

## TRABAJO PRÁCTICO - BIG DATA

*Dataset: Healthcare*

En el ámbito hospitalario, la rapidez en la toma de decisiones médicas es crucial para mejorar la atención al paciente y optimizar los recursos clínicos.

Actualmente, los resultados de ciertos test médicos pueden demorar, lo que retrasa diagnósticos y tratamientos. Ante este contexto, surge la necesidad de contar con herramientas que permitan anticipar el resultado probable de un test médico, incluso antes de que esté disponible, utilizando variables clínicas y administrativas recogidas durante la internación. Esta predicción permitiría a los profesionales de la salud priorizar casos, prever complicaciones y tomar decisiones proactivas, especialmente en contextos de alta demanda o recursos limitados.

### **Definición del problema**

El objetivo del presente trabajo es desarrollar un modelo predictivo capaz de anticipar el resultado de un test médico realizado durante la internación de un paciente. Para ello, se utilizará como variable objetivo la columna "Test Results", presente en el dataset "Healthcare Dataset".

Se buscará determinar, a partir de datos clínicos y demográficos del paciente, si el resultado será "Normal", "Abnormal" o "Inconclusive", permitiendo así tomar decisiones preventivas antes de contar con el resultado físico del estudio.

La variable "Test Results" es una variable categórica de clasificación multiclase, ya que adopta tres posibles valores no numéricos ni ordinales. Cada uno representa una categoría discreta:

- ★ Normal
- ★ Abnormal
- ★ Inconclusive

Al tratarse de una variable cualitativa con tres categorías, el enfoque del modelo será de clasificación multiclase.

Anticipar el resultado de un test médico mediante un modelo predictivo representa una herramienta valiosa para los sistemas de salud. Este tipo de predicción puede contribuir significativamente a mejorar la toma de decisiones

clínicas, especialmente en contextos donde los tiempos de espera para los resultados de estudios pueden demorar el diagnóstico y el tratamiento oportuno.

En particular, si el modelo predice con alta probabilidad que un paciente tendrá un resultado "Abnormal", el equipo médico puede actuar de manera proactiva, solicitando exámenes adicionales, monitoreos más frecuentes o incluso iniciando tratamientos preventivos antes de contar con la confirmación definitiva. Esto permite reducir riesgos, mejorar los tiempos de respuesta ante casos críticos y brindar una atención más personalizada.

Además, esta herramienta puede resultar especialmente útil en situaciones de alta demanda hospitalaria, donde los recursos son limitados y se requiere priorizar pacientes según su nivel de riesgo. Por ejemplo, si los laboratorios están colapsados o los resultados se demoran por cuestiones logísticas, disponer de una predicción preliminar basada en los datos disponibles puede ser determinante para decidir a quién atender primero o a qué pacientes monitorear más de cerca.

Por otro lado, una predicción de resultado "Normal" también tiene valor, ya que puede evitar estudios innecesarios, reducir costos y optimizar el uso de equipos y personal.

En síntesis, predecir los resultados de un test médico no solo tiene un valor técnico, sino también clínico, operativo y estratégico dentro de una institución de salud.

### **Beneficios del uso del modelo**

- Anticipación diagnóstica: Permite actuar antes de tener el resultado real del test.
- Priorización inteligente: Ayuda a identificar pacientes de riesgo ("Abnormal") para atención más rápida y seguimiento intensivo.
- Eficiencia operativa: Optimiza uso de recursos médicos y evita estudios innecesarios en casos "Normal".
- Atención personalizada: Mejora la calidad del cuidado al adaptar las decisiones a cada paciente.

- Reducción de costos: Menor cantidad de estudios repetidos y menor tiempo de internación innecesaria

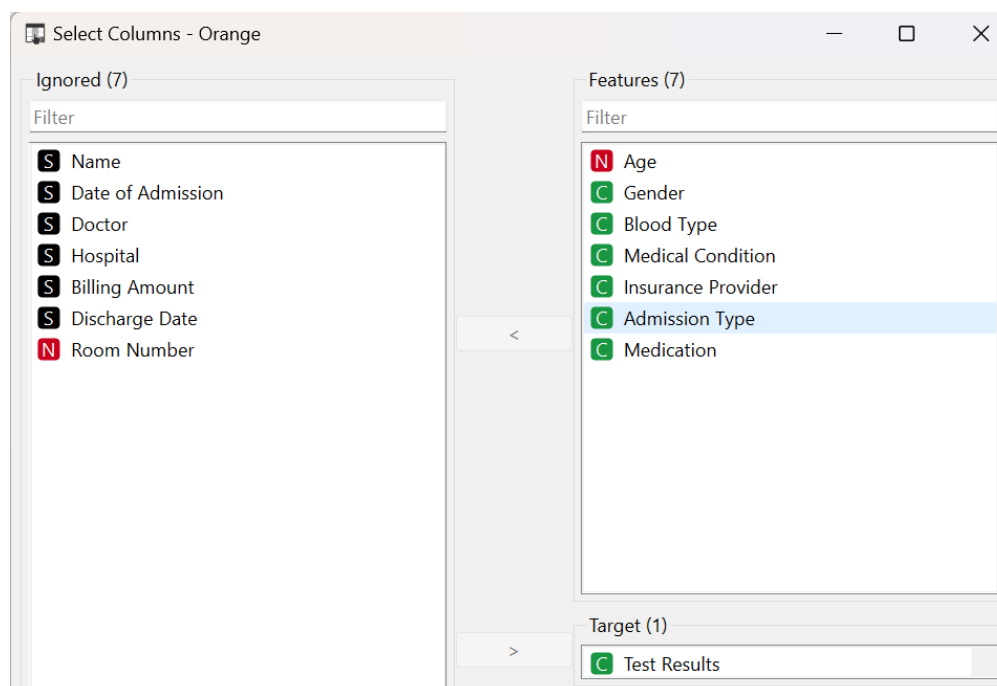
## EDA

Este dataset está compuesto por 55.500 filas y 15 columnas, de ellas 7 son categóricas, 2 numericas y 6 de texto. No hay valores faltantes

El análisis exploratorio de este conjunto de datos se ve limitado debido a que la mayoría de las variables son de tipo categórico. Si bien estas variables pueden contener información valiosa, su tratamiento presenta ciertas limitaciones tanto en el análisis como en la modelización y se reduce la posibilidad de generar visualizaciones gráficas significativas, ya que no permiten calcular métricas estadísticas como media, mediana o dispersión, ni representar relaciones de forma continua. Esto dificulta la identificación de patrones visuales, correlaciones o tendencias claras dentro de los datos. Esto limitó el uso de modelos más robustos y la exploración visual profunda, lo que posiblemente contribuyó al bajo poder predictivo observado en los modelos evaluados.






Variables independientes utilizadas

Para predecir el resultado del test médico, se utilizarán como variables predictoras las siguientes columnas del dataset:



La variable target es = Test Results

Valores de cada variable

	Name	Type	Role	Values
1	Age	 numeric	feature	
2	Gender	 categorical	feature	Female, Male
3	Blood Type	 categorical	feature	A+, A-, AB+, AB-, B+, B-, O+, O-
4	Medical ...	 categorical	feature	Arthritis, Asthma, Cancer, Diabetes, Hypertension, Obesity
5	Insurance ...	 categorical	feature	Aetna, Blue Cross, Cigna, Medicare, UnitedHealthcare
6	Room Number	 numeric	feature	
7	Admission Type	 categorical	feature	Elective, Emergency, Urgent
8	Medication	 categorical	feature	Aspirin, Ibuprofen, Lipitor, Paracetamol, Penicillin

Estas variables contienen información relevante sobre el estado clínico y contexto del paciente al momento de la internación, y se espera que tengan relación con los resultados de los estudios médicos.

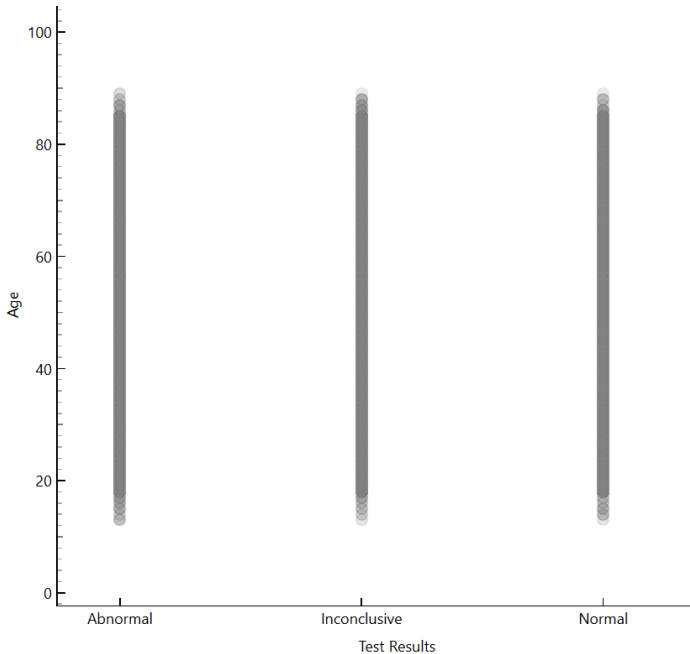
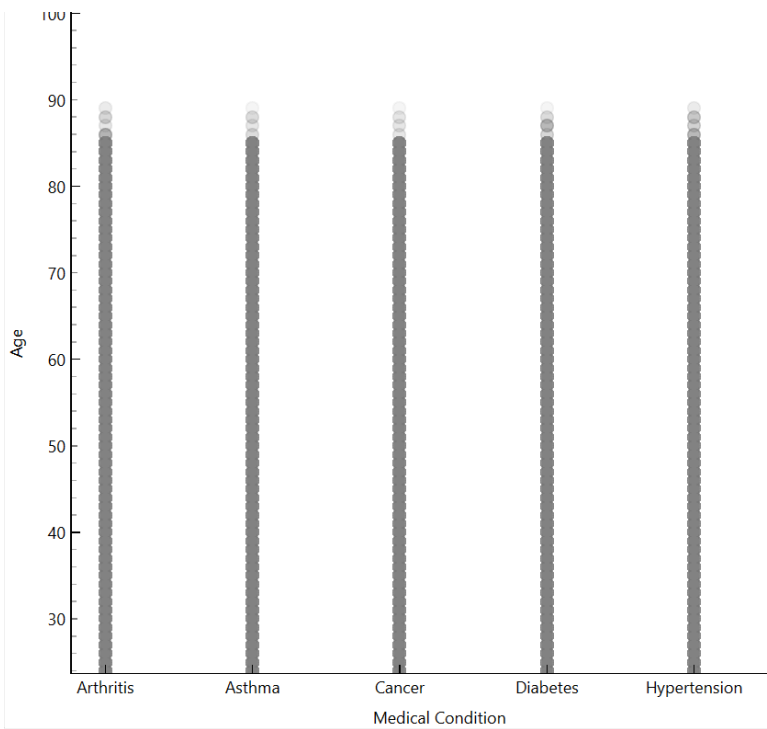
### Hipótesis sobre la predicción

A partir del análisis de las variables disponibles en el dataset, se seleccionaron como variables independientes aquellas que se consideraron potencialmente relevantes para predecir el resultado del test médico. En particular, se identificaron como más influyentes las variables Age, Medical Condition y Admission Type (clasificada como Emergency, Elective o Urgent).

1. Age es una variable numérica continua que aporta información sobre el perfil etario del paciente.
2. Medical Condition refleja el diagnóstico o condición médica registrada en la internación. Aunque es una variable categórica, se considera altamente informativa, ya que las condiciones clínicas del paciente son un factor determinante en la evolución y resultados de estudios médicos.
3. Admission Type, que distingue entre admisiones de tipo emergency, elective y urgent, también se identificó como relevante. Este dato aporta contexto sobre la gravedad del ingreso, lo cual puede estar correlacionado con la probabilidad de obtener un resultado "Abnormal" en los tests realizados.

Estas variables, combinadas, aportan una base clínicamente coherente para construir un modelo predictivo, y fueron incluidas en los modelos desarrollados para evaluar su capacidad de anticipar el resultado del test.

Gráficos:





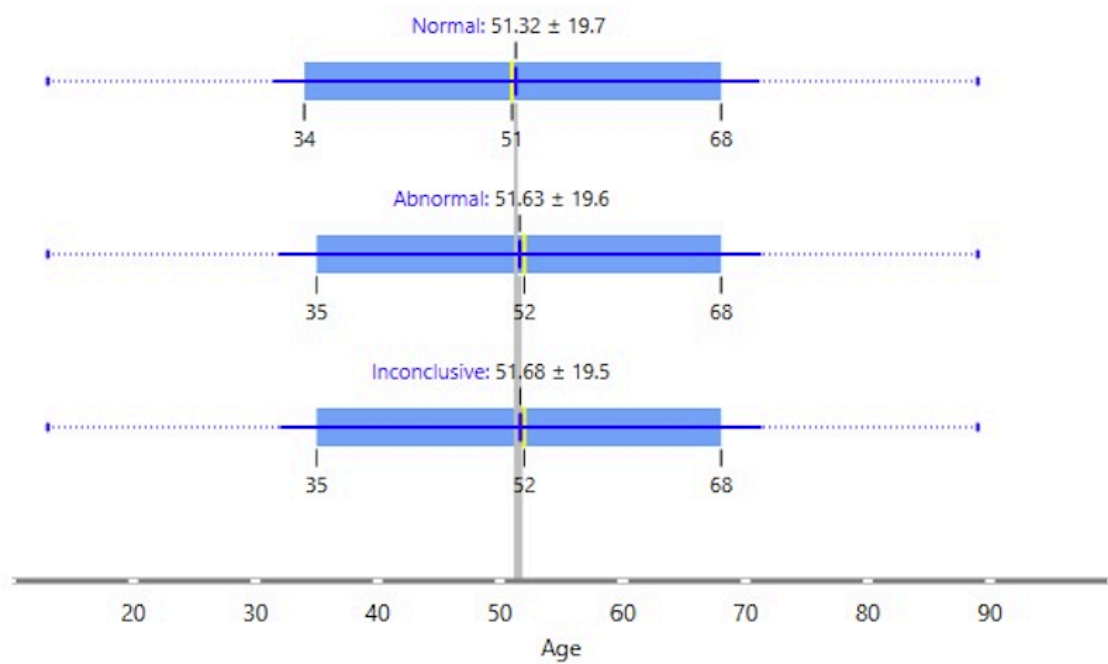
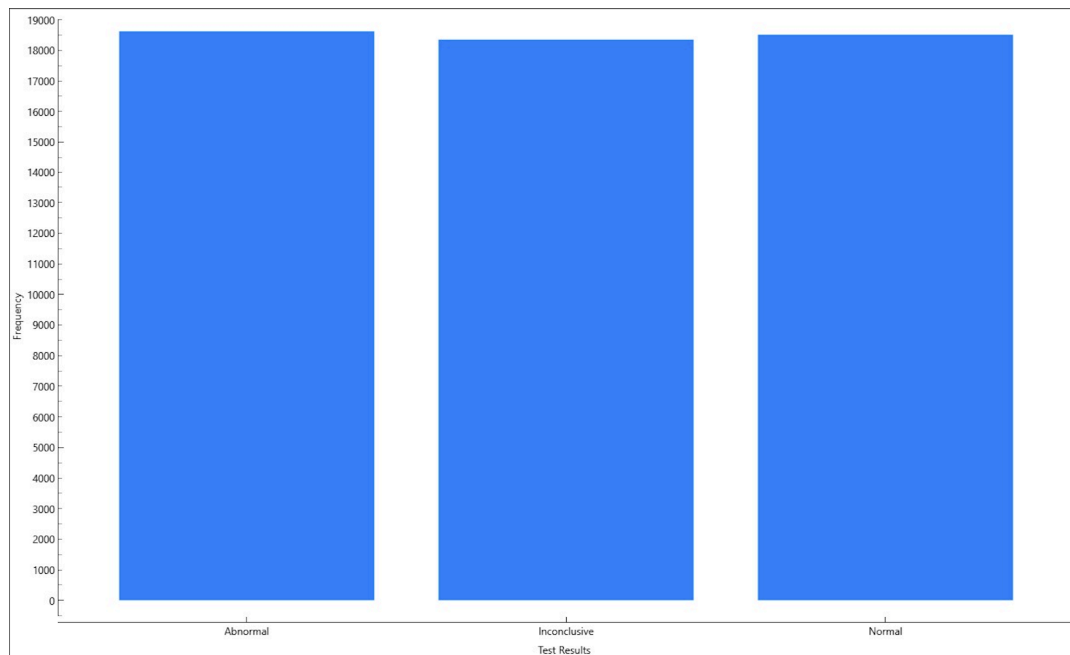
Haciendo el gráfico de “feature statistics” la variable AGE por ser numérica no presenta valores faltantes, por lo que pudo ser utilizada sin necesidad de imputaciones.

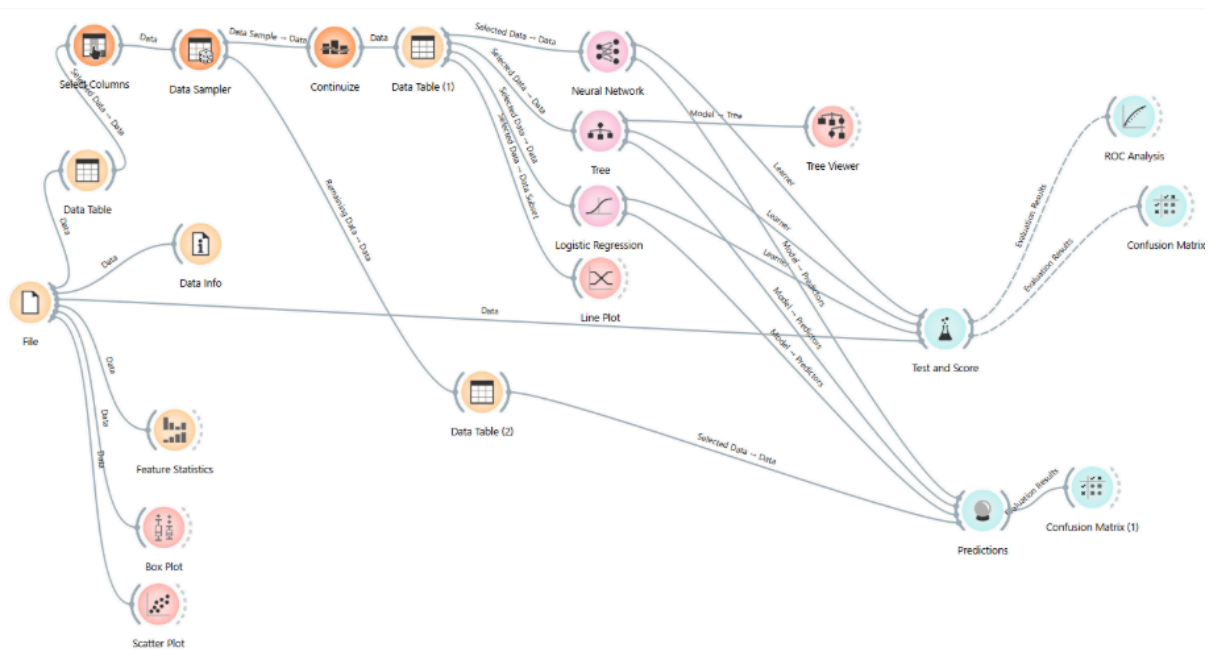
- La edad promedio de los pacientes es de 51,54 años,
- La mediana de 52 años
- La moda es de 38 años, lo cual indica una distribución aproximadamente simétrica.
- Los valores oscilan entre los 13 años (mínimo) y los 89 años (máximo), abarcando una población predominantemente adulta.

El gráfico de distribución muestra cómo se reparte la variable Test results a lo largo de distintos tramos de edad, con una presencia equilibrada de las tres clases objetivo (Abnormal, Inconclusive, Normal) en todos los rangos.

Esto permite anticipar que, si bien la edad no define por sí sola un resultado de test, podría aportar información relevante al momento de predecirlo, especialmente cuando se la combina con otras variables.

En conclusión, Age es una variable con buena variabilidad y potencial valor predictivo, por lo cual fue incluida en los modelos construidos.





## Análisis de Métricas

Para evaluar el rendimiento de los modelos predictivos construidos, se utilizaron diferentes métricas: AUC, accuracy (CA), F1 Score, precisión, recall y MCC. Cada una de estas métricas brinda información complementaria sobre el comportamiento del modelo, especialmente en problemas de clasificación multiclase como el presente.

En este caso, el contexto médico en el que se enmarca el problema es fundamental para decidir qué métricas priorizar. Al tratarse de un hospital, minimizar tanto los falsos negativos como los falsos positivos es crítico, ya que ambos errores pueden tener consecuencias graves.

Los falsos negativos representan un riesgo clínico importante, ya que un paciente con resultado "Abnormal" podría no recibir atención prioritaria o seguimiento temprano. Minimizar estos casos es prioritario para evitar complicaciones en pacientes críticos. Para estos casos, el Recall es una métrica clave.

Los falsos positivos son pacientes "Normal" o "Inconclusive" clasificados como "Abnormal". Podrían recibir estudios adicionales innecesarios, generando sobrecarga de recursos. La precisión permite controlar este efecto.

Por esta razón, la métrica prioritaria para la evaluación fue el F1 Score, ya que representa el balance entre la precisión y el recall, y penaliza ambos tipos de error por igual. En segundo lugar, se consideró particularmente importante el



recall, porque refleja la capacidad del modelo para detectar correctamente los casos positivos reales (en este caso, pacientes con resultados “Abnormal”).

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.548	0.386	0.386	0.386	0.386	0.080
Neural Network	0.514	0.346	0.345	0.345	0.346	0.018
Logistic Regression	0.499	0.334	0.330	0.334	0.334	0.000

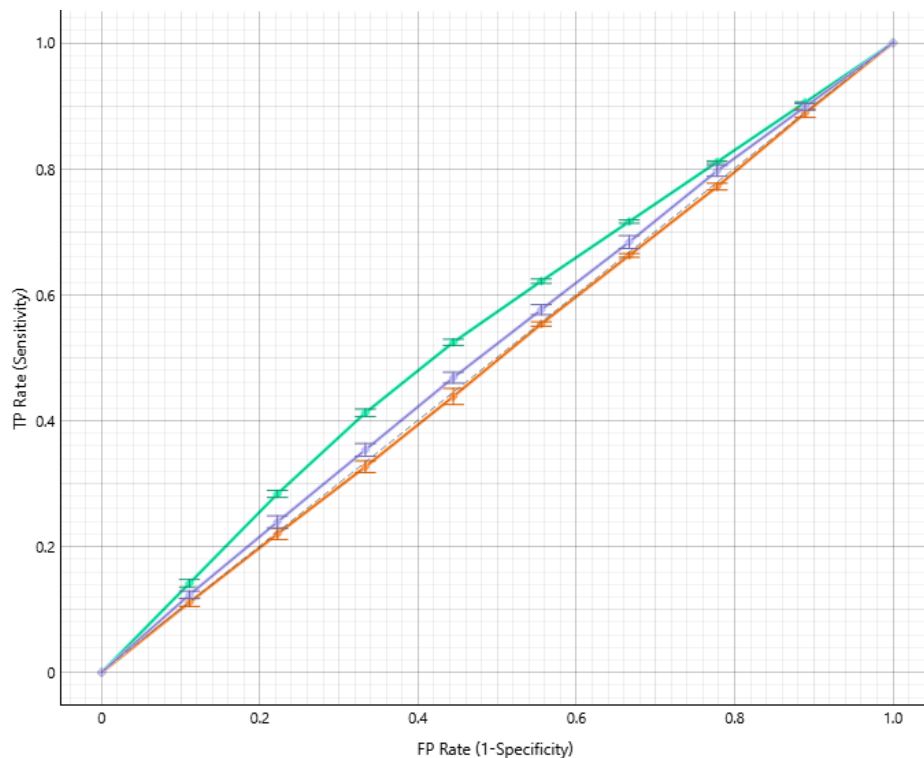
### Análisis de modelos

Se entrenaron y evaluaron tres modelos de clasificación supervisada: regresión logística, red neuronal y árbol de decisión, con el objetivo de anticipar el resultado de un test médico durante la internación de un paciente.

- 1. Regresión logística:** Este modelo fue el de peor rendimiento. Su F1 Score como su recall son bajos, lo cual muestra que sus predicciones no aportan valor respecto al azar.
- 2. Red neuronal:** mostró una leve mejora respecto a la regresión logística, pero su rendimiento sigue siendo limitado. Su F1 y recall son ligeramente más altos, pero la diferencia no es significativa.
- 3. Árbol de decisión:** fue el modelo con mejor rendimiento general. Si bien sus métricas aún son moderadas, es el único modelo que supera consistentemente a los otros en todas las métricas clave, en especial F1 y recall. Además, su estructura permite interpretar fácilmente qué variables influyen en la predicción, lo que es una ventaja en entornos médicos donde la transparencia del modelo es importante.

En función de las métricas seleccionadas, el modelo de árbol de decisión resultó ser el más adecuado, alcanzando los mejores valores y mostrando mayor robustez que los otros algoritmos evaluados. El árbol de decisión se considera el modelo más útil y confiable dentro del conjunto probado, tanto por su rendimiento como por su capacidad de interpretación clínica, aunque presenta oportunidades de mejora.

En segundo lugar, las curvas ROC para la clase "Abnormal" muestran que ninguno de los modelos logra una discriminación clara entre los casos positivos y negativos. Las curvas se ubican muy próximas a la línea de azar (diagonal), lo que evidencia una baja capacidad predictiva en términos de sensibilidad y especificidad.



Si bien el árbol de decisión presenta una curva ligeramente más elevada que el resto, la diferencia es mínima y no representa una mejora sustancial en términos clínicos. Este resultado refuerza la conclusión de que, aunque el árbol fue el mejor modelo entre los evaluados, su capacidad para predecir la clase "Abnormal" sigue siendo limitada, y podría mejorarse con ajustes adicionales o el uso de modelos más complejos y técnicas avanzadas.

### **Confusion matrix del conjunto entrenamiento**

Las siguientes matrices comparan los resultados reales con los obtenidos en cada modelo. Se puede ver qué tan bien "aprendió" el modelo para predecir correctamente.

### **Tree**

		Predicted			
		Abnormal	Inconclusive	Normal	$\Sigma$
Actual	Abnormal	7753	5575	5299	18627
	Inconclusive	6116	7079	5161	18356
	Normal	6296	5609	6612	18517
$\Sigma$		20165	18263	17072	55500

En este cuadro de doble entrada, puede verse la debilidad con los “falsos negativos”:

- De los 18.627 casos reales de “Abnormal”, el modelo predijo correctamente 7.753, mientras que cometió 10.874 errores, clasificándolos como “Inconclusive” (5.575 casos) o “Normal” (5.299 casos).
- En cuanto a los 18.356 casos reales de “Inconclusive”, el modelo acertó 7.079 veces, pero falló en 11.277, clasificándolos como “Abnormal” (6.116) o “Normal” (5.161).
- Finalmente, de los 18.517 casos reales de “Normal”, el modelo clasificó correctamente 6.612, y se equivocó en 11.905, etiquetándolos como “Abnormal” (6.296) o “Inconclusive” (5.609).

En todos los casos, el número de predicciones incorrectas supera ampliamente al número de aciertos. La clase “Abnormal” fue la que obtuvo el mayor número absoluto de predicciones correctas (7.753), pero aun así representa menos de la mitad de los casos reales de esa categoría.

Esto evidencia que el modelo tiene dificultades importantes para distinguir entre las clases, y que existe una gran cantidad de confusiones cruzadas, especialmente entre “Abnormal” e “Inconclusive”.

## Regresión Logística

		Predicted			
		Abnormal	Inconclusive	Normal	$\Sigma$
Actual	Abnormal	7902	4481	6244	<b>18627</b>
	Inconclusive	7832	4472	6052	<b>18356</b>
	Normal	7944	4410	6163	<b>18517</b>
$\Sigma$		<b>23678</b>	<b>13363</b>	<b>18459</b>	<b>55500</b>

En este cuadro de doble entrada, puede verse la debilidad con los “falsos negativos”:

- De los 18.627 casos reales de “Abnormal”, el modelo predijo correctamente 7.902, mientras que cometió 10.725 errores, clasificándolos como “Inconclusive” en 4.481 casos y como “Normal” en 6.244 casos.
- Para los 18.356 casos reales de “Inconclusive”, el modelo acertó 4.472 veces, y se equivocó en 13.884, clasificándolos erróneamente como “Abnormal” (7.832) o “Normal” (6.052).
- En cuanto a los 18.517 casos reales de “Normal”, el modelo predijo correctamente 6.163, pero se equivocó en 12.354 ocasiones, asignando la clase “Abnormal” en 7.944 casos y “Inconclusive” en 4.410.

En resumen, el modelo de regresión logística muestra un comportamiento muy similar al azar. En las tres clases, los errores superan ampliamente a los aciertos, siendo especialmente bajo el desempeño en la clase “Inconclusive”, donde más del 75% de los casos fueron mal clasificados.

Además, se observa una tendencia a sobrepredecir la clase “Abnormal”, ya que fue asignada en 23.678 casos (la más alta), aún cuando solo hay 18.627 casos reales de esa clase. Esto genera muchos falsos positivos, lo cual es preocupante en un contexto médico.

## Red Neuronal

		Predicted			
		Abnormal	Inconclusive	Normal	$\Sigma$
Actual	Abnormal	7134	5158	6335	18627
	Inconclusive	6565	5471	6320	18356
	Normal	6668	5278	6571	18517
$\Sigma$		20367	15907	19226	55500

En este cuadro de doble entrada, puede verse la debilidad con los “falsos negativos”:

- De los 18.627 casos reales de “Abnormal”, el modelo predijo correctamente 7.134, mientras que cometió 11.493 errores, clasificándolos como “Inconclusive” (5.158 casos) o “Normal” (6.335 casos).
- De los 18.356 casos reales de “Inconclusive”, el modelo acertó 5.471 veces y se equivocó en 12.885 ocasiones, clasificándolos como “Abnormal” (6.565) o “Normal” (6.320).
- En el caso de los 18.517 casos reales de “Normal”, el modelo clasificó correctamente 6.571, pero cometió 11.946 errores, etiquetándolos como “Abnormal” (6.668) o “Inconclusive” (5.278).

Al igual que en los otros modelos, el número de errores supera ampliamente al de aciertos en todas las clases. La red neuronal predice bien entre el 35% y 38% de los casos reales, con una ligera mejora en la clase “Normal”.

También se observa que la clase “Abnormal” fue sobre predicha en 20.367 casos, lo que refleja una tendencia del modelo a sobrediagnosticar resultados negativos, generando un gran número de falsos positivos. Esto, en contexto médico, podría derivar en tratamientos innecesarios o preocupación injustificada para pacientes sanos.

Las matrices de confusión presentadas corresponden al análisis del rendimiento de cada modelo durante la etapa de entrenamiento. En este caso, los datos utilizados provienen del mismo conjunto que fue dividido para entrenar y validar internamente los modelos.

Analizar la matriz de confusión en esta etapa permite entender cómo se comporta el modelo cuando se enfrenta a datos que ya “conoce” parcialmente, es decir, que provienen de la misma distribución utilizada para su ajuste. A través de estas matrices, es posible identificar con mayor detalle cuántos casos de cada clase fueron clasificados correctamente y en qué medida el modelo se equivoca, mostrando si tiende a confundir ciertas clases más que otras.

Este análisis es útil para detectar desbalance de clases, sesgos del modelo, y para evaluar si el algoritmo está aprendiendo patrones relevantes o simplemente respondiendo al azar. Sin embargo, es importante aclarar que este tipo de evaluación no refleja la capacidad real de generalización del modelo, por lo que debe complementarse con pruebas sobre un conjunto de datos independientes (testeo real) para validar su comportamiento en la práctica.

## TEST

### Tree

		Predicted			
		Abnormal	Inconclusive	Normal	$\Sigma$
Actual	Abnormal	1550	1126	1073	3749
	Inconclusive	1286	1292	1065	3643
	Normal	1250	1160	1298	3708
$\Sigma$		4086	3578	3436	11100

En la matriz de confusión correspondiente al modelo de árbol de decisión evaluado sobre el conjunto de testeo, se observa lo siguiente:

- De los 3.749 casos reales de “Abnormal”, el modelo predijo correctamente 1.550, mientras que cometió 2.199 errores, clasificándolos erróneamente como “Inconclusive” (1.126 casos) o “Normal” (1.073 casos).
- En los 3.643 casos reales de “Inconclusive”, el modelo acertó 1.292 veces, pero se equivocó en 2.351, etiquetándolos como “Abnormal” en 1.286 casos y como “Normal” en 1.065.
- Por último, de los 3.708 casos reales de “Normal”, solo 1.298 fueron correctamente clasificados, mientras que 2.410 se clasificaron mal: 1.250 como “Abnormal” y 1.160 como “Inconclusive”.

En todos los casos, el número de errores supera ampliamente al de aciertos, manteniéndose la tendencia observada durante el entrenamiento. El modelo no logra identificar correctamente más del 42% de los casos por clase, lo que refleja una capacidad limitada de predicción sobre datos nuevos.

## Regresión logística

		Predicted			$\Sigma$
		Abnormal	Inconclusive	Normal	
Actual	Abnormal	1442	1025	1282	3749
	Inconclusive	1415	1004	1224	3643
	Normal	1518	992	1198	3708
$\Sigma$		4375	3021	3704	11100

En la matriz de confusión correspondiente al modelo de regresión logística evaluado sobre el conjunto de test, se observa lo siguiente:

- De los 3.749 casos reales de "Abnormal", el modelo predijo correctamente 1.442, pero cometió 2.307 errores, clasificándolos como "Inconclusive" en 1.025 casos y como "Normal" en 1.282.
- En cuanto a los 3.643 casos reales de "Inconclusive", el modelo acertó 1.004 veces, mientras que falló en 2.639 ocasiones, etiquetándolos como "Abnormal" (1.415) o "Normal" (1.224).
- Por último, de los 3.708 casos reales de "Normal", predijo correctamente 1.198, pero se equivocó en 2.510, clasificándolos como "Abnormal" (1.518) o "Inconclusive" (992).

En los tres casos, el número de errores es significativamente mayor al de aciertos. Además, se nota una fuerte tendencia del modelo a predecir "Abnormal", ya que esa clase fue asignada 4.375 veces, a pesar de haber solo 3.749 casos reales con ese valor. Esto genera muchos falsos positivos, especialmente en las clases "Normal" e "Inconclusive".

## Red neuronal

		Predicted			
		Abnormal	Inconclusive	Normal	$\Sigma$
Actual	Abnormal	1014	1497	1238	3749
	Inconclusive	916	1504	1223	3643
	Normal	910	1482	1316	3708
$\Sigma$		2840	4483	3777	11100

En la matriz de confusión del modelo de red neuronal evaluado sobre el conjunto de test, se observa lo siguiente:

- De los 3.749 casos reales de "Abnormal", el modelo clasificó correctamente 1.014, pero se equivocó en 2.735 casos, prediciendo como "Inconclusive" a 1.497 y como "Normal" a 1.238.
- Para los 3.643 casos reales de "Inconclusive", el modelo acertó 1.504 veces, mientras que cometió 2.139 errores, clasificándolos como "Abnormal" en 916 casos y como "Normal" en 1.223.
- En los 3.708 casos reales de "Normal", predijo correctamente 1.316 casos, pero se equivocó en 2.392, asignando la etiqueta "Abnormal" en 910 casos y "Inconclusive" en 1.482.

A diferencia de los otros modelos, la red neuronal predice más frecuentemente la clase "Inconclusive" (4.483 veces), lo que representa un posible sesgo hacia esa categoría. A su vez, los errores son significativos en todas las clases, con entre 58% y 73% de clasificaciones incorrectas por clase.

En general, el modelo muestra bajo nivel de precisión y una distribución de errores bastante equilibrada, pero sin lograr destacar en ninguna clase. Esto indica que la red neuronal, aunque levemente más balanceada, tampoco logra generalizar bien sobre datos nuevos.

## Resumen confusion matrix

El análisis conjunto de las matrices de confusión de los tres modelos evaluados (árbol de decisión, regresión logística y red neuronal) muestra un patrón común preocupante: todos los modelos cometen una alta proporción de errores en la



clasificación de los resultados del test médico, especialmente al confundir pacientes verdaderamente enfermos con otras categorías.

En particular, se observa que los modelos tienden a clasificar erróneamente muchos casos reales de "Abnormal" como "Inconclusive" o incluso "Normal". En algunos modelos, más del 50% de los pacientes con resultado "Abnormal" fueron mal clasificados. Esta situación es crítica en el contexto hospitalario, ya que implica el riesgo de no detectar correctamente a pacientes que podrían requerir atención urgente o tratamiento inmediato.

Por otro lado, también se registran numerosos falsos positivos, donde pacientes sanos (por ejemplo, con resultado real "Normal") son clasificados como "Abnormal". Si bien este tipo de error puede parecer menos grave, en el entorno clínico puede conducir a sobre diagnósticos, tratamientos innecesarios, estrés en el paciente y mal uso de recursos.

Aunque los modelos logran cierto grado de predicción, la tasa de errores es demasiado alta para su uso como herramienta de apoyo clínico confiable en su estado actual.

	Tree	error	Logistic Regression	error	Neural Network	error	Test Results
1	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.33 : 0.33 : 0.34 → Normal	0.657	0.30 : 0.54 : 0.16 → Inconclusive	0.835	Normal
2	1.00 : 0.00 : 0.00 → Abnormal	0.000	0.34 : 0.32 : 0.34 → Abnormal	0.660	0.40 : 0.30 : 0.30 → Abnormal	0.596	Abnormal
3	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.32 : 0.34 : 0.34 → Inconclusive	0.665	0.30 : 0.32 : 0.39 → Normal	0.614	Normal
4	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.34 : 0.32 : 0.34 → Normal	0.682	0.34 : 0.28 : 0.39 → Normal	0.723	Inconclusive
5	1.00 : 0.00 : 0.00 → Abnormal	0.000	0.35 : 0.31 : 0.34 → Abnormal	0.654	0.48 : 0.21 : 0.31 → Abnormal	0.519	Abnormal
6	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.36 : 0.32 : 0.33 → Abnormal	0.684	0.24 : 0.25 : 0.50 → Normal	0.747	Inconclusive
7	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.34 : 0.33 : 0.33 → Abnormal	0.668	0.19 : 0.55 : 0.26 → Inconclusive	0.738	Normal
8	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.33 : 0.34 : 0.33 → Inconclusive	0.661	0.42 : 0.30 : 0.28 → Abnormal	0.698	Inconclusive
9	1.00 : 0.00 : 0.00 → Abnormal	0.000	0.35 : 0.32 : 0.33 → Abnormal	0.654	0.27 : 0.40 : 0.33 → Inconclusive	0.727	Abnormal
10	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.34 : 0.34 : 0.32 → Abnormal	0.677	0.23 : 0.44 : 0.33 → Inconclusive	0.668	Normal
11	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.34 : 0.33 : 0.34 → Normal	0.663	0.36 : 0.37 : 0.27 → Inconclusive	0.733	Normal
12	1.00 : 0.00 : 0.00 → Abnormal	0.000	0.33 : 0.32 : 0.35 → Normal	0.667	0.43 : 0.18 : 0.40 → Abnormal	0.573	Abnormal
13	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.32 : 0.33 : 0.35 → Normal	0.667	0.24 : 0.26 : 0.50 → Normal	0.739	Inconclusive
14	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.36 : 0.33 : 0.31 → Abnormal	0.686	0.34 : 0.24 : 0.42 → Normal	0.580	Normal
15	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.33 : 0.33 : 0.34 → Normal	0.668	0.19 : 0.33 : 0.48 → Normal	0.675	Inconclusive
16	1.00 : 0.00 : 0.00 → Abnormal	1.000	0.34 : 0.32 : 0.34 → Normal	0.683	0.59 : 0.13 : 0.28 → Abnormal	0.865	Inconclusive
17	1.00 : 0.00 : 0.00 → Abnormal	0.000	0.35 : 0.33 : 0.32 → Abnormal	0.654	0.23 : 0.43 : 0.35 → Inconclusive	0.774	Abnormal
1136	0.00 : 0.00 : 1.00 → Normal	0.000	0.35 : 0.32 : 0.33 → Abnormal	0.669	0.22 : 0.43 : 0.35 → Inconclusive	0.646	Normal
1137	0.00 : 0.00 : 1.00 → Normal	0.000	0.35 : 0.33 : 0.32 → Abnormal	0.677	0.30 : 0.45 : 0.25 → Inconclusive	0.751	Normal
1138	0.00 : 0.00 : 1.00 → Normal	0.000	0.33 : 0.33 : 0.33 → Abnormal	0.667	0.13 : 0.36 : 0.51 → Normal	0.488	Normal
1139	0.00 : 0.00 : 1.00 → Normal	1.000	0.33 : 0.32 : 0.35 → Normal	0.680	0.30 : 0.43 : 0.28 → Inconclusive	0.573	Inconclusive
1140	0.00 : 0.00 : 1.00 → Normal	0.000	0.34 : 0.32 : 0.34 → Normal	0.662	0.33 : 0.49 : 0.18 → Inconclusive	0.822	Normal
1141	0.00 : 0.00 : 1.00 → Normal	0.000	0.32 : 0.33 : 0.35 → Normal	0.653	0.26 : 0.41 : 0.33 → Inconclusive	0.666	Normal
1142	0.00 : 0.00 : 1.00 → Normal	0.000	0.33 : 0.34 : 0.33 → Inconclusive	0.666	0.21 : 0.41 : 0.39 → Inconclusive	0.613	Normal
1143	0.00 : 0.00 : 1.00 → Normal	0.000	0.32 : 0.34 : 0.34 → Inconclusive	0.663	0.27 : 0.52 : 0.21 → Inconclusive	0.792	Normal
1144	0.00 : 0.00 : 1.00 → Normal	1.000	0.32 : 0.34 : 0.34 → Normal	0.677	0.28 : 0.34 : 0.38 → Normal	0.723	Abnormal
1145	0.00 : 0.00 : 1.00 → Normal	1.000	0.33 : 0.32 : 0.34 → Normal	0.677	0.16 : 0.16 : 0.68 → Normal	0.836	Inconclusive
1146	0.00 : 0.00 : 1.00 → Normal	1.000	0.34 : 0.33 : 0.33 → Abnormal	0.670	0.34 : 0.25 : 0.41 → Normal	0.755	Inconclusive
1147	0.00 : 0.00 : 1.00 → Normal	1.000	0.34 : 0.33 : 0.33 → Abnormal	0.660	0.33 : 0.55 : 0.13 → Inconclusive	0.673	Abnormal
1148	0.00 : 0.00 : 1.00 → Normal	0.000	0.33 : 0.34 : 0.34 → Inconclusive	0.664	0.36 : 0.51 : 0.12 → Inconclusive	0.876	Normal
1149	0.00 : 0.00 : 1.00 → Normal	0.000	0.36 : 0.33 : 0.31 → Abnormal	0.692	0.35 : 0.25 : 0.40 → Normal	0.600	Normal
1150	0.00 : 0.00 : 1.00 → Normal	0.000	0.34 : 0.33 : 0.33 → Abnormal	0.667	0.18 : 0.36 : 0.46 → Normal	0.536	Normal
1151	0.00 : 0.00 : 1.00 → Normal	1.000	0.34 : 0.33 : 0.33 → Abnormal	0.663	0.33 : 0.27 : 0.40 → Normal	0.671	Abnormal
1152	0.00 : 0.00 : 1.00 → Normal	0.000	0.33 : 0.33 : 0.34 → Normal	0.662	0.40 : 0.25 : 0.35 → Abnormal	0.647	Normal
1153	0.00 : 0.00 : 1.00 → Normal	1.000	0.34 : 0.33 : 0.33 → Abnormal	0.670	0.30 : 0.51 : 0.20 → Inconclusive	0.490	Normal

La imagen muestra resultados de los modelos. En la línea 1144, seleccionada en color azul, se puede ver un “falso negativo”, en donde el paciente obtuvo resultado de test “Abnormal”, pero los tres modelos predijeron “Normal”. Esto muestra la debilidad del modelo, especialmente en un contexto donde la predicción puede ser determinante en la salud de una persona.

