

Modelo Predictivo de Alzheimer

Desarrollando herramientas para predecir la probabilidad de que una persona desarrolle la enfermedad de Alzheimer, permitiendo una intervención temprana y mejorando el pronóstico de los pacientes.



¿Qué es el Alzheimer?

El Alzheimer es una **enfermedad cerebral neurodegenerativa progresiva** que destruye lentamente la memoria y las habilidades de pensamiento, afectando gravemente a la capacidad de realizar tareas diarias.

Se manifiesta inicialmente con olvidos de eventos recientes, evolucionando gradualmente.

Áreas afectadas

- Memoria a corto plazo
- Razonamiento lógico
- Capacidad de lenguaje
- Comportamiento social
- Autonomía personal

Dataset y Variables Clave

El conjunto de datos inicial contiene **35 variables clínicas** recopiladas de evaluaciones médicas. Hemos identificado las características más relevantes para nuestro modelo predictivo.



Deterioro cognitivo

Evaluación del funcionamiento mental y capacidad de procesamiento



Evaluación funcional

Medición de habilidades para tareas cotidianas



Problemas de memoria

Indicadores de pérdida de memoria a corto y largo plazo



Problemas de conducta

Cambios en comportamiento y personalidad



Actividades diarias

Dificultades en autocuidado y tareas rutinarias



Diagnóstico

Variable objetivo de nuestro modelo predictivo

Proceso de Limpieza de Datos

01

Definición del target

Establecimiento de la variable objetivo: Diagnóstico de Alzheimer

02

Eliminación de variables

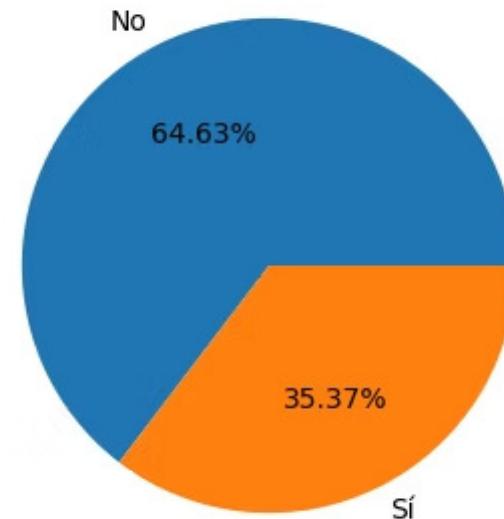
Descarte de columnas no predictivas: identificadores de paciente y doctor

03

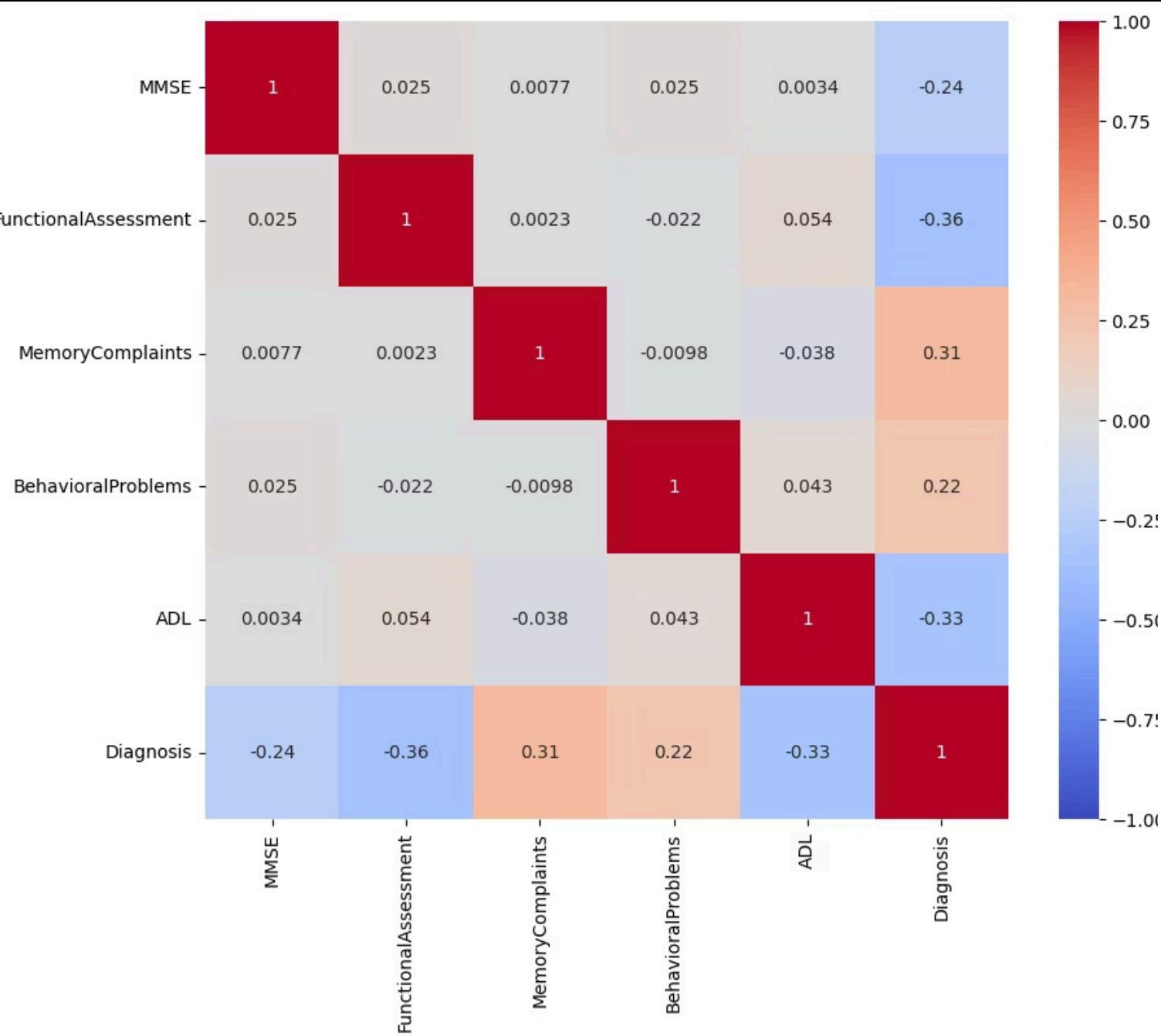
Análisis de distribución

Verificación del balance de clases en la variable objetivo

Distribución variable objetivo



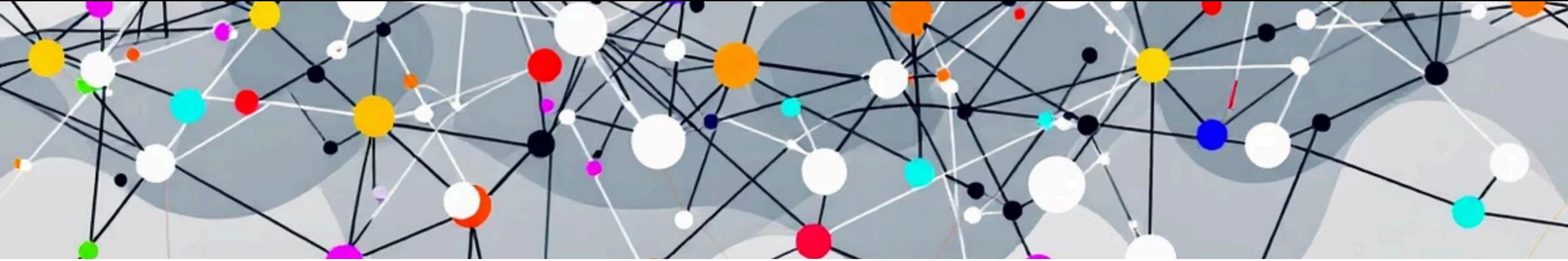
La preparación de datos es fundamental para garantizar la fiabilidad y precisión de nuestros modelos predictivos.



04 Análisis de Correlaciones

Identificamos las relaciones entre variables y el diagnóstico de Alzheimer mediante análisis de correlación.

- Las variables con mayor correlación positiva incluyen deterioro cognitivo, problemas de memoria y dificultades funcionales, confirmando su relevancia clínica.



Arquitectura de Modelos Evaluados

Para la selección del mejor modelo, evaluamos **8 diferentes** con enfoques supervisados, no supervisados y de deep learning.

Modelos Supervisados Clásicos

- Logistic Regression
- Decision Tree
- Support Vector Classifier
- Ada Boost
- Gradient Boosting
- Random Forest

Preprocesamiento KMeans

- KMeans + Random Forest
- KMeans + Logistic Regression

Deep Learning

- Red Neuronal Sequential

Modelos Descartados: Análisis

KMeans + Supervisados

Problema: Recall muy bajo para la clase positiva

Mayor complejidad sin mejora significativa en rendimiento

Red Neuronal

Problema: Resultados similares a Logistic Regression

Arquitectura más compleja con menor interpretabilidad clínica

Logistic Regression

Problema: Recall inferior a otros modelos

Aunque presenta buen AUC y sirve como baseline robusto

Decision Tree y SVC

Problema Decision Tree: Inestable con alto riesgo de overfitting

Problema SVC: Accuracy y AUC inferiores comparados con ensemble methods

Top 3: Los Finalistas

Ada Boost

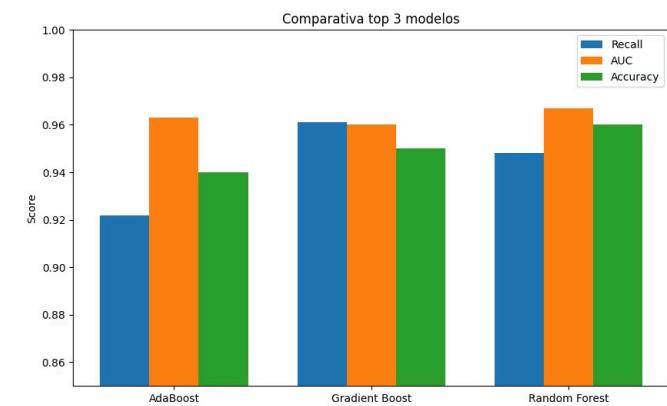
- ✓ Recall alto consistente
- ✓ AUC muy elevado
- ✓ Excelente equilibrio entre métricas

Gradient Boosting

- ✓ Mejor recall de los tres finalistas
- ✓ Accuracy muy alta
- ✗ Mayor complejidad y riesgo de overfitting

Random Forest

- ✓ Recall muy alto para clase positiva
- ✓ Mejor AUC de los tres
- ✓ Mayor estabilidad y facilidad de implementación



Conclusiones del Estudio

En el contexto del diagnóstico de Alzheimer, el **criterio más importante es maximizar el recall de la clase positiva**, minimizando los falsos negativos para no dejar casos sin detectar.



Objetivo primario

Minimizar falsos negativos: personas con Alzheimer no detectadas



Resultados obtenidos

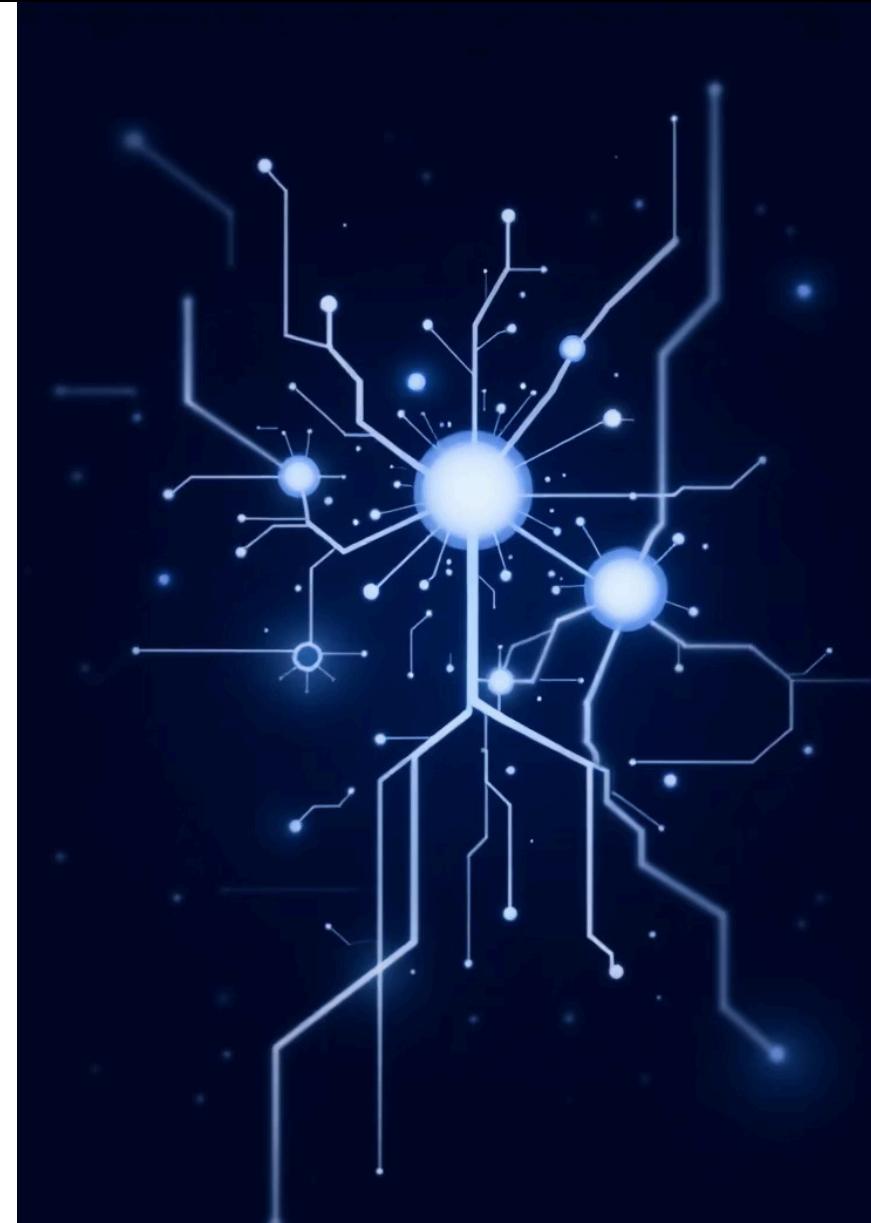
Múltiples modelos con rendimiento muy alto en todas las métricas



Modelo final: Random Forest

Mejor equilibrio entre recall positivo, AUC y estabilidad operativa

En contexto clínico, Random Forest ofrece uno de los valores más bajos de falsos negativos sin sacrificar la capacidad discriminativa global, convirtiéndolo en la opción más segura para implementación real.





¡Gracias!

Rebeca Pérez

Modelo predictivo de Alzheimer mediante Machine Learning