BENEMERITA UNIVERSIDAD AUTONOMA DE PUEBLA

Reporte del Servicio Social



**Análisis de un electroencefalograma aplicado a una prueba.**

Salvador Alejandro Uribe Castellanos

**INDICE**

[**Introducción** 2](#_Toc57204284)

[**Descripción del conjunto de datos** 3](#_Toc57204285)

[**Procesamiento** 5](#_Toc57204286)

[**Resultados** 6](#_Toc57204287)

[**Comparación entre proyectos** 7](#_Toc57204288)

[Tesis basada 7](#_Toc57204289)

[Trabajo de mi compañera 7](#_Toc57204290)

[**Conclusiones** 7](#_Toc57204291)

[**Bibliografía** 8](#_Toc57204292)

# **Introducción**

El análisis de datos consiste en inspeccionar, limpiar y transformar datos que a simple vista no es posible realizarlo para la obtención de información útil y concisa para llegar a conclusiones y apoyo en la toma de decisiones. El análisis de datos tiene múltiples facetas con diferentes enfoques como para los negocios, la ciencia y los dominios de las ciencias sociales.

La electroencefalografía (EEG) es una exploración neurofisiológica que se basa en el registro de la actividad bioeléctrica cerebral en condiciones basales de reposo, en vigilia o sueño, y durante diversas activaciones mediante un equipo de electroencefalografía como se muestra en la **Ilustración 1** .

El **objetivo general** de este trabajo es analizar un conjunto de datos de varias pruebas de un electroencefalograma para poder realizar un clasificador para así saber que hizo el usuario con el simple hecho de tener el estudio encefalografico. Este estudio suele ser algo complicado y con una efectividad no tan alta como en otras pruebas ya que hay muchas variables que pueden alterar el resultado ya que prácticamente estaríamos leyendo la mente con el simple estudio.



**Ilustración 1.-** Ilustración de un estudio electroencefalógrafo.

# **Descripción del conjunto de datos**

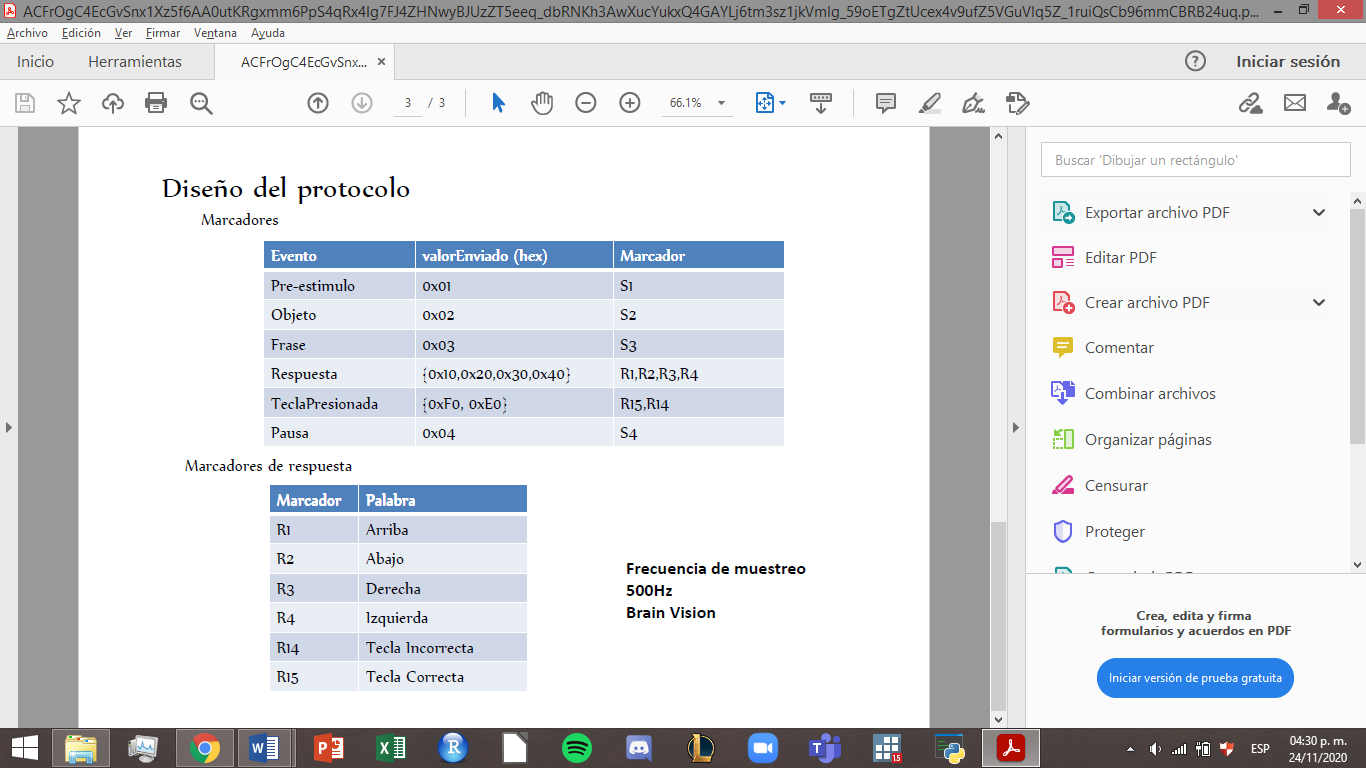
Dado que el tiempo es limitado optamos por un conjunto de datos “limpio” haciendo referencia a que no tuvimos que realizar alguna técnica de antemano para su manejo ni mucho menos realizar el estudio electroencefalógrafo por el coste que esto generaría ya que en esta etapa del análisis de datos suele ser muy tardada y sin tanto goce de conocimiento ya que son herramientas simples así que por estas razones simplificamos esa parte ya que el profesor nos asignó un conjunto de datos listos para aplicarle métodos estadísticos con sus respectiva descripción para entender lo que se hacía.

El estudio trata de una prueba de 6 segundos realizada a 20 personas diferentes con 150 repeticiones por persona en el cual el primer medio segundo es un pre-estimulo, el siguiente segundo y medio es la presentación de un objeto (este objeto se encuentra arriba, abajo, izquierda o a la derecha) el siguiente segundo y medio se le presenta la oración “el objeto estaba en “ , el siguiente segundo y medio el usuario debe de decir la oración completa con su respectiva respuesta y el ultimo medio segundo se pausa la prueba para volver a iniciarlo. El estudio arroja si el usuario acertó en el resultado o no y todos los datos son tomados por medio de un electroencefalograma. El estudio del electroencefalograma se hizo a una frecuencia de muestreo de 500 Hz con 63 electrodos.



**Ilustración 2.-** Descripción visual de la prueba del electroencefalograma.

Para tener un mayor entendimiento de la prueba, se le insertaron un par de indicadores a los archivos para saber en qué punto de la prueba empieza y termina cada sección en donde S1,S2 y S3 son las banderas de pre-estimulo, objeto y frase respectivamente, R1 a R4 para saber cual fue el resultado del usuario, R15 para saber si fue correcto o R14 si fue incorrecta la respuesta del usuario y por ultimo S4 para la pausa aplicada al final de la prueba, así como se explica en la **Ilustración 3**.

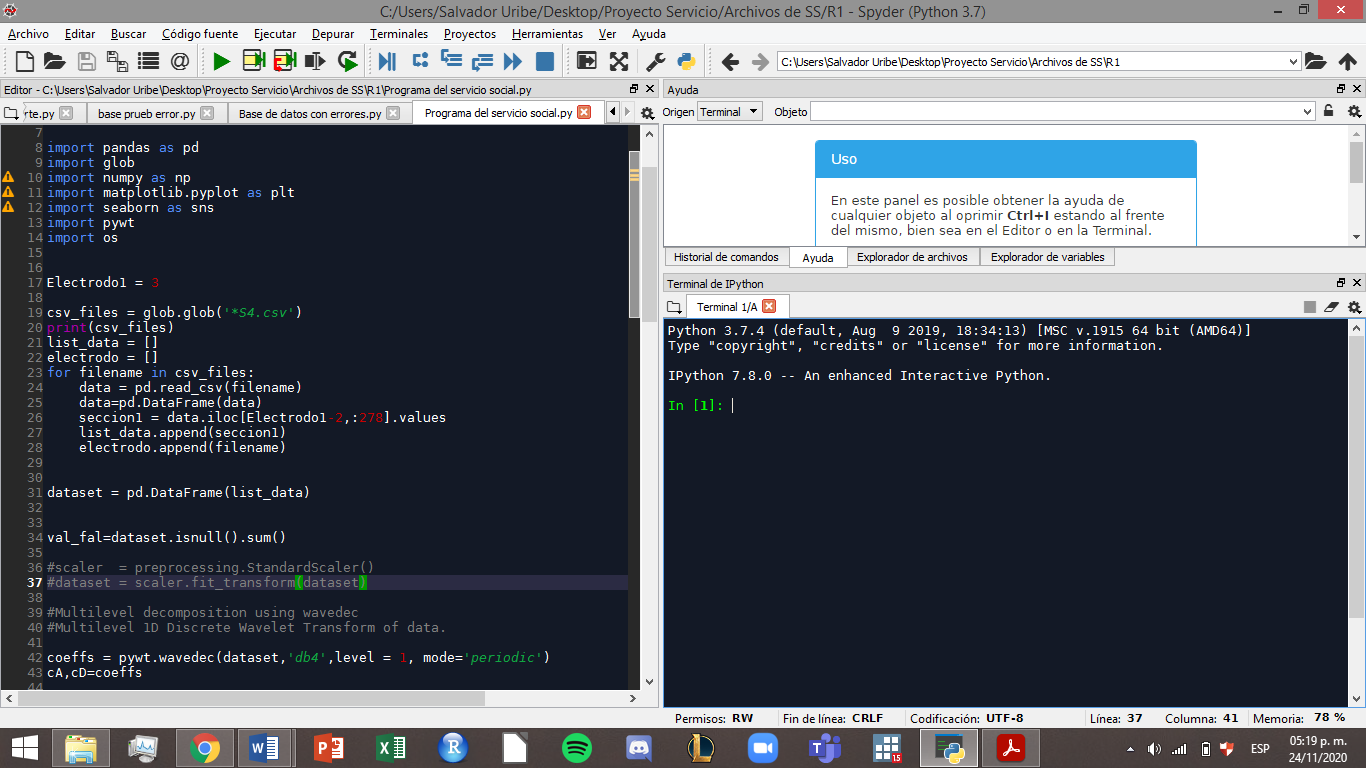
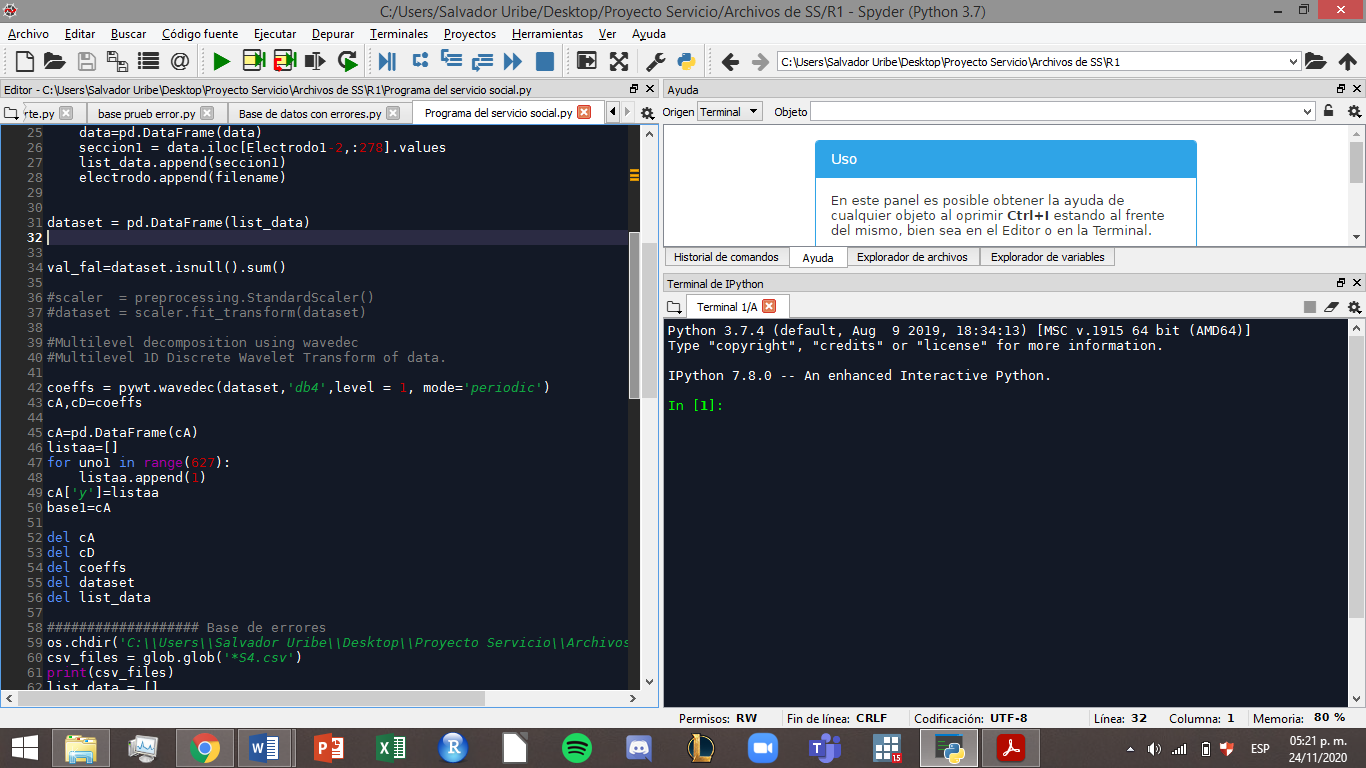


**Ilustración 3**.- Diseño del protocolo en la prueba.

Los archivos ya tratados que nos brindó el profesor se dividen en carpetas con el nombre de la respuesta elegida por el usuario teniendo en cuenta que todos los archivos fueron solamente los contestados correctamente ya que al final de cada documento en la sección del nombre viene puro R15. Cada archivo de valor separado por comas tiene la asignación del número de usuario elegido, el número de prueba y la parte de la sección que fue analizada por ejemplo Ale\_piloto\_1\_94\_R1-R15 que hace referencia al primer usuario, numero de prueba y con la sección se respuesta de pronunciación imaginada con la respuesta R1 y como tiene el R15 tenemos que es correcta.

# **Procesamiento**

Para el tratamiento de los datos utilicé el lenguaje de programación spyder de python para tener un buen manejo de los datos. La base de datos es demasiado grande para su eficaz manejo optamos por elegir todas las pruebas en donde todos los usuario respondieron con la opción R1 y esta haya sido correcta. Primero se forman los datos a estudiar, en una matriz se seleccionan todos los datos correctos de R1 y se verifica que no haya ningún dato faltante, le apliqué la transformada de wavelet con los parámetros db4 con nivel 1 para obtener una matriz más reducida con las frecuencias más relevantes divididas en dos matrices las altas(cA) y bajas (cD), dado que nosotros queremos realizar un clasificador en la base de datos también deben de ir muestras erróneas ya sea en donde haya salido R1 pero incorrecto o resultados del mismo electrodo pero como respuestas correctas R2,R3 o R4 , por practicidad se utilizó como muestras erróneas los datos arrojados de R2 en el mismo periodo de tiempo para que puedan ser comparados, esta misma matriz también se le hace el mismo tratamiento de la transformada de wavelet. En la **Ilustración 4** se muestra el tratamiento hecho para el electrodo 3 asignando datos solo del conjunto correcto que es el R1.

**Ilustración 4.-**Programa para la selección de datos y procesamiento.

Una vez que tengamos una matriz en donde estén los datos de cada electrodo con muestras correctas e incorrectas se le asigna una etiqueta a cada uno para saber si es correcto o no, la cual llamaremos Y, por comodidad se le etiquetó con los números 1 y 0 haciendo referencia a los correctos e incorrectos respectivamente. Como en todo proceso de análisis de datos a esta matriz se le da un tratamiento de estandarización, en este caso utilicé el standarScaler, una vez teniendo en forma la base de datos le apliqué la máquina de vectores de soporte (svm) y para poder encontrar el mejor kernel posible, fue haciéndolo de prueba y error ya que no siempre con el kernel sugerido será con el que uno tenga los mejores resultados. Concluí que el mejor kernel para esta base de datos fue la polinomial ya que era con la cual tenía una mejor predicción de los datos.

# **Resultados**

Para el electrodo 3 de la prueba haciendo referencia a R1 como los correctos y R2 como los datos incorrectos llegué a que el clasificador tiene un 58.16% de efectividad lo cual dado mis conocimientos aplicados es relativamente bueno porque no todos los electrodos aportan con la misma información dado que existen partes del cerebro que dan más datos que otros así que con este mismo modelo se podría encontrar una mejor predicción sobre electrodos de la misma prueba pero con mayor aporte de información.

# **Comparación entre proyectos**

## Tesis basada

Para tener más conocimiento sobre el análisis de datos enfocado a estudios electroencefalógrafos tuve como referencia una tesis de un egresado de la UAM (Muñoz, 2014), gracias a esto me fue más fácil encontrar los métodos estadísticos más aplicados a este tipo de estudios así que de esta manera tomé los parámetros de la transformada wavelet y de la máquina de vectores de soporte, claro está que estos parámetros solo los utilicé como sugerencia ya que hice unas modificaciones para tener un mejor estimador de los datos.

## Trabajo de mi compañera

A pesar de que mi proyecto y el de mi compañera haya sido enfocado sobre el mismo conjunto de datos existen diferencias en el proceso de cada uno y por lo mismo cada quien llego a resultados distintos como se presentan en la siguiente tabla resumen.

|  |  |  |
| --- | --- | --- |
|  | Mi trabajo | Trabajo de Ana |
| Respuesta correcta ,incorrecta | R1 – R2 | R2-R3 |
| Electrodo seleccionado | 3 | 10 |
| Primer método aplicado | Wavelet | Wavelet biortogonal |
| Parámetros del primer método | 'db4'  'periodic' | “bior2.4”  'symmetric' |
| Segundo método aplicado | Estandarización | Estandarización |
| Tercer método aplicado | Máquina de vectores de soporte | Regresión logística |
| Métrica de exactitud | 58.16% | 71% |

# **Conclusiones**

Este pequeño proyecto para mí ha sido muy importante ya que el análisis de datos es una rama muy importante de la actuaría y me ha hecho comprender lo importante que es recabar los datos para darle un significado y aprovechar toda la información posible hoy en día ya que existen muchas herramientas estadísticas para aplicar a los conjuntos de datos pero no todas son las mejores para cada uno ya que existen muchas suposiciones de antema lo cual hace que cada técnica estadística sea especial y útil.

# 

# **Bibliografía**

Muñoz, C. N. (2014). *Estudio de Técnicas de análisis y clasificación de señales EEG.* Universidad Autonoma de Madrid, Madrid. <https://repositorio.uam.es/bitstream/handle/10486/660477/henriquez_munnoz_claudia_nureibis_tfm.pdf?sequence=1>