



TRABAJO PRÁCTICO DE MATEMÁTICAS III

Año 2022



Visual Studio Code



Denise Martin



Universidad
Nacional
de San Martín

Tabla de Contenidos

1. Introducción
2. Palabras Claves
3. Procesado de la Información
4. Establecimiento de variables
5. Elección de gráficos
6. Conclusiones

1. Introducción

El trabajo práctico consiste en la elección y procesamiento de un dataset para el estudio y posible predicción de una hipótesis.

En este caso se selecciono uno de kaggle:

<https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers?resource=download>

Allí la hipótesis se registra como: ¿Se puede predecir la Demencia o Alzheimer's?

Y se toma en una muestra de 416 y 150 sujetos de ambos sexos, y en todos los casos diestros. Los estudios realizados en ellos se detallan en la siguiente sección

2. Palabras Claves

- ◆ **Demencia:** no es una enfermedad específica, sino una serie de síntomas que afectan la capacidad de una persona de realizar actividades cotidianas, estas incluyen:
 - Un deterioro de la memoria
 - Cambios en las habilidades de pensamiento
 - Poco criterio y poca capacidad de razonamiento
 - Disminución de la concentración y la atención
 - Cambios en el lenguaje
 - Cambios en el comportamiento
- ◆ **Alzheimer's:** es un tipo de Demencia, catalogada como patología ya que produce cambios en el cerebro. Primero afecta a la parte del cerebro asociada con el aprendizaje, por lo que los primeros síntomas suelen incluir cambios en la memoria, en el pensamiento y en la capacidad de razonamiento. A medida que la enfermedad avanza, los síntomas se agravan e incluyen confusión, cambios en el comportamiento y otras dificultades
- ◆ **Mini Mental State Evaluation:** es una herramienta que provee información cognitiva sobre la persona evaluada en un score de 0 a 30
<https://oxfordmedicaleducation.com/geriatrics/mini-mental-state-examination-mmse/>
- ◆ **Clinical Dementia Rating:** es una tabla cuantitativa de evaluación de grado de demencia
<https://knightadrc.wustl.edu/professionals-clinicians/cdr-dementia-staging-instrument/cdr-scoring-table/>
- ◆ **Estimated Total Intracranial Volume:** es un método complejo de medición de espacio intracraneal
<https://surfer.nmr.mgh.harvard.edu/fswiki/eTIV>
- ◆ **Normalized Whole-Brain Volume:** es un método de medición de tamaño total del cerebro (y de allí tamaño de la materia gris y blanca). Esta medición puede brindar datos acerca de la conectividad de las distintas regiones del cerebro.
<https://library.neura.edu.au/schizophrenia/physical-features/brain-regions/whole-brain-volume/>

3. Procesado de la Información

Se determina la elección del dataset correspondiente al Corte Longitudinal de las Resonancias Magnéticas (MRI). El mismo cuenta con 150 sujetos de edades 60 a 96, de ambos sexos y todos con uso de mano dominante derecha.

Dicho dataset fue modificado para contener los nombres de las columnas completos.

El preprocesado dejó: 'ID', 'Hand', 'Estimated Total Intracranial Volume', 'ASF' y 'Delay' fuera del dataset, ya que no van a utilizarse.

La columna de medición de Volumen Total del Cerebro se encuentra con valores de ceros, lo cual es una imposibilidad. Para evitar aumentar más los errores con el completado de valores NaN descripto a continuación, se decide eliminarlos. El dataset sigue siendo considerable en número así que podemos prescindir de esos valores.

Para no perder muchos datos, se localizó a los NAN o na, y se reemplazó sus valores mediante la estrategia de Moda (la selección de la misma se justifica por ser la forma más ‘equilibrada’ de reutilizar datos, en caso de Media, podría afianzar un resultado incorrecto). En el caso de las Evaluaciones de Demencia la Moda es cero, esto puede eventualmente traer problemas en el procesado, sin embargo, es la estrategia que mejor resultado podría traer.

La selección de Train y Test, se realiza en una proporción 80-20

Caso 1: Como se intenta probar que el Volumen Completo Cerebral tiene correlación con Demencia, se extraen de la matriz de variables independientes: Volumen Completo Cerebral y se la relaciona con los datos de la variable dependiente: Rating de Demencia.

Caso 2: Como se intenta probar que la Edad tiene correlación con Demencia, se extraen de la matriz de variables independientes: Edad y se la relaciona con los datos de la variable dependiente: Rating de Demencia.

Caso Bonus: aquí se toman ambos dataset (Cross section and Logitudinal section), para relacionar todos los datos pertinentes a la posibilidad de predicción de Demencia.

4. Establecimiento de Variables

La **variable dependiente** utilizada en este proyecto es: Clinical Dementia Rating

Las **variables independientes** son: M/F, Age, Socio Economic Status , Mini Mental State Examination y Normalized Whole-Brain Volume

Para la Regresión Lineal se utiliza:

Caso 1: Normalized Whole-Brain Volume

La **variable dependiente** utilizada en este proyecto es: Clinical Dementia Rating

La **variable independiente** es: Normalized Whole-Brain Volume

Caso 2: Age

La **variable dependiente** utilizada en este proyecto es: Clinical Dementia Rating

La **variable independiente** es: Age

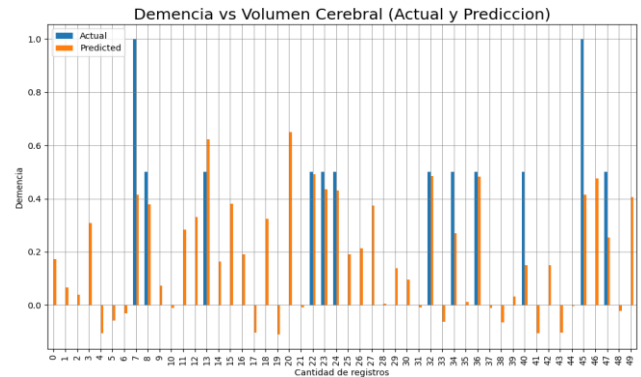
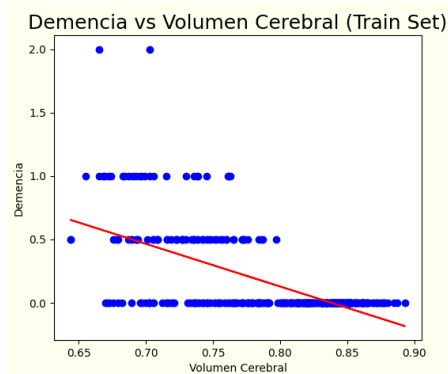
Caso Bonus:

Aquí se utilizan todas las columnas de los dataset, enfrentadas en un cuadro de doble entrada

5. Elección de Gráficos

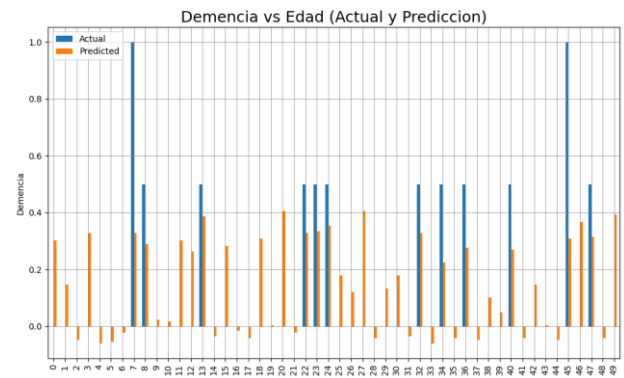
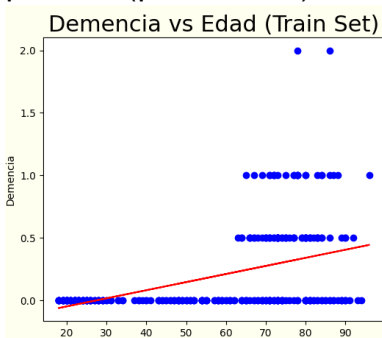
Caso 1: Normalized Whole-Brain Volume

Se utiliza un diagrama de dispersión de datos para observar mejor la tendencia
Se corrobora con un Grafico de Barras entre los datos de entrenamiento y de prueba (predicción)



Caso 2: Age

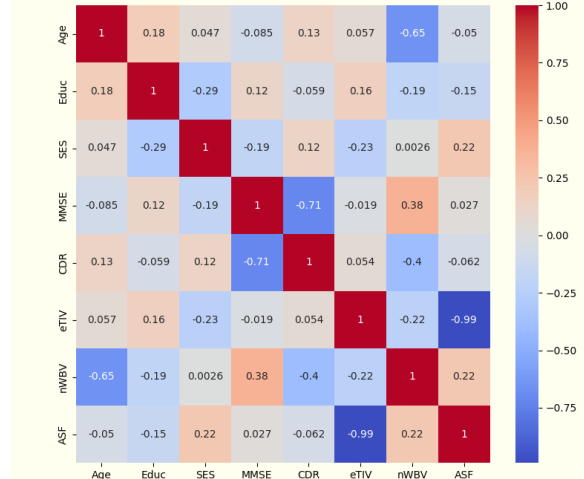
Se utiliza un diagrama de dispersión de datos para observar mejor la tendencia
Se corrobora con un Grafico de Barras entre los datos de entrenamiento y de prueba (predicción)



Caso Bonus:

Se utiliza un Mapa de Calor para visualizar la relación de los distintos estudios y herramientas

Influencia de Factores en la Prediccion de Demencia



6. Conclusiones

Caso 1: Normalized Whole-Brain Volume

La grafica muestra una tendencia: que a mayor Volumen cerebral existe una menor probabilidad de Demencia

Si bien hay una tendencia, no es posible predecir la demencia desde el Volumen del cerebro ya que los grupos de datos en el gráfico de dispersión se encuentran alineados al eje x y no con una pronunciada y balanceada pendiente.

En el grafico de barras es aún más evidente que en 50 casos las predicciones correctas han sido mínimas

No es concluyente.

Caso 2: Age

La grafica muestra una tendencia: que a mayor Edad existe una mayor probabilidad de Demencia

Si bien hay una tendencia, no es posible predecir la demencia desde el Volumen del cerebro ya que los grupos de datos en el gráfico de dispersión se encuentran alineados al eje x y no con una pronunciada y balanceada pendiente.

En el gráfico de barras es aún más evidente que en 50 casos las predicciones correctas han sido mínimas

No es concluyente.

Conclusión General:

Si bien en ambos casos hay una clara tendencia, las predicciones no son concluyentes, por lo que se infiere que se necesitan más datos para poder llevar a cabo una relación que permita predecir la aparición de un caso de Demencia. La falta de algunos datos como Nan y la necesidad de suplantar esos espacios por la moda, también pudo traer aparejado una serie de errores que se traducen a la conclusión. Tampoco se utiliza el total de datos en la gráfica de barras, y se utilizó solamente el archivo de Cortes Longitudinales; por lo que es posible, entonces, encontrar datos distintos, o más concluyentes, si se analizaran ambos datasets con todos los datos.

Conclusión Bonus:

Es posible aquí, visualizar (con ambos dataset), como la evaluación cognitiva (MMSE) y el volumen completo normalizado de cerebro, podrían ser los mejores métodos para estudio con el fin de predecir futuros casos de Demencia.

Corroborando la conclusión general de falta de datos.