



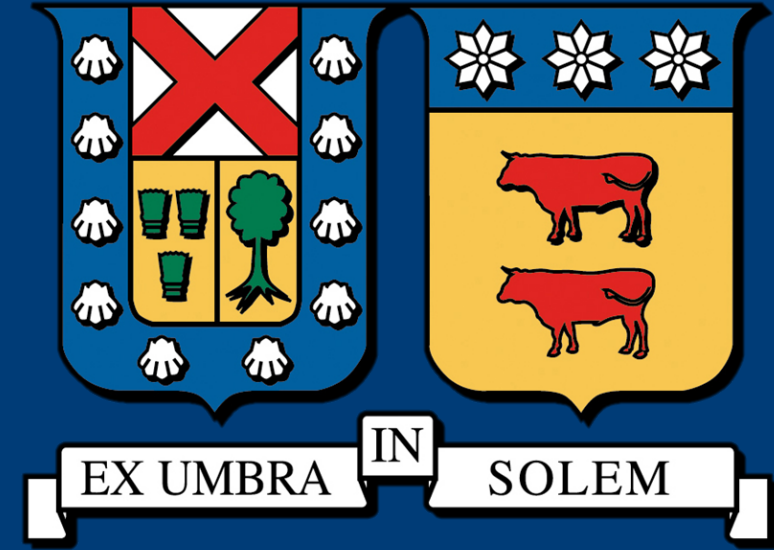
UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

DEPARTAMENTO
DE MATEMÁTICA

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



Introducción

Las **enfermedades neurodegenerativas**, como el Alzheimer, son afecciones incurables que resultan de la degeneración progresiva de las células nerviosas, afectando tanto los movimientos como las habilidades mentales. El Alzheimer, por ejemplo, conlleva un declive gradual en funciones como la memoria, el pensamiento y el juicio.



Figure 1. Ilustración de la enfermedad de Alzheimer.

A pesar de la **ausencia de cura**, es crucial gestionar el declive durante su progresión. Con el aumento de la esperanza de vida a nivel mundial, se anticipa un drástico aumento en la incidencia de estas enfermedades en las próximas décadas, subrayando la urgencia de mejorar los métodos de diagnóstico temprano.

Resumen de los datos y objetivo

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** es generar un modelo de aprendizaje de máquina que sea capaz de diagnosticar la enfermedad de Alzheimer en las personas mediante la realización de tareas escritas a mano.

¿Cómo debemos enfocar el trabajo?

En el proceso de desarrollar un modelo para diagnosticar enfermedades como el Alzheimer, es razonable buscar una **implementación sencilla** que no requiera procedimientos extensos y, al mismo tiempo, ofrezca **resultados efectivos** para que pueda ser ofrecida como una **solución real**.

Esto ofrece un reto para este proyecto debido a que no siempre *más es mejor*, debido a que un exceso de información en la construcción de un modelo de aprendizaje de máquinas puede ser redundante y empeora la efectividad de este, pero a su vez disminuir mucho esta puede provocar que el modelo no tenga capacidad de diagnosticar correctamente.

Enfoque propuesto

Utilizando una herramienta matemática llamada **información mutua**, se puede dar un indicativo de "cuánta información" tiene cada tarea al momento de diagnosticar.

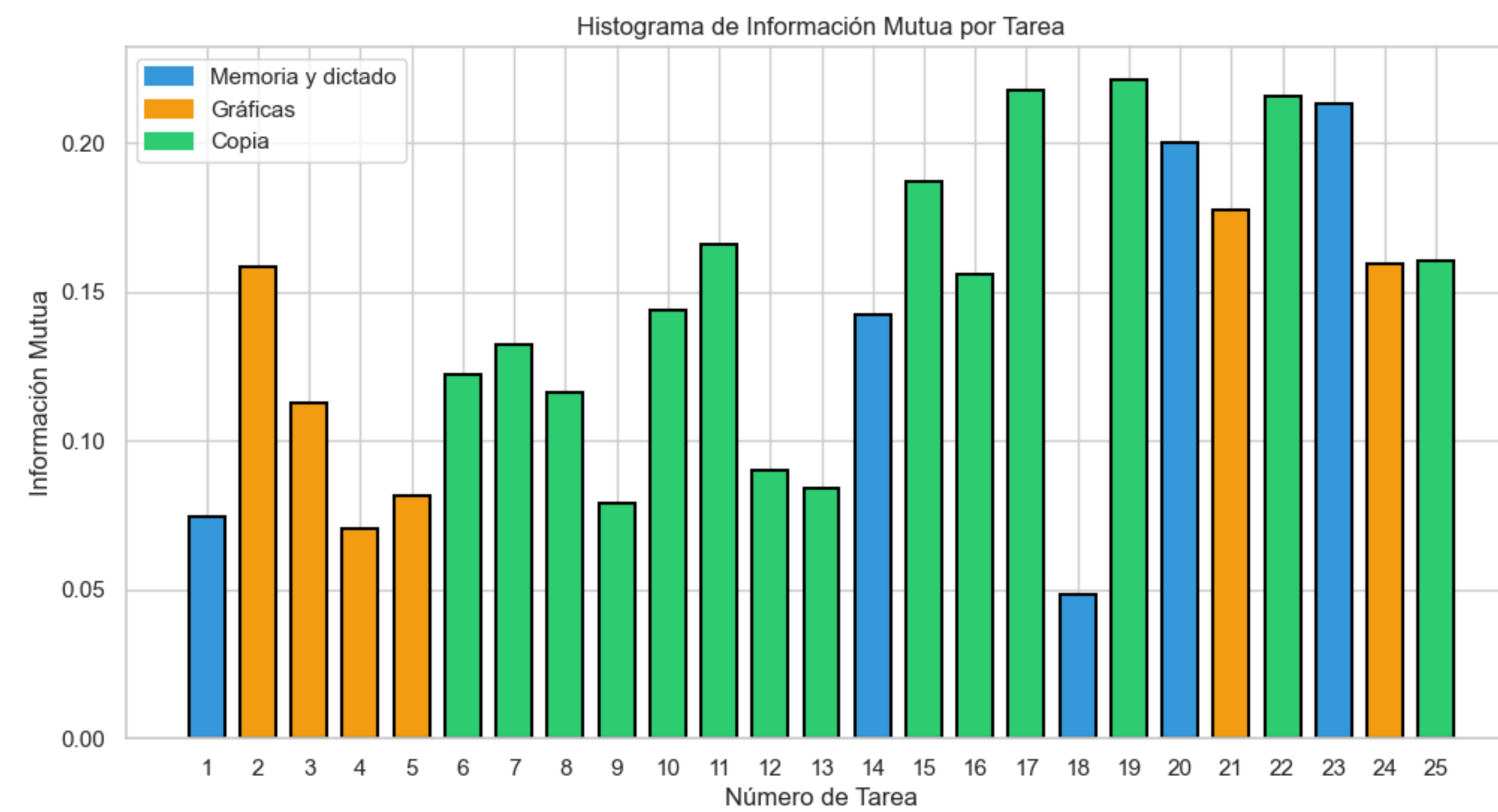


Figure 2. Información Mutua de tareas respecto al diagnóstico.

Para trabajar con menos tareas y buscar mantener la "información", se propuso utilizar para cada categoría aquellas dos tareas con mayor información mutua para diagnosticar. Para poder realizar comparación se utilizan dos modelos distintos de aprendizaje de máquinas, ambos con el objetivo de diagnosticar correctamente pero con **diferencias claves**.

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.

- Ventaja:** Permite generar modelos sofisticados muy buenos para clasificar.
- Problema:** Por lo general carecen de capacidad interpretativa.

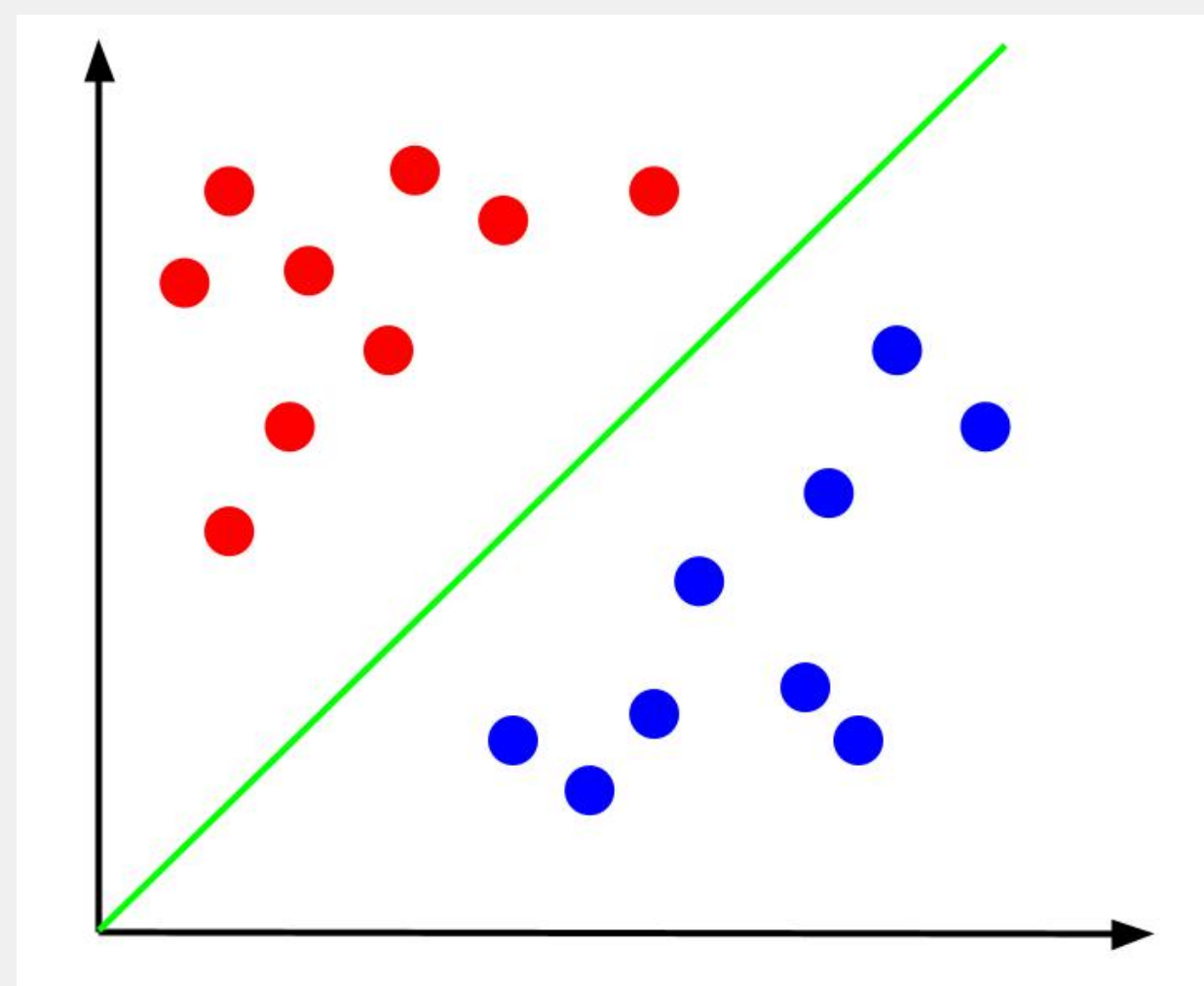
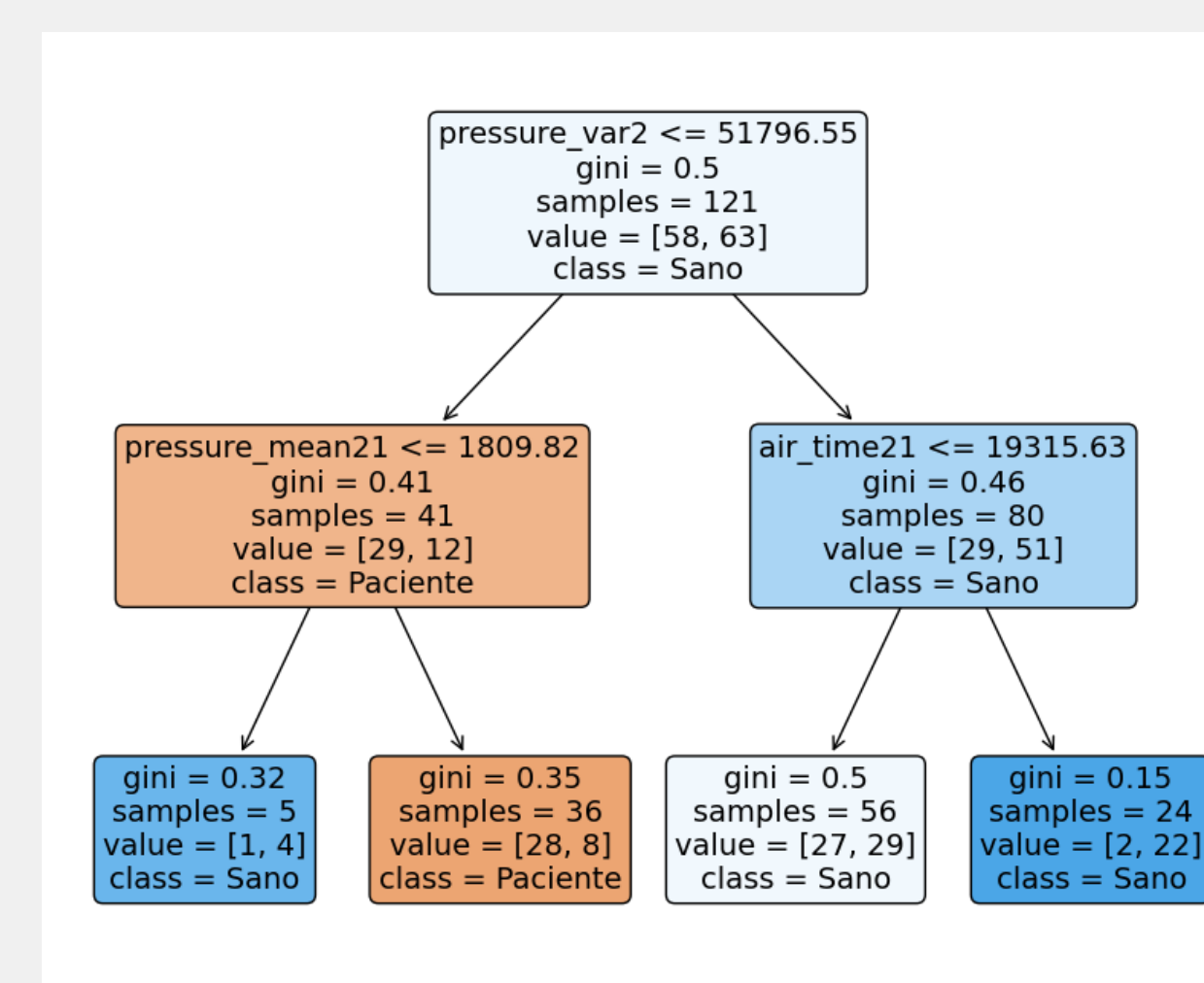


Figure 3. Ilustración de modelo SVM lineal y Ejemplo de uno de los árboles generados en el trabajo.

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

- Ventaja:** Son fáciles de interpretar permitiendo tomar acción.
- Problema:** Menos capacidad predictiva y alta sensibilidad a los datos.



Resultados

Se empleó validación cruzada de k etapas ($k = 5$) para ajustar clasificadores (**SVM** y **árboles de decisión**) con diversos parámetros para cada categoría de tarea. Los clasificadores finales fueron consolidados mediante ensambles de mayoría simple.

Se pueden resumir los resultados obtenidos mediante mostrar las **matrices de confusión**, las cuales permiten visualizar los aciertos y errores de los modelos al diagnosticar:

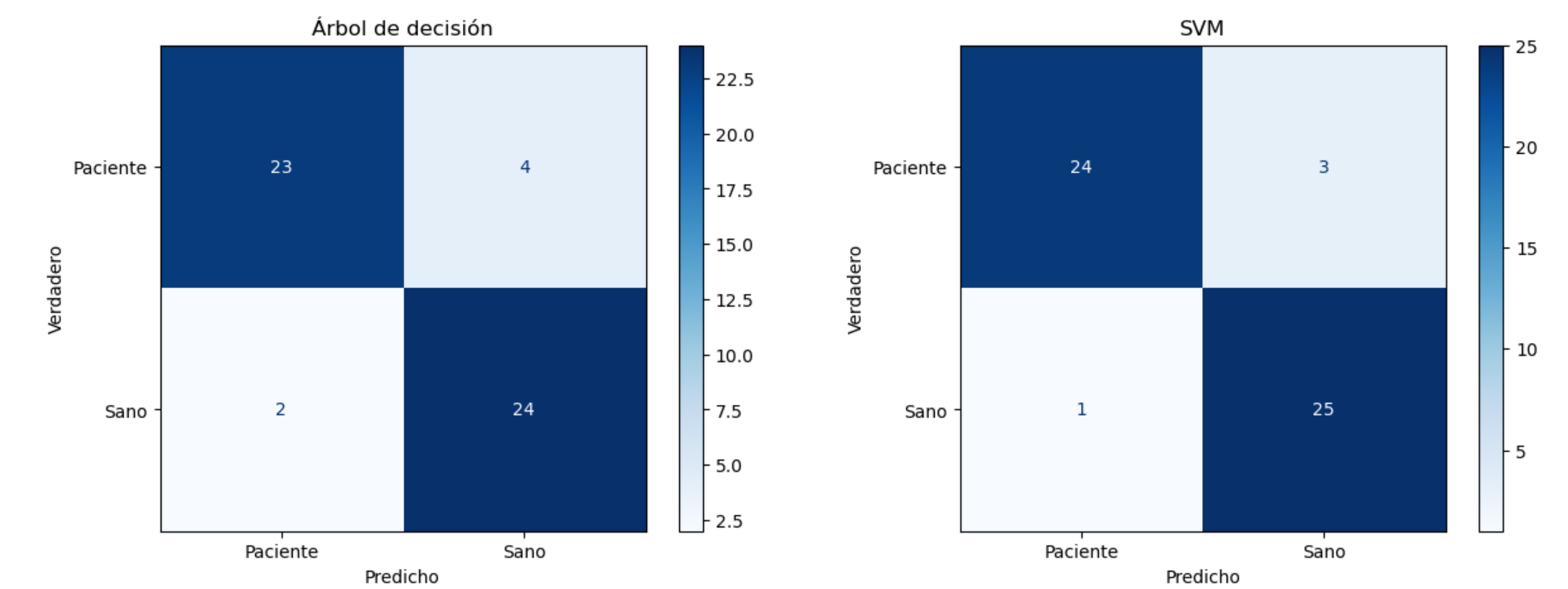


Figure 4. Matrices de confusión de los modelos.

La diagonal principal de las matrices (de color azul oscuro) muestra las personas que fueron diagnosticadas correctamente por el modelo únicamente basándose en su información entregada al realizar las tareas.

Se puede observar que el modelo de **SVM** es mejor para diagnosticar dado que comete menos errores (diagonal de color más claro), mientras que por otro lado mediante utilizar **árboles de decisión**, se obtiene un modelo un poco menos capaz pero que permite generar una ilustración que permite tener mayor entendimiento del problema (ver figura 3).

Conclusiones

Ciertamente, los modelos de aprendizaje automático ofrecen una **herramienta potente para poder diagnosticar** aún cuando se realizó una disminución de las tareas utilizadas, permitiendo implementarse con menos dificultades.

Por otra parte, se recalca que buena parte del trabajo se centra en **comprender bien la problemática** para ofrecer una solución adecuada y realista concorde a las necesidades.

Finalmente, debemos comprender que los resultados **no** indican que un modelo es mejor que otro, sino que ambos tienen sus **ventajas y desventajas** y deben ser tomadas en cuenta.

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.
- Francesco Fontanella. DARWIN. UCI Machine Learning Repository, 2022. DOI: <https://doi.org/10.24432/C55D0K>.