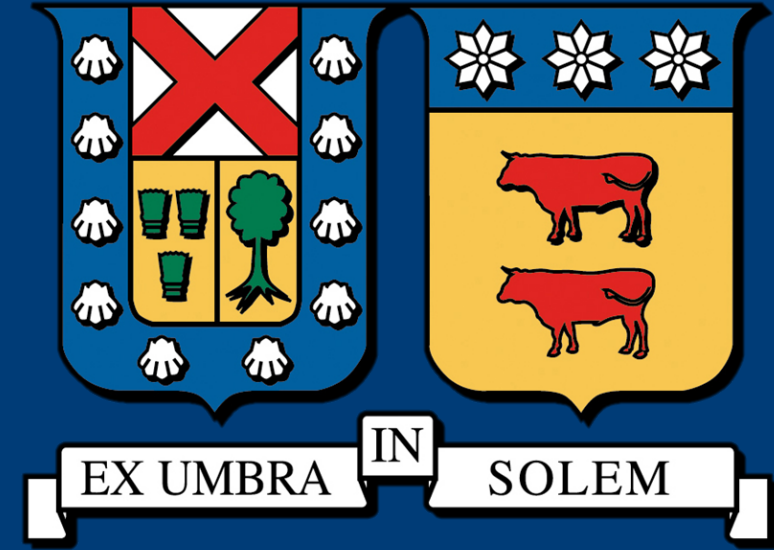




Más allá de las palabras: Cómo la escritura en línea puede ser la clave para detectar el Alzheimer

Diego Astaburuaga, David Rivas, Diego Vivallo

Universidad Técnica Federico Santa María



Resumen

Las **enfermedades neurodegenerativas** son afecciones incurables y debilitantes causadas por la degeneración progresiva de las células nerviosas. Pueden afectar los movimientos y/o las habilidades mentales, siendo el Alzheimer y la enfermedad de Parkinson las más comunes entre ellas.

El Alzheimer produce un declive lento y progresivo en funciones mentales como la memoria, el pensamiento, el juicio y las habilidades de aprendizaje.



Figure 1. Ilustración de la enfermedad Alzheimer.

No hay cura para estas enfermedades y el declive solo puede ser gestionado de alguna manera durante su progresión. Debido al aumento de la esperanza de vida en todo el mundo, se espera que la incidencia de estas enfermedades aumente drásticamente en las próximas décadas. Esto crea una necesidad crítica de mejorar los enfoques actualmente utilizados para diagnosticarlas lo más temprano posible.

Introducción

El conjunto de datos DARWIN (Diagnosis Alzheimer With haNdwriting) contiene muestras de escritura a mano de personas afectadas por la enfermedad de Alzheimer y un grupo de control. El conjunto de datos incluye 25 tareas que evalúan las habilidades de 174 participantes (89 pacientes con Alzheimer (P) y 85 personas sanas (H, Saludable)). Las categorías de tareas son:

- Memoria y dictado (M):
- Gráficas (G):
- Copia (C):

El objetivo principal es presentar un modelo que ayude a predecir si una persona tiene Alzheimer. Como objetivo secundario, se trabajará con las tareas realizadas más importantes para la detección de la enfermedad, manteniendo las mismas áreas cerebrales involucradas. Esto permitiría optimizar los gastos en la adquisición de datos y tener una detección más temprana de la enfermedad de Alzheimer.

Análisis exploratorio de los datos

Teniendo en cuenta que cada tarea tiene asociada una categoría, se decide separar cada tarea según sus categorías respectivas y utilizar la que tenga mayor información mutua en promedio con respecto a la variable objetivo. Para cada una de las tres tareas seleccionadas se realizará su respectivo análisis de datos, selección de Features (en caso de ser necesario), selección de modelo, para luego poder entrenarlo a cada tarea.

Repositorio: <https://github.com/Darkrayyss/Proyecto-Final-MAT281>

Modelos matemáticos

Agregar cosas como el modelamiento del problema, sacar la información mutua, reducir la cantidad de tareas, evitar independencia lineal talvez, utilizar un modelo distinto para cada tarea y luego combinarlos, utilizando árboles en un caso y SVM en otro para comparar.

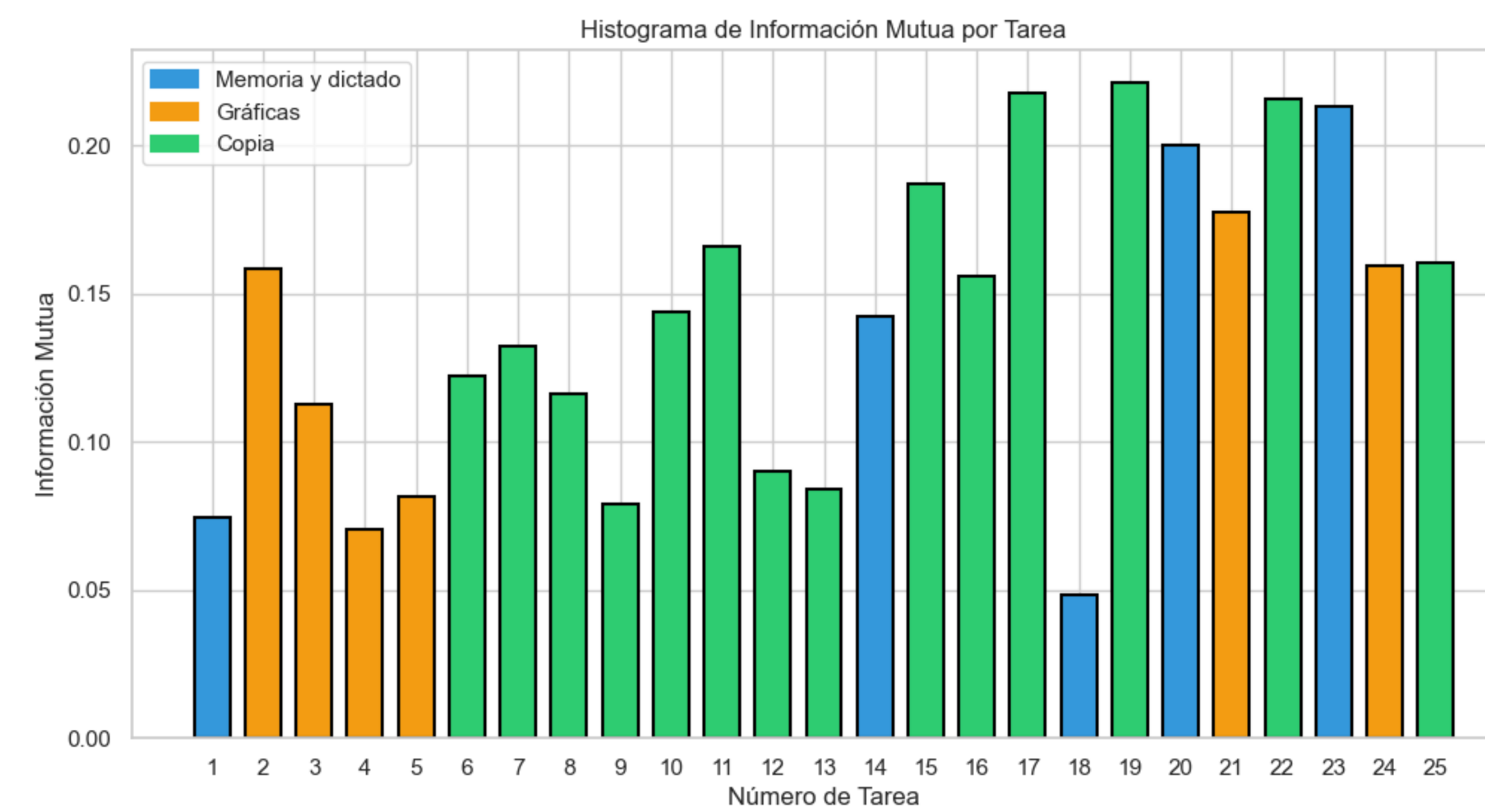


Figure 2. Información Mutua de tareas respecto al objetivo.

Máquinas de Vectores de Soporte (SVM) y Árboles de Decisión

SVM: Busca ubicar los datos en un espacio de alta dimensión donde pueda trazar un plano que divida ambas categorías.

- Pro's:** Permite generar modelos sofisticados muy buenos para clasificar.
- Contras:** Por lo general carecen de capacidad interpretativa.

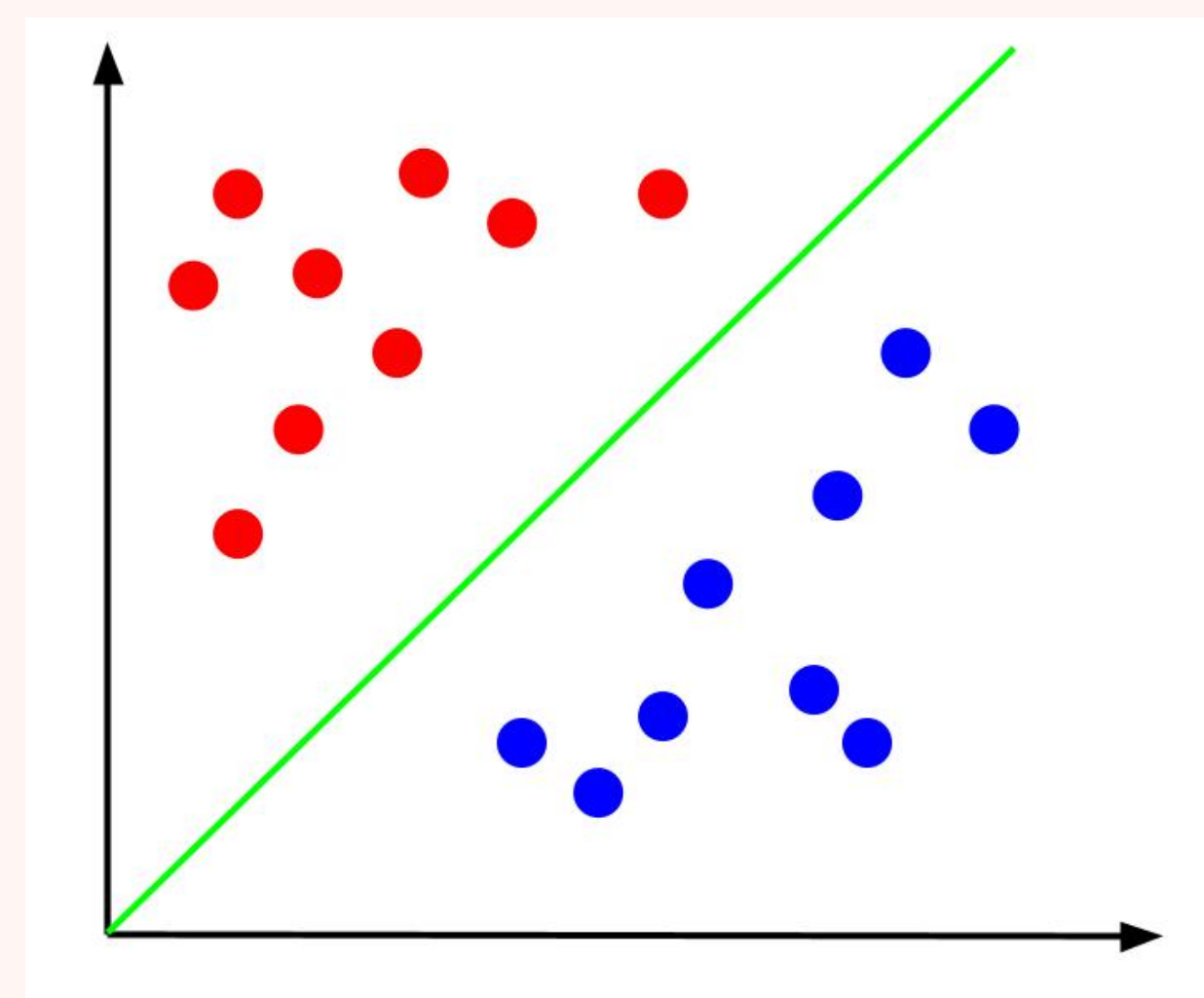
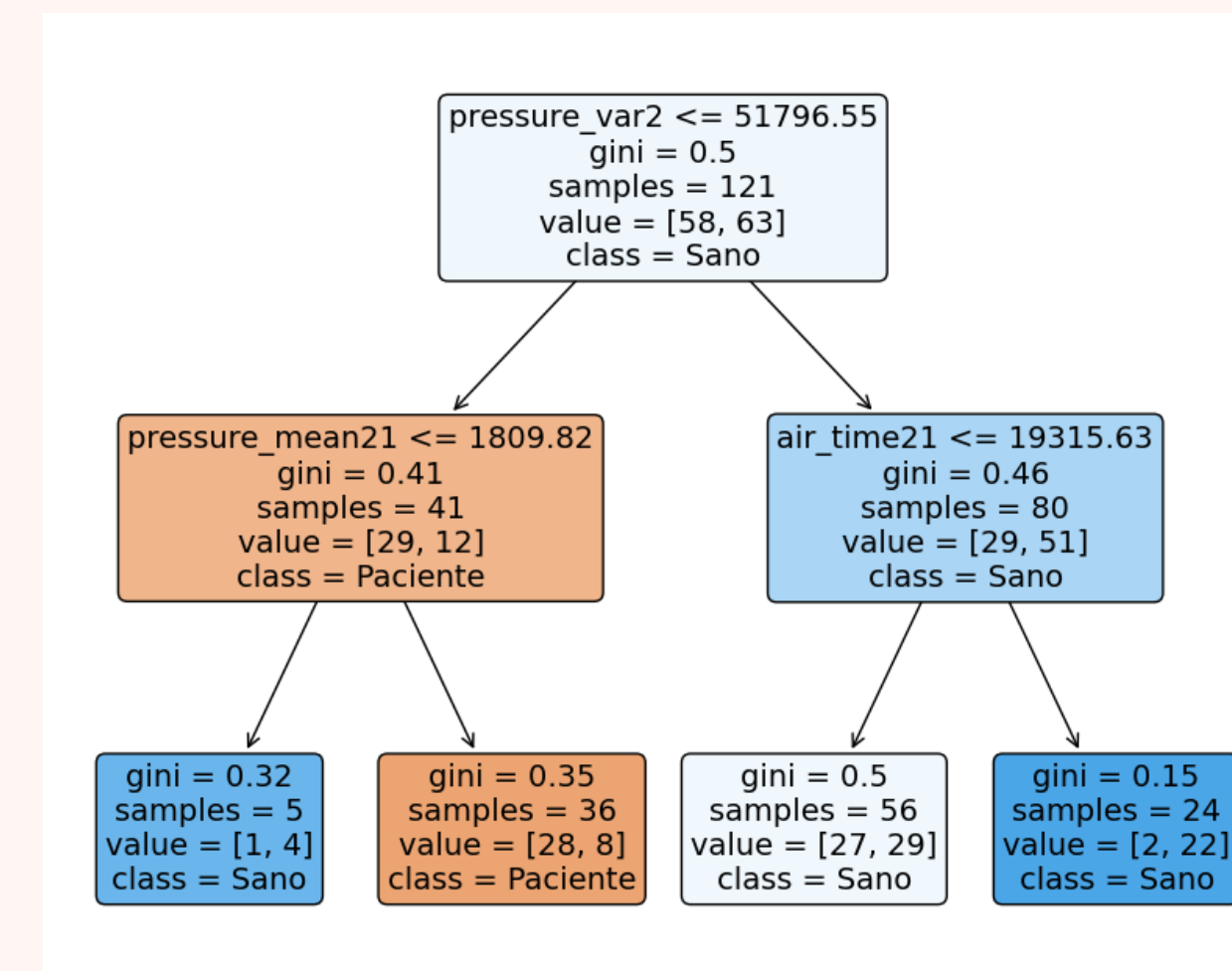


Figure 3. Ilustración de modelo SVM lineal y Ejemplo de uno de los árboles generados.

Árboles de Decisión: Estudia las variables presentes y las utiliza como regla de decisión para discernir entre las clases.

- Pro's:** Son fáciles de interpretar permitiendo tomar acción.
- Contras:** Menos capacidad predictiva y alta sensibilidad a los datos.



Resultados

	F1	Accuracy	Recall	Precision
SVM	0.926	0.925	0.962	0.893
Árbol de decisión	0.889	0.887	0.923	0.857

Considerar poner una matriz de confusión. Debe estar en español y ser entendible.

Conclusiones

Es posible desarrollar técnicas que permitan reducir la cantidad de tareas a realizar para la detección del Alzheimer pero siempre teniendo en cuenta el trade-off entre capacidad predictiva e interpretabilidad. Además, si bien muchas veces más simplicidad es bueno para el desarrollo de un trabajo, el exceso de esta puede provocar pérdida de información relevante y, por ende, un mal desempeño. *esto fue escrito como idea, no está pensado en ser utilizado en el momento final, no borrar este mensaje*

Distribución de Clases Objetivo

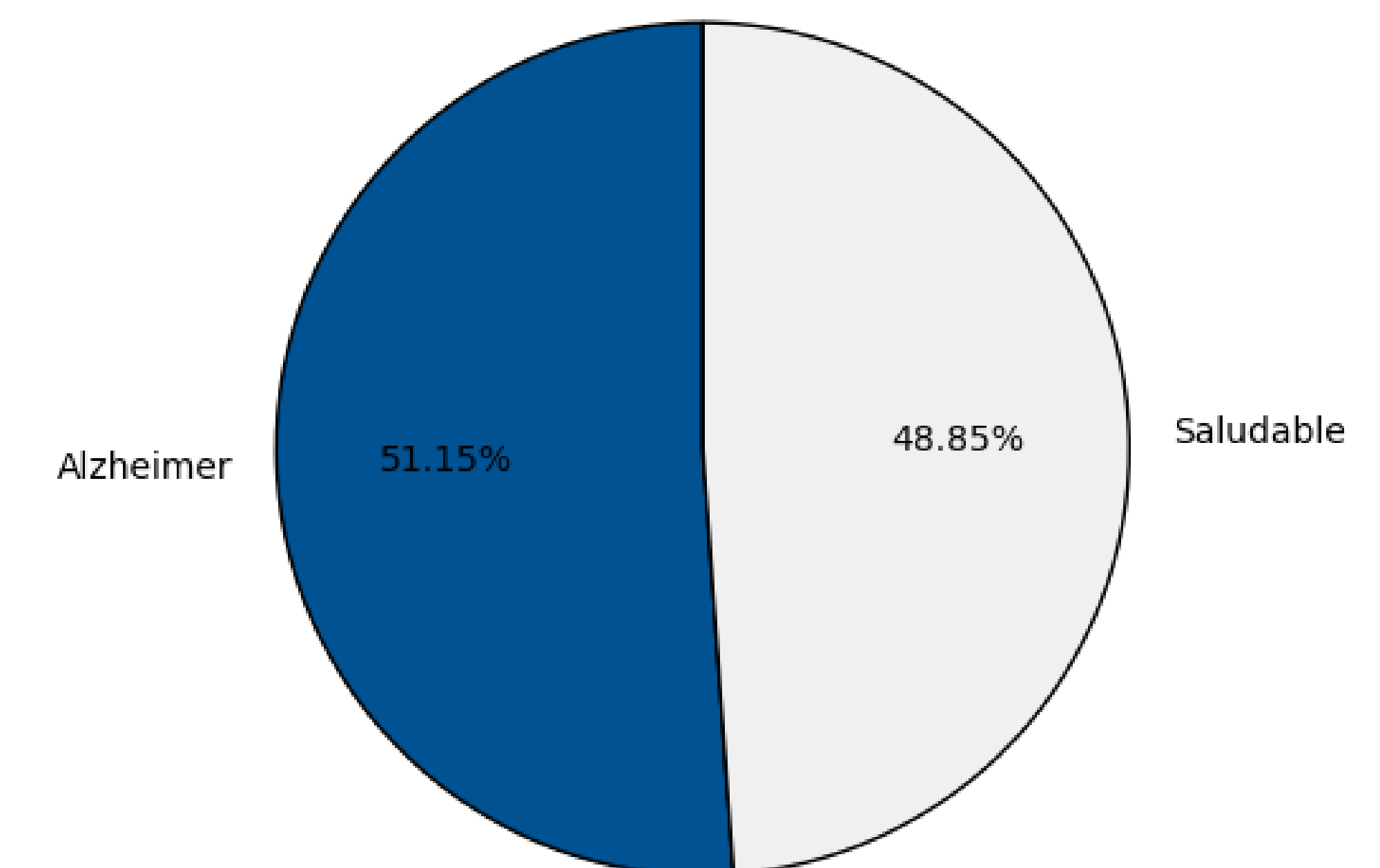


Figure 4. Gráfico de la distribución de las clases.

Referencias

- Nicole D. Cilia, Giuseppe De Gregorio, Claudio De Stefano, Francesco Fontanella, Angelo Marcelli, and Antonio Parziale. Diagnosing alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. *Engineering Applications of Artificial Intelligence*, 111:104822, 2022.
- Nicole Dalia Cilia, Claudio De Stefano, Francesco Fontanella, and Alessandra Scotto Di Freca. An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis. *Procedia Computer Science*, 141:466–471, 2018. The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018) / The 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018) / Affiliated Workshops.
- Francesco Fontanella. DARWIN. UCI Machine Learning Repository, 2022. DOI: <https://doi.org/10.24432/C55D0K>.