

# Propuesta de proyecto final: Desarrollo de un modelo clasificador de potenciales evocados de EEG como parte de una interfaz cerebro computadora.

Moreno-Sánchez Ángel A. [2],<sup>a</sup>

<sup>a</sup> Diplomado en Ciencia de Datos, Universidad Nacional Autónoma de México. 30 de mayo de 2022.

**Resumen.-** La implementación de un modelo de clasificación robusto y con alta precisión es fundamental para el desarrollo de una interfaz cerebro-máquina capaz de interpretar las señales cerebrales motoras de una persona cuya aplicación podría en un futuro permitir la inclusión social de personas con discapacidades asociadas al movimiento. En este sentido, se plantea la selección e implementación de un modelo de machine learning validado y optimizado para la identificación y clasificación de potenciales cerebrales evocados relacionados al movimiento en segmentos de registros de electroencefalografía (EEG), orientado a la integración en un dispositivo de interfaz cerebro-máquina (BCI). Los registros provienen del trabajo publicado por Kaya et al. (2018) [2] donde se obtuvo un dataset de registros EEG para BCI con potenciales corticales lentos, el cual contiene al rededor de 60,000 de imágenes motoras en 4 paradigmas de interacción asociados a la proyección mental de movimiento de extremidades como las manos y piernas.

## 1. Esquema del proyecto

El desarrollo del proyecto seguirá una metodología similar a la presentada por Mishchenko et al (2018) [4] contando con la siguiente secuencia de desarrollo

### 1.1. Elección y adquisición de datos

El primer paso consistirá en elegir uno de los paradigmas disponibles en el dataset, del cual dependerá el número de clases a identificar teniendo 3 clases para el paradigma CLA donde se cuentan con 3 estados mentales (en reposo, movimiento de mano izquierda o derecha), mientras que para el paradigma HaLT se tienen los mismos estados mentales añadiendo otros 3 distintos (movimiento de pie izquierdo y derecho y lengua). Una vez seleccionado el paradigma se deberán descargar las sesiones correspondientes a los 13 pacientes evaluados en la base de datos.

### 1.2. Extracción y Definición de variables de entrada

Posteriormente, se realizará un proceso de extracción de datos a formato libre csv, ya que éstos provienen de un objeto de MATLAB. Al importarse se definirán las variables de entrada como los segmentos de registro correspondientes a un periodo definido tras el estímulo del paradigma, esto para cada canal de EEG. La selección de los segmentos cortos se realizará haciendo uso de las etiquetas de control del experimento donde se indica en que punto de la sesión de registro se dió el estímulo y su clase correspondiente. El producto

de esta sección será un dataset con los segmentos cortos de los canales de EEG etiquetados al tipo de estímulo motriz.

### 1.3. Selección y generación de características

Se realizará un acercamiento frecuencial, en términos de procesamiento de señales, para la realización de la clasificación, en este sentido, se generarán características de los datos de entrada que describan las propiedades espectro-temporales de las señales cortas de EEG. Para ello se emplearán 3 técnicas distintas: Descomposición por Análisis de Espectro Singular [5], Transformada Continua Wavelet [1] y Transformada de Fourier por Amplitudes [4]. Los resultados obtenidos de las distintas técnicas, conjunto de componentes ortogonales, espectrograma y conjunto de amplitudes en frecuencia respectivamente, serán las características que los modelos de clasificación utilizarán para discriminar entre las distintas clases. La intención de utilizar múltiples técnicas de análisis de EEG es poder comparar entre ellas de acuerdo a las métricas de negocio de los distintos modelos.

### 1.4. Modelado y elección de parámetros

Siguiendo la metodología para un desarrollo de modelo de machine learning, tras realizar el preprocesamiento de los datos (normalización o escalamiento, verificación de balance entre clases, etc.) se entrenarán, validarán los siguientes algoritmos:

1. *Convolutional Neural Network* (CNN)

2. *Artificial Neural Network* (Perceptrón multi-capas)
3. *Support Vector Classifier* (SVC)
4. *XGBoost Classifier*
5. *Random Forest Classifier*

Cada algoritmo será validado usando el estándar *Cross-Validation* y tendrá una optimización de parámetros para cada uno de los acercamientos planteados en la sección anterior, teniendo como métrica de negocio objetivo obtener la mayor precisión (accuracy) posible cuidando mantener un alto puntaje f1 por cada clase ya que una incorrecta clasificación de estímulo puede entorpecer el uso de la BCI.

### 1.5. Optimización de modelo y resultados

Finalmente, se iterará el proceso realizando ajustes finos tales como la selección de los intervalos de los segmentos cortos del registro EEG, una reducción de características, etc. Obteniendo al final una tabla comparativa entre cada modelo planteado y sus métricas de negocio validadas.

## 2. Justificación del proyecto

El desarrollo de BCI es una de las aplicaciones multidisciplinarias de vanguardia en el mundo, las cuales tienen por objetivo lograr una conexión directa entre el Sistema Nervioso Central (SNC) y un dispositivo informático, omitiendo el manejo y control de los dispositivos de manera mecánica. En los últimos años se han logrado avances que han permitido la implementación de BCI en un amplio espectro de aplicaciones desde el ámbito clínico, como parte de una prótesis de extremidad, hasta el entretenimiento, como una interfaz de reproducción de música sin auriculares.

Particularmente, este proyecto se enfoca en la aplicación de BCI, y en específico su modelo de clasificación, en el ámbito clínico como una alternativa de movimiento prostético. Tan solo en México cerca de seis millones de personas presentan algún tipo de discapacidad, y son precisamente las enfermedades neuromusculares (ENM) una de las causas más frecuentes de invalidez [3], por ello la importancia del desarrollo de sistemas BCI capaces de descifrar la intención del usuario para comunicarse o moverse, a través del análisis de señales eléctricas cerebrales y su posterior conversión a comandos para controlar dispositivos en tiempo real.

## 3. Conjunto de datos para usar

El conjunto de datos que se empleará es una base de registros EEG enfocado a potenciales corticales lentos a lo largo de 75 sesiones de 13 participantes involucrados en 4 tipos de paradigmas. Cuenta, en promedio, con 4.8h de registros de EEG y 4600 imágenes motrices por participante. La base de datos se encuentra validada y verificada por pares en términos de uniformidad, cobertura lateral y longitudinal y complejidad de interacción. Los datos pueden ser consultados y descargados de este sitio [Base de datos](#)

## 4. Resultados esperados del proyecto

Se espera obtener una comparación entre distintos modelos y acercamientos para la clasificación de potenciales evocados motrices de la cual se pueda seleccionar la mejor combinación de algoritmos y acercamientos para una interfaz BCI. De ser posible, se desea poner el modelo en producción dentro de un contenedor Docker para, eventualmente, implementarlo en una plataforma de análisis en tiempo real.

## Referencias

- [1] Raymundo Cassani y Tiago H Falk. "Spectrotemporal modeling of biomedical signals: Theoretical foundation and applications". En: *Reference Module in Biomedical Sciences*. Elsevier, 2018.
- [2] Murat Kaya y col. "A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces". En: *Scientific data* 5.1 (2018), págs. 1-16.
- [3] Alexandra B Luna-Angulo y col. "Diagnóstico molecular de enfermedades neuromusculares en el Instituto Nacional de Rehabilitación, situación actual y perspectivas". En: *Investigación en Discapacidad* 5.1 (2016), págs. 9-26.
- [4] Yuriy Mishchenko y col. "Developing a three-to six-state EEG-based brain-computer interface for a virtual robotic manipulator control". En: *IEEE transactions on biomedical engineering* 66.4 (2018), págs. 977-987.
- [5] Saeid Sanei y Hossein Hassani. *Singular spectrum analysis of biomedical signals*. CRC press, 2015.