measures voltage fluctuations resulting from [ionic current](https://en.wikipedia.org/wiki/Ion_current) within the [neurons](https://en.wikipedia.org/wiki/Neurons) of the [brain](https://en.wikipedia.org/wiki/Brain). Clinically, EEG refers to the recording of the brain's spontaneous electrical activity over a period of time, as recorded from multiple [electrodes](https://en.wikipedia.org/wiki/Electrode) placed on the scalp.[[1]](https://en.wikipedia.org/wiki/Electroencephalography#cite_note-Niedermeyer-1) Diagnostic applications generally focus either on [event-related potentials](https://en.wikipedia.org/wiki/Event-related_potential) or on the [spectral content](https://en.wikipedia.org/wiki/Frequency_spectrum) of EEG. The former investigates potential fluctuations time locked to an event, such as 'stimulus onset' or 'button press'. The latter analyses the type of [neural oscillations](https://en.wikipedia.org/wiki/Neural_oscillation) (popularly called "brain waves") that can be observed in EEG signals in the frequency domain.

# **2. Setting up the Environment**

***Variables Description:***

* trial number: number of the trial (obviously)
* sensor position: position of electrode placed on subject's scalp (based on International 10-20 system)
* sample num: 0-255
* sensor value: value in microvolts
* subject identifier: a - Alcoholic; c - Control
* matching condition:
  + S1 obj - a single object shown;
  + S2 match - object 2 shown in a matching condition (S1 was identical to S2),
  + S2 nomatch - object 2 shown in a non matching condition (S1 differed from S2)
* channel: channel number (0-63). Basically, it's the same as sensor position column, so one of these columns can be dropped
* name: a serial code assigned to each subject
* time: inverse of sample num measured in seconds

I have changed some sensor positions so they match the basics for head topography visualization. Also, I have removed those positions which equal to **X**, **Y** and **ND** since I couldn’t figure out which regions they respond to.

# ***3. Preprocessing:***

The first thing we will do is to do the preprocessing of the data, we will take into account the following recommendations for this:

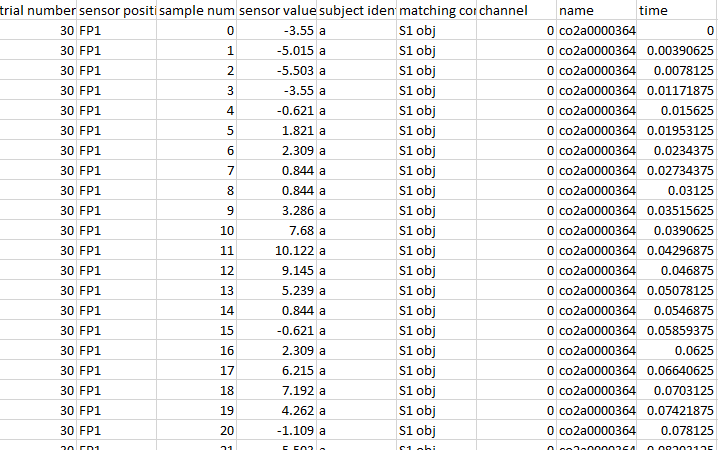
• We are not interested in an "id" of the subject, as it can cause overfitting

• We are very interested in which sensors are the important sensors, for this we will take the conclusions given by the previous study that are 'FPZ', 'FP2', 'AF3', 'AFZ', 'AF4', 'F5', 'F3' , 'F1', 'FZ', 'FC5', 'FC3', 'FCZ', 'T7', 'CZ', 'C4', 'C6', 'TP7', 'CP5', 'CP3', ' CP1 ',' CPZ ',' CP2 ',' CP4 ',' CP8 ',' P5 ',' P1 ',' PZ ',' P2 ',' P4 ',' P6 ',' P8 ',' PO7 ' , 'PO4', 'O1'

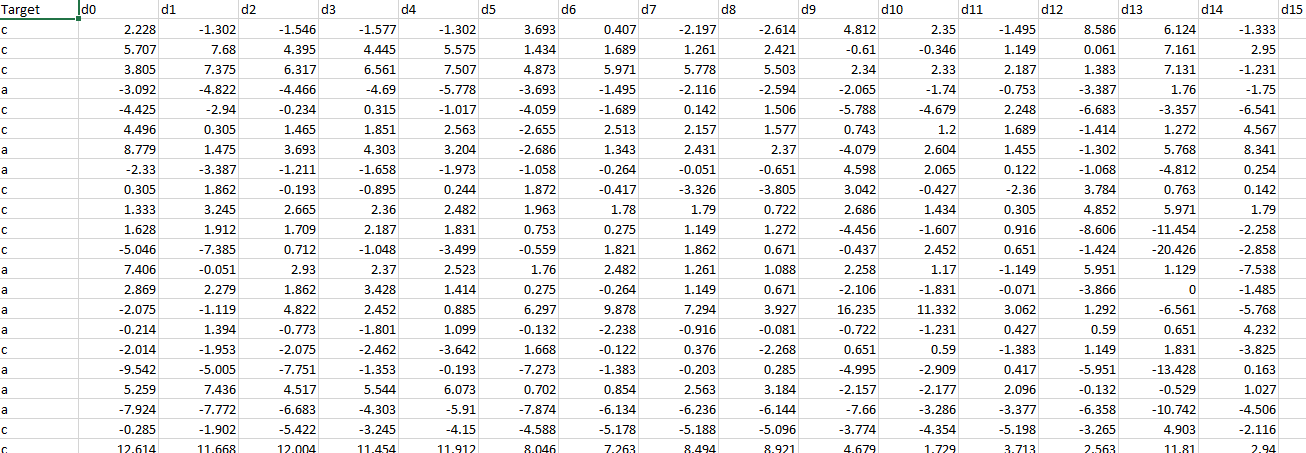
• Analyzing the data we realized that the position of the sensor and the channel, is the same then for reasons of processing we have decided to eliminate the field channel.

• For a better performance of the algorithm it has been decided to leave everything in numerical values.

The way the data was before was the following:

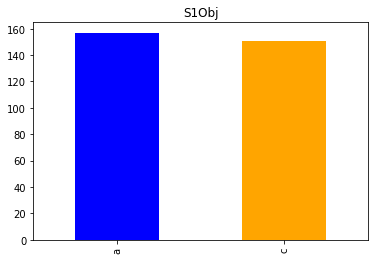


And now it's like that:



So now instead of each file being a test, each record is a test.

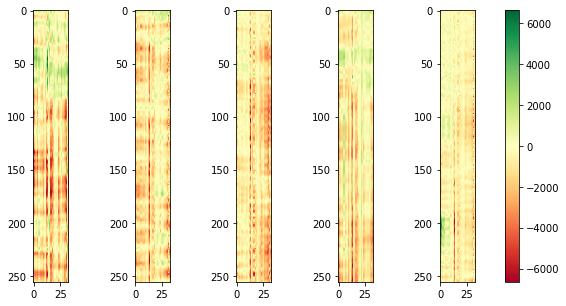
Reviewing the data, we realize that the data is very ideal that almost half of alcoholics and half are non-alcoholic



If we want to see the data in image

form to see the electroencephalogram it would look like this

* No alcoholic
* Alcoholic
* Alcoholic
* No alcoholic
* Alcoholic



# ***4.Training & Testing:***

In this case we will make 3 classifiers, each one will be trained by 3 types of test.

Model SVMs1

Confusion matrix

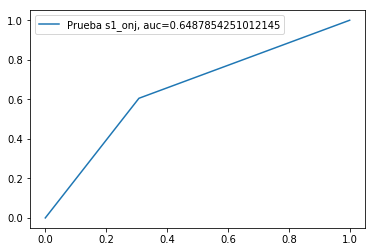
34 5

15 23

Accuracy 0.7402597402597403

recall 0.6052631578947368

precision 0.8214285714285714



Model SVMs2

Confusion matrix

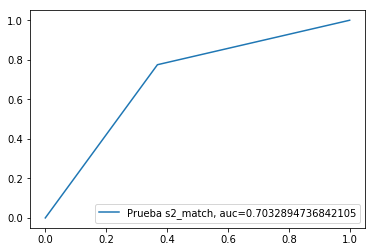
24 14

9 31

Accuracy: 0.7051282051282052

Recall: 0.775

Precisión: 0.6888888888888889



Modelo SVMs3

Confusion matrix

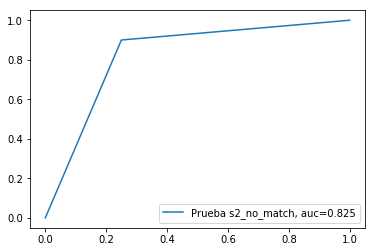
27 9

4 36

Accuracy:0.8289473684210527

Recall: 0.9

Precision: 0.8



# ***Random tests:***

Having the trained model we will see how certain random people classify.

We will take 3 people, one for each classifier, and we will see how it classifies them.

El algoritmo dijo que si es alcohólico y en realidad no es alcohólico

El algoritmo dijo que no es alcohólico y en realidad no es alcohólico

El algoritmo dijo que si es alcohólico y en realidad si es alcohólico

# ***5.Conclusions:***

If we use only the 3rd classifier we will have better results, because it has the best testing results, that is the most conclusive test to classify someone as alcoholic or not.