

1. GUÍA DE PROYECTO

Manejar datos de señales EEG para clasificar emociones. Se les proporcionará el dataset DEAP el cual contiene un total de 32 archivos.dat Cada grupo de estudiantes deberá analizar la clasificación de emociones de 5 archivos (acorde lo asignado a cada grupo).

- Información del dataset DEAP <https://www.eecs.qmul.ac.uk/mmv/datasets/deap/>
- Archivos .dat [DatasetDEAP](#)

1. La carga de datos de los archivos.dat se debe realizarlo con la librería pickle (véase Fig 1).

```
def read_file(filename):  
    x = pickle._Unpickler(open(filename, 'rb'))  
    x.encoding = 'latin1'  
    p = x.load()  
    return p
```

Fig 1. Librería pickle para archivos.dat

2. Seguido se deben leer los archivos únicamente de los datos asignados al grupo para su posterior análisis (véase Fig 2).

```
files = []  
for n in range(1, 2):  
    s = ''  
    if n < 10:  
        s += '0'  
    s += str(n)  
    files.append(s)  
print(files)
```

Fig 2. Carga de archivos específicos

3. La matriz que contienen los archivos .dat corresponde a las características / data: 40x40x8064 (videos, canales, características) y sus respectivas etiquetas: 40 x 4 (video, etiquetas), de estas etiquetas solo se van a utilizar 2 (valencia y arousal). Por tanto, deberá almacenar la data como se muestra en la Fig 3.

```
labels = []  
data = []  
  
for i in files:  
    filename = "data/s" + i + ".dat"  
    trial = read_file(filename)  
    labels.append(trial['labels'])  
    data.append(trial['data'])  
  
# Re-shape arrays into desired shapes  
labels = np.array(labels)  
labels = labels.flatten()  
labels = labels.reshape(40, 4)  
  
data = np.array(data)  
data = data.flatten()  
data = data.reshape(40, 40, 8064)  
  
print("Labels: ", labels.shape) # trial x label  
print("Data: ", data.shape) # trial x channel x data
```

Fig 3. División de datos en Labels y Data

- De la parte de Labels deberá tomar únicamente las etiquetas valencia y arousal, para el posterior análisis. Cabe mencionar que la data se encuentra desequilibrada, y de acuerdo a la teoría las características de la data deben estar presentes por igual tanto en train como en test. Por tanto, hay que implementar algún modo de etiquetar la data por igual (**One Hot Encoding**).
- Una vez tenga las etiquetas listas debe considerar únicamente 32 de los 40 canales que contienen la data, dando origen a una matriz de data: 40x32x8064. Se la forma como se muestra en Fig 4.

```
# Data tomando en cuenta solo 32 canales
eeg_data = []
for i in range (len(data)):
    for j in range (len(eeg_channels)):
        eeg_data.append(data[i,j])
eeg_data = np.reshape(eeg_data, (len(data), len(eeg_channels), len(data[0,0])))
print(eeg_data.shape)

(40, 32, 8064)
```

Fig 4. Considerar 32 canales de señales EEG.

- Al obtener la matriz de datos y la matriz de etiquetas proceda a **extraer características** de la data, para esto puede utilizar medidas estadísticas como: media, varianza, mediana, curtosis, skewness entre otras, de modo que mejore el resultado de clasificación.
- Cuando tenga la data resultante de la extracción de características unifique la matriz, realice un reshape y procesa con la división de la data para train y test.
- Aplique los algoritmos de aprendizaje supervisado y compare los modelos obtenidos mediante el uso de métricas de evaluación (Accuracy, F1, matriz de confusión, recall, etc). Configure los algoritmos con diferentes parámetros con el fin de mejorar los resultados para valencia y arousal. Puede intentar mejorar el resultado normalizando los datos.
- Coloque sus conclusiones y recomendaciones del trabajo realizado.

2. RECOMENDACIONES

- El día 07 de marzo, siendo la última clase del semestre se reserva para entrega del proyecto final y su documentación (estilo tesis).
- Cabe recordar que la nota del funcionamiento del proyecto final es grupal y la defensa del mismo individual, por ende el día de la presentación se procederá a una ronda de preguntas donde los integrantes del grupo demostraran los conocimientos adquiridos a lo largo del semestre y de la materia.

3. RÚBRICA

Algoritmo

1. Carga de datos - **1pto**
2. Separación en data y labels - **1pto**
3. One hot encoding – **2 ptos**
4. Separar datos de 32 canales - **1pto**
5. Extracción de características – **2 ptos**
6. Algoritmos de aprendizaje supervisado - **2ptos**
7. Métricas de evaluación del modelo - **1pto**

Informe

1. Tabla de contenido
2. Tabla de imágenes/tablas
3. Referencias cruzadas en imágenes
4. Marco teórico - **2ptos**
5. Conclusiones y recomendaciones - **3ptos**
6. Bibliografía - **2ptos**