



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN
EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

**Aplicación de Machine Learning para la
Predicción de la Mortalidad en Cirugía Cardíaca**

Autor:

D. Telmo Miguel Medina

Tutores:

Dr. Dña. Isabel de la Torre Díez

Cotutores:

Dr. Dña. Susel Góngora Alonso

Dr. Dña M^a Lourdes del Río Solá

Junio 2025

TÍTULO: **Aplicación de Machine Learning para la Predicción de la Mortalidad en Cirugía Cardíaca**

AUTOR: **D. Telmo Miguel Medina**

TUTOR: **Dr. Dña. Isabel de la Torre Díez**

COTUTOR: **Dr. Dña. Susel Góngora Alonso**
 Dr. Dña. M^a Lourdes del Río Solá (HCUV)

DEPARTAMENTO:

TRIBUNAL

PRESIDENTE: D.

VOCAL: D.

SECRETARIO D.

FECHA: Junio 2025

CALIFICACIÓN:

RESUMEN

El presente TFM tiene el objetivo de desarrollar una herramienta predictiva de la mortalidad a 30 días tras una cirugía cardíaca, utilizando algoritmos de ML aplicados a datos clínicos reales.

El proyecto ha seguido una metodología rigurosa que abarca el análisis exploratorio de los datos, la selección de variables clínicas relevantes, la evaluación comparativa de múltiples modelos y la implementación del mejor modelo (*XGBoost*) en una aplicación web funcional.

Los resultados obtenidos muestran un alto rendimiento predictivo para un dataset clínico real de 325 pacientes sometidos a cirugía cardíaca, destacando métricas como *AUC-ROC* (0,964), *recall* (0,9) y una buena calibración del modelo (*Brier Score* = 0,082).

Además se ha prestado especial atención a la interpretabilidad del modelo y su posible integración futura en entornos clínicos. Este trabajo demuestra el valor del uso del ML para apoyar decisiones médicas complejas y plantea una base sólida para su futura validación y mejora.

PALABRAS CLAVE

Predicción mortalidad, cirugía cardíaca, ML, *XGBoost*, modelo clínico, análisis de datos clínicos

ABSTRACT

This TFM Thesis aims to develop a predictive tool for 30-day mortality following cardiac surgery, using machine learning algorithms applied to real clinical data.

The project followed a rigorous methodology encompassing exploratory data analysis, selection of relevant clinical variables, comparative evaluation of multiple models, and the deployment of the best-performing model (*XGBoost*) in a functional web application.

The results demonstrate high predictive performance on a clinical dataset of 325 patients who underwent cardiac surgery, with strong metrics such as *AUC-ROC* (0.964), *recall* (0.9), and strong probability calibration (*Brier Score* = 0.082).

Special attention has been given to the interpretability of the model and its potential integration into clinical settings. This work highlights the value of machine learning in supporting complex medical decisions and provides a solid foundation for future validation and improvement.

KEYWORDS

Mortality prediction, cardiac surgery, machine learning, XGBoost, clinical model, clinical data analysis

ACRÓNIMOS

30DM - 30-day mortality – Mortalidad a 30 días (tras cirugía cardiovascular)
180DM - 30-day mortality – Mortalidad a 180 días (tras cirugía cardiovascular)
CHD – Congenital Heart Disease
ROC – Receiver Operating Characteristic – Tasa de verdaderos positivos frente tasa de falsos positivos
AUC-ROC / AROC – Area Under Curve ROC - Área bajo la curva ROC
ML – Machine Learning – Aprendizaje Automático
SVM – Support Vector Machine - Máquina de Soporte Vectorial
RF – Random Forest – Bosque Aleatorio
DT – Decision Tree – Árbol de Decisión
LR – Logistic Regression – Regresión Logística
ANN – Artificial Neuronal Network - Red Neuronal Artificial
DRF – Distributed Random Forest – Bosque Aleatorio Distribuido
XRT – Extremely Randomized Trees – Árboles Extremadamente Aleatorizados
GLM – Generalized Linear Model – Modelo Lineal Generalizado
EuroSCORE – European System for Cardiac Operative Risk Evaluation - Sistema de predicción de mortalidad para pacientes de cirugía cardíaca
TIMI – Thrombolysis in Myocardial Infarction – Sistema de puntuación para evaluar riesgo en infarto de miocardio
GRACE – Global Registry of Acute Coronary Events – Registro Global de Eventos Coronarios Agudos
MSE – Mean Squared Error – Error Cuadrático Medio
RMSE - Root Mean Squared Error – Raíz del Error Cuadrático Medio
F1-score – Métrica que combina precisión y recall en un promedio armonizado
PPV – Positive Predictive Value – Valor Predictivo Positivo
NPV – Negative Predictive Value – Valor Predictivo Negativo
DCA – Decision Curve Analysis – Análisis de Curva de Decisión
SHAP – Shapley Additive exPlanations – Método de interpretación de modelos de ML
LIME – Local Interpretable Model-agnostic Explanations – Técnica de interpretación de modelos de caja negra
HCUV – Hospital Clínico Universitario de Valladolid
EDA – Exploratory Data Analysis – Análisis Exploratorio de Datos

ÍNDICE

1.	INTRODUCCIÓN.....	9
1.1.	ANTECEDENTES	9
1.2.	OBJETIVOS	10
1.3.	ALCANCE	10
1.4.	METODOLOGÍA.....	11
1.4.1.	ESTADO DEL ARTE.....	11
1.4.2.	ANÁLISIS DE LOS DATOS	11
1.4.3.	DESARROLLO DEL MODELO	11
1.4.4.	DESARROLLO DE LA APLICACIÓN.....	12
1.4.5.	VALIDACIÓN	12
1.5.	INDICES CLÍNICOS	12
1.5.1.	INTRODUCCIÓN.....	12
1.5.2.	CONCEPTOS PREVIOS.....	13
1.5.3.	INDICES EMPLEADOS.....	14
2.	ESTADO DEL ARTE	20
2.1.	INTRODUCCIÓN	20
2.2.	METODOLOGÍA.....	20
2.2.1.	PROTOCOLO Y REGISTRO	20
2.2.2.	CRITERIOS DE ELIGIBILIDAD.....	20
2.2.3.	FUENTES DE INFORMACIÓN.....	20
2.2.4.	BÚSQUEDA.....	21
2.2.5.	SELECCIÓN DE LOS ESTUDIOS.....	21
2.2.6.	PROCESO DE EXTRACCIÓN DE DATOS	21
2.2.7.	LISTA DE DATOS.....	21
2.3.	RESULTADOS	22
2.3.1.	SELECCIÓN DE ESTUDIOS.....	22
2.3.2.	CARACTERÍSTICAS DE LOS ESTUDIOS.....	23
2.3.3.	RIESGO DE SESGO EN LOS ESTUDIOS.....	23
2.3.4.	RESULTADOS DE LOS ESTUDIOS INDIVIDUALES.....	24
2.3.5.	SÍNTESIS DE LOS RESULTADOS.....	24
2.3.6.	RIESGO DE SESGO EN LOS ESTUDIOS INDIVIDUALES.....	26
2.3.7.	MEDIDAS DE RESUMEN.....	27
2.3.8.	SÍNTESIS DE RESULTADOS.....	28
2.3.9.	RIESGO DE SESGO ENTRE LOS ESTUDIOS.....	28
2.4.	DISCUSIÓN	29

2.4.1.	RESUMEN DE LA EVIDENCIA.....	29
2.4.2.	LIMITACIONES.....	31
2.4.3.	CONCLUSIONES.....	32
3.	ANÁLISIS DE DATOS.....	33
3.1.	METODOLOGÍA Y CRITERIOS.....	33
3.1.1.	CRIBADO ESPECIALISTA	34
3.1.2.	CARGA Y PREPROCESAMIENTO.....	34
3.1.3.	ANÁLISIS ESTADÍSTICO VARIABLES NUMÉRICAS.....	34
3.1.4.	ANÁLISIS ESTADÍSTICO VARIABLES CATEGÓRICAS.....	38
3.2.	DATASET	39
3.3.	CRIBADO ESPECILISTA.....	40
3.4.	CARGA Y PREPROCESAMIENTO	43
3.4.1.	ANÁLISIS DE FALTANTES.....	43
3.4.2.	ANÁLISIS DE DUPLICADOS.....	44
3.4.3.	ERRORES DE CODIFICACIÓN.....	44
3.5.	ANÁLISIS ESTADÍSTICO VARIABLES NUMÉRICAS	44
3.5.1.	ANÁLISIS EXPLORATORIO (EDA).....	44
3.5.2.	TEST DE NORMALIDAD.....	48
3.5.3.	ANÁLISIS DE HOMOCEDASTICIDAD.....	52
3.6.	ANÁLISIS ESTADÍSTICO VARIABLES CATEGÓRICAS	55
3.6.1.	ANÁLISIS EXPLORATORIO (EDA).....	55
3.6.2.	ANÁLISIS DE HOMOCEDASTICIDAD.....	61
3.7.	ANÁLISIS ESTADÍSTICO VARIABLE OBJETIVO.....	62
3.7.1.	ANÁLISIS EXPLORATORIO (EDA).....	62
3.8.	SELECCIÓN DE CARACTERÍSTICAS	63
3.8.1.	PRE-SELECCIÓN CON RANDOM FOREST.....	63
3.8.2.	SELECCIÓN DEFINITIVA.....	64
4.	SELECCIÓN DEL MODELO	65
4.1.	INTRODUCCIÓN.....	65
4.2.	CONCEPOS PREVIOS.....	65
4.2.1.	SMOTE.....	65
4.2.2.	MODELOS.....	65
4.2.3.	VOTING CLASSIFIER.....	68
4.3.	EQUILIBRADO DE LAS CLASES.....	68
4.4.	PRE-SELECCIÓN DEL MODELO	68
4.4.1.	SVM.....	69
4.4.2.	XGBOOST.....	69
4.4.3.	RANDOM FOREST	70
4.4.4.	CATBOOST.....	70

4.4.5.	KNN.....	71
4.4.6.	MLP.....	71
4.4.7.	XGBOOST + CATBOOST (Voting Classifier)	72
4.4.8.	RANDOM FOREST + CATBOOST (Voting Classifier)	72
4.4.9.	COMPARATIVA	73
4.4.10.	SELECCIÓN	76
4.5.	OPTIMIZACIÓN MODELO SELECCIONADO	77
5.	DESARROLLO APLICACIÓN	78
5.1.	INTRODUCCIÓN	78
5.2.	REQUERIMIENTOS	78
5.2.1.	REQUERIMIENTOS FUNCIONALES	78
5.2.2.	REQUERIMIENTOS NO FUNCIONALES.....	79
5.2.3.	REQUERIMIENTOS ÉTICO-LEGALES	81
5.3.	TECNOLOGÍAS	81
5.3.1.	PYTHON	81
5.3.2.	STREAMLIT.....	82
5.3.3.	SCIKIT-LEARN.....	82
5.4.	MÓDULOS Y ESTRUCTURA.....	83
5.5.	DESPLIEGUE	83
5.6.	REPOSITORIO	83
5.7.	MANUAL DE USUARIO	84
5.8.	LICENCIA.....	85
5.9.	ADVERTENCIAS Y CONDICIONES DE USO	85
6.	CONCLUSIONES.....	86
6.1.	CONCLUSIONES	86
6.2.	LINEAS FUTURAS	87
7.	REFERENCIAS	88
7.1.	REFERENCIAS BIBLIOGRÁFICAS.....	88
7.2.	ÍNDICE DE ILUSTRACIONES	92
7.3.	ÍNDICE DE TABLAS	94

1. INTRODUCCIÓN

1.1. ANTECEDENTES

La predicción temprana de la mortalidad en pacientes con enfermedades cardiovasculares es un área de creciente interés en la medicina clínica y la investigación en salud. Las enfermedades del corazón siguen siendo la principal causa de muerte a nivel mundial, según datos de la Organización Mundial de la Salud [1], y representan una carga significativa tanto para los sistemas de salud como para las familias afectadas.

En este contexto, la estimación precisa del riesgo de muerte en un horizonte temporal breve, como los primeros 30 días tras el ingreso hospitalario, resulta esencial para orientar decisiones clínicas, optimizar recursos y mejorar los resultados en salud. En las últimas décadas, diversos modelos de predicción han sido desarrollados utilizando técnicas estadísticas clásicas como la regresión logística, particularmente en cohortes de pacientes con insuficiencia cardíaca, infarto agudo de miocardio o enfermedad coronaria. Estos modelos, como el *TIMI* o el *GRACE* [2]/[3], han sido ampliamente utilizados en la práctica clínica. Sin embargo, presentan limitaciones en cuanto a su capacidad de generalización, ya que suelen estar entrenados sobre poblaciones específicas, y su rendimiento puede verse afectado en contextos con características distintas.

Además, tienden a asumir relaciones lineales entre variables, lo cual no siempre refleja la complejidad de los procesos fisiopatológicos subyacentes. Con el avance de la ciencia de datos y el ML, ha sido posible desarrollar modelos más flexibles y potentes que capturan relaciones no lineales y patrones complejos en los datos clínicos. Modelos como los *DT*, *RF* [4], *XGBoost* [5], y redes neuronales [3] han mostrado mejoras en precisión predictiva en comparación con los modelos tradicionales. En estudios recientes, el uso de datos clínicos electrónicos combinados con algoritmos de machine learning ha permitido obtener modelos que predicen con mayor exactitud la mortalidad a corto plazo en pacientes con enfermedades cardiovasculares, al integrar múltiples variables clínicas, demográficas y de laboratorio.

A pesar de estos avances, la implementación de estos modelos en la práctica clínica sigue siendo limitada, en parte debido a la falta de transparencia (especialmente en modelos tipo "caja negra") y a problemas de generalización fuera de los entornos

donde fueron entrenados. Por ello, se hace necesario desarrollar e investigar modelos que no solo alcancen un alto rendimiento predictivo, sino que también sean interpretables y evaluados rigurosamente en cuanto a sesgo, representatividad de la muestra y validez externa.

En este trabajo se abordará esta necesidad mediante el desarrollo de un modelo de predicción de la mortalidad a 30 días en pacientes con problemas cardíacos, utilizando técnicas de ciencia de datos sobre una base clínica real, con énfasis en la validación cruzada, la interpretabilidad del modelo y la evaluación del sesgo potencial en los datos.

1.2. OBJETIVOS

El objetivo principal del proyecto está centrado en desarrollar una aplicación para la predicción de la mortalidad a 30 días en cirugía cardíaca utilizando herramientas de *ML*.

1.3. ALCANCE

Para el desarrollo de la aplicación se desarrollará un modelo predictivo basado en los datos de pacientes intervenidos en cirugía cardíaca, a los cuales de manera previa se les hará un estudio del estado del arte, un análisis estadístico completo, una selección de característica, selección y optimización del modelo predictivo y exportación e implementación del modelo en una aplicación web de simple uso.

1.4. METODOLOGÍA

Para el desarrollo de la aplicación se seguirá una metodología basada en cinco etapas bien definidas, que permiten abordar de manera ordenada y rigurosa el desarrollo requerido.



Ilustración 1. Metodología (Elaboración Propia)

1.4.1. ESTADO DEL ARTE

Revisión bibliográfica y contextualización del problema, dentro de esta etapa se realizará:

- Búsqueda y análisis de estudios previos sobre la 30DM en cirugía cardíaca
- Revisión de modelos de puntuación (*scoring*) existentes (*EuroSCORE*, *STS...*)
- Revisión de técnicas de ML aplicadas en el ámbito clínico

1.4.2. ANÁLISIS DE LOS DATOS

El objetivo de esta fase es la por un lado disponer de un conjunto de datos (*dataset*) clínicos que relacionen la 30DM en cirugía cardíaca con variables fisiológicas medibles y por otro lado la de realizar un entendimiento profundo de los datos (relación entre variables, correlaciones, variables predictoras, visualización mediante gráficos...)

Se empleará como herramienta principal Python con librerías de análisis avanzado de datos como *numpy*, *pandas*, *scipy*, o *scikit-learn*.

1.4.3. DESARROLLO DEL MODELO

En el desarrollo del modelo se trabajará en identificar que modelo de ML y con qué hiper-parámetros se comporta mejor para el *dataset* utilizado, para ello se subdividirá en cuatro fases:

- 1º - Separación de los datos de prueba (*test*) y entrenamiento (*train*)

2º - Selección y comparación de diferentes algoritmos de *ML* mediante validación cruzada utilizando *K-folds*.

3º - Ajuste de los hiper-parámetros y evaluación del desempeño de cada modelo (métricas como precisión, recall, F1-score, ROC-AUC...)

4º - Interpretación del modelo (importancia de variables, SHAP...)

1.4.4. DESARROLLO DE LA APLICACIÓN

Una vez definido el modelo de aprendizaje y las variables necesarias para su correcto funcionamiento, se desarrollará una aplicación web con los siguientes hitos dentro del desarrollo:

- Diseño de una interfaz sencilla e intuitiva que permita la entrada de los datos clínicos por parte del usuario
- Integración del modelo entrenado en la aplicación para realizar predicciones en tiempo real
- Visualización de resultado de manera comprensible (probabilidad de mortalidad, aplicabilidad del modelo...)
- Validación funcional de la aplicación

1.4.5. VALIDACIÓN

Como punto final del desarrollo se somete a la aplicación a un proceso de validación donde se realizará:

- Pruebas con datos simulados o reales para comprobar la robustez del sistema
- Recogida del *feedback* de los usuarios (médicos, investigadores...)
- Análisis de casos de error o predicciones erróneas del modelo

1.5. INDICES CLÍNICOS

1.5.1. INTRODUCCIÓN

Los índices clínicos son indicadores que evalúan el estado de salud de un individuo o población, permitiendo medir el nivel de salud y sus cambios

El presente apartado hace una revisión de los principales índices utilizados en el estudio de predicción de la mortalidad, desde su definición más teórica hasta la determinación del valor propio del índice.

1.5.2. CONCEPTOS PREVIOS

1.5.2.1 FRAGILIDAD

La fragilidad [6] se define como un estado clínico relacionado con el envejecimiento, caracterizado por una disminución de la reserva funcional y fisiológica en diversos órganos y sistemas. Esta situación reduce la capacidad del individuo para enfrentarse a factores estresantes, tanto crónicos como agudos, lo que lo vuelve más vulnerable.

Por este motivo, se asocia con un mayor riesgo de eventos adversos en la salud, como caídas, recuperación más lenta o con secuelas tras enfermedades, cirugías o efectos adversos de medicamentos, así como con un incremento en las hospitalizaciones, la institucionalización y la mortalidad. Resulta especialmente significativa su relación con el avance hacia la discapacidad o dependencia.

La fragilidad es considerada un síndrome geriátrico porque tiene múltiples causas, se identifica por ciertas características clínicas específicas y se aborda con tratamientos multifactoriales. Esto implica también en un aumento de los costes generales del sistema sanitario.

La fragilidad y la funcionalidad están estrechamente vinculadas, ya que la fragilidad suele conducir a un deterioro de la funcionalidad. Ambas comparten estrategias de intervención enfocadas en lograr un envejecimiento activo y sin discapacidad. En términos conceptuales, la fragilidad es considerada una etapa previa a la discapacidad. No obstante, en la práctica también se incluye dentro del grupo de personas frágiles a aquellas con dificultades iniciales en actividades instrumentales o avanzadas de la vida diaria, las cuales muchas veces solo se detectan mediante una evaluación específica y detallada.

1.5.2.2 EuroSCORE

El EuroSCORE (*European System for Cardiac Operative Risk Evaluation*) [7] es un índice clínico ampliamente utilizado para predecir el riesgo de mortalidad quirúrgica en pacientes sometidos a cirugía cardíaca. Es un modelo de predicción de riesgo desarrollado a partir de datos de pacientes europeos sometidos a cirugía cardíaca. Estima la probabilidad de mortalidad intraoperatoria o en los primeros 30 días postoperatorios.

El modelo incluye múltiples variables agrupadas en diferentes categorías como los datos del paciente (edad, sexo, creatinina, enfermedad pulmonar, diabetes...), estado clínico (endocarditis activa, reanimación previa, disfunción en órganos...) y datos de la cirugía (urgencia del procedimiento, tipo de cirugía, reoperación...)

1.5.3. INDICES EMPLEADOS

1.5.3.1 INDICE DE BARTHEL

El índice de *Barthel* [8] es instrumento utilizado por los profesionales de la salud para la valoración funcional de un paciente y el seguimiento de la evolución de este.

El índice funciona con una escala que mide el nivel de dependencia/independencia del paciente en la realización de actividades cotidianas, entendidas como básicas, midiendo de una forma cuantitativa el grado de dependencia.

ACTIVIDAD	DESCRIPCIÓN	PUNTUACIÓN
Comer	Incapaz	0
	Necesita ayuda para cortar, extender mantequilla, usar condimentos, etc.	5
	Independiente (la comida está al alcance de la mano)	10
Trasladarse entre la silla y la cama	Incapaz, no se mantiene sentado	0
	Necesita ayuda importante (una persona entrenada o dos personas), puede estar sentado	5
	Necesita algo de ayuda (una pequeña ayuda física o ayuda verbal)	10
	Independiente	15
Aseo personal	Necesita ayuda con el aseo personal	0

	Independiente para lavarse la cara, las manos y los dientes, peinarse y afeitarse	5
Uso del retrete	Dependiente	0
	Necesita alguna ayuda, pero puede hacer algo solo	5
	Independiente (entrar y salir, limpiarse y vestirse)	10
Bañarse y/o Ducharse	Inmóvil	0
	Independiente para bañarse y/o ducharse	5
Desplazarse	Inmóvil	0
	Independiente en silla de ruedas en 50m	5
	Anda con pequeña ayuda de una persona (física o verbal)	10
	Independiente al menos 50m, con cualquier tipo de muleta, excepto andador	15
Subir y bajar escaleras	Incapaz	0
	Necesita ayuda física o verbal, puede llevar cualquier tipo de muleta	5
	Independiente para subir y bajar	10
Vestirse y desvestirse	Dependiente	0
	Necesita ayuda, pero puede hacer la mitad aproximadamente, sin ayuda	5
	Independiente, incluyendo botones, cremalleras, cordones, etc.	10
Control de heces	Incontinente (o necesita que le suministren enema)	0
	Accidente excepcional (uno/semana)	5
	Continente	10
Control de orina	Incontinente, o sondado incapaz de cambiarse la bolsa	0
	Accidente excepcional (uno/semana)	5
	Continente, durante al menos 7 días	10
Total	0-100 puntos (0-90 si usa silla de ruedas)	-

Tabla 1. Puntuaciones del Índice de Barthel (Elaboración Propia)

En función de la puntuación obtenida, se clasifica la dependencia/independencia del paciente según el criterio de *Shal* [9] de la siguiente manera

PUNTUACIÓN	RANGO
0-20	Dependencia Total
21-60	Dependencia Severa
61-90	Dependencia Moderada
91-99	Dependencia Escasa
100	Independencia

Tabla 2. Clasificación Índice de Barthel

1.5.3.2 INDICE DE EDMONTON

El índice *Edmonton* [10] mide la fragilidad en adultos mayores, utilizando nueve dominios a través de un sistema de puntuación de 0 a 17 puntos según la siguiente tabla:

ÁMBITO	PREGUNTA	PUNTUACIÓN
Cognición	¿Qué día es hoy? ¿En qué estación estamos?	0 = ambas correctas 1 = una correcta 2 = ninguna correcta
Estado general de salud	¿Cómo describiría su salud en general?	0 = excelente/buena 1 = regular 2 = mala
Independencia funcional	¿Necesita ayuda para hacer compras?	0 = independiente 1 = necesita ayuda 2 = dependiente
Apoyo social	¿Tiene alguien que lo apoye regularmente?	0 = siempre 1 = algunas veces 2 = nunca
Uso de medicamentos	¿Cuántos medicamentos toma al día?	0 = 0 a 4 1 = 5 o más
Nutrición	¿Ha perdido peso sin proponérselo en los últimos 6 meses?	0 = no 1 = si
Estado de ánimo	¿Se siente frecuentemente triste o deprimido?	0 = no 1 = si
Continencia	¿Ha tenido pérdidas involuntarias de orina o heces en el último mes?	0 = no 1 = si

Funcional	¿Tiene dificultad para caminar, usar escaleras o hacer tareas diarias?	0 = no 1 = si
Rendimiento físico	¿Puede levantarse de una silla sin usar los brazos?	0 = <10 seg 1 = 10 a 20 seg 2 = > 20 seg

Tabla 3. Puntuación Índice de Edmonton (Elaboración Propia)

En función de la puntuación obtenida, se clasifica la fragilidad según la siguiente tabla:

PUNTUACIÓN	RANGO
0-4	No frágil / Pre-frágil
5-6	Fragilidad leve
7-8	Fragilidad moderada
>9	Fragilidad severa

Tabla 4. Clasificación Índice de Edmonton (Elaboración Propia)

1.5.3.3 INDICE DE FRAIL

El índice de *Frail* [11] es una herramienta clínica para la determinación de la fragilidad de un paciente. Se basa en un cuestionario de cinco preguntas cada una relativa a un dominio diferente:

- Fatigabilidad
- Resistencia
- Deambulación
- Comorbilidad
- Pérdida de peso

ÁMBITO	PREGUNTA	RESPUESTA	PUNTUACIÓN
Fatigabilidad	¿Qué parte del tiempo durante las últimas 4 semanas se sintió cansado/a?	Todo el tiempo	1
		La mayoría de las veces	1
		Parte del tiempo	0
		Un poco el tiempo	0
		En ningún momento	0

Resistencia	¿Tiene alguna dificultad para subir 10 escalones sin descansar por usted mismo/a, sin ningún tipo de ayuda?	Si	1
		No	0
Deambulación	Por sí mismo/a y sin el uso de ayudas, ¿tiene alguna dificultad para caminar varios cientos de metros?	Si	1
		No	0
Comorbilidad	Tiene o ha tenido: <ul style="list-style-type: none"> • Hipertensión • Diabetes • Cáncer • EPOC • Ataques cardíacos • Insuficiencia cardíaca congestiva • Angina de pecho • Asma • Artritis • Ictus • Enfermedad renal 	Si (de 0 a 4)	0
		Si (de 5 a 11)	1
Pérdida de peso	¿Peso actual? ¿Peso hace 1 año? $\frac{Peso_{hace\ 1\ año} - Peso_{actual}}{Peso_{hace\ 1\ año}} * 100$	>5	1
		<5	0

Tabla 5. Puntuación Índice de Frail

La puntuación oscila entre 0 y 5, si el resultado es ≥ 1 se considera probabilidad de fragilidad.

1.5.3.4 INDICE DE KATZ

El índice de *Katz* [12] es una métrica utilizada para evaluar la independencia de una persona enferma en la realización de actividades cotidianas básicas. Este índice analiza seis puntos relacionados con actividades diarias y evalúa la dependencia o independencia de cada uno de ellos.

ACTIVIDAD	DESCRIPCIÓN	PUNTUACIÓN
Baño	Independiente – Se baña solo o precisa de ayuda para lavar alguna zona, como espalda, o una extremidad con minusvalía	0
	Dependiente – Precisa de ayuda para lavas más de una zona para salir o entrar de la bañera, o no puede bañarse solo	1
Vestido	Independiente – Saca ropa de cajones y armarios, se la pone y abrocha. Se excluye el acto de atarse los zapatos	0
	Dependiente – No se viste por sí mismo, o permanece parcialmente desvestido	1
Uso del WC	Independiente – Va al WC solo, se arregla la ropa y se limpia	0
	Dependiente – Necesita ayuda para ir al WC	1
Movilidad	Independiente – Se levanta y acuesta en la cama por sí mismo, y puede levantarse de una silla por sí mismo	0
	Dependiente – Precisa ayuda para levantarse y acostarse en la cama o silla. No realiza uno o más desplazamientos	1
Continencia	Independiente – Control completo de micción y defecación	0
	Dependiente – Incontinencia parcial o total de la micción o defecación	1
Alimentación	Independiente – Lleva el alimento a la boca desde el plato o equivalente	0
	Dependiente – Precisa de ayuda para comer	1

Tabla 6. Puntuación Índice de Katz

2. ESTADO DEL ARTE

2.1. INTRODUCCIÓN

El presente apartado tiene el objetivo de dar a conocer el estado del arte de la estimación de la 30DM en pacientes con problema cardíacos, aplicando métodos de predicción avanzados de ML.

Se pretende conocer cuáles son los principales métodos utilizados en la predicción de la 30DM, conocer el rendimiento que ofrece cada uno de estos sistemas y definir las variables críticas que más influyen en la estimación del 30DM.

2.2. METODOLOGÍA

Para la realización del estudio se seguirá la metodología *PRISMA* (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*) [13][14][15]

2.2.1. PROTOCOLO Y REGISTRO

No existe un protocolo claro ni definido en este tipo de estudios, ya que dependen, en gran medida, de las pruebas médicas a las que haya podido ser sometido cada paciente, lo que genera un amplio abanico en cuanto a las pruebas, análisis y estudios a las que ha estado sujeto cada paciente.

2.2.2. CRITERIOS DE ELIGIBILIDAD

En las publicaciones analizadas se buscarán aquellas en las que los datos analizados:

- Correspondan a registros de entre 300 y 1.000 sujetos
- Se haya aplicado, como mínimo, una técnica/modelo de ML
- Los resultados estén analizados con más de una métrica de evaluación
- Año de publicación sea posterior a 2020 (inclusive)

2.2.3. FUENTES DE INFORMACIÓN

Las plataforma utilizadas para la búsqueda serán las siguientes:

- *Sci-Space*

- *Semantic Scholar*
- *Paper Digest*
- *SciSummary*

2.2.4. BÚSQUEDA

La estrategia de búsqueda seguida será la de analizar en todas las Fuentes de Información indicada todas las publicaciones que contengan:

- El Título/Abstract contenga las siguientes palabras clave: “30”, “day”, “mortality”, “cardiac”, “machine”, “learning”
- 2020 <= Año Publicación <= 2025
- Tipo de publicación: Artículo, Capítulo de Libro y Libro

2.2.5. SELECCIÓN DE LOS ESTUDIOS

Para todo el conjunto de datos, se realizará un cribado para las publicaciones que no cumplan las siguientes condiciones:

- N° de autores + Revisores > 3
- Tamaño del dataset > 300 registros
- Utilización de herramientas de alto nivel (*MATLAB, Python, R...*)
- Datasets revisados, con origen definido, con referencias clínicas...

De igual modo, se realizará un cribado con las publicaciones duplicadas en las diferentes Fuentes de Información

2.2.6. PROCESO DE EXTRACCIÓN DE DATOS

Para la extracción de los datos de las publicaciones se realizará de manera independiente y se utilizará un formulario (EXCEL) en el que se irán anotando los datos indicados en la “Lista de Datos”. Los datos serán posteriormente revisados por una tercera persona

2.2.7. LISTA DE DATOS

Los datos que van a formar parte del Formulario para la Extracción de Datos y que se utilizarán para la Selección de los Estudios serán:

- Autor/es
- Fecha
- Dataset (fuente y tamaño)
- Herramientas utilizadas para el análisis
- Técnicas/modelos empleados
- Métricas de Evaluación
- Resultados obtenidos

2.3. RESULTADOS

2.3.1. SELECCIÓN DE ESTUDIOS

Nº de registros Identificados: 30

Nº de Registros Duplicados: 7

Nº de Registros excluidos tras primer filtro (por nombres de título o nº de autores): 0

Nº de Registros excluido por el criterio selección: 5

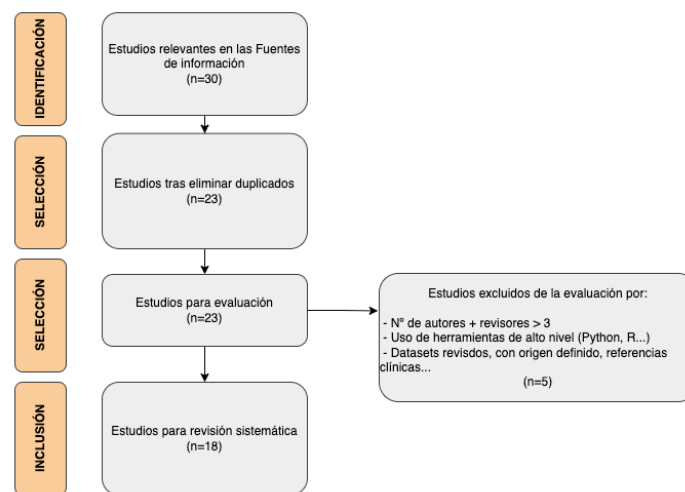


Ilustración 2. Selección de Estudios (Elaboración Propia)

2.3.2. CARACTERÍSTICAS DE LOS ESTUDIOS

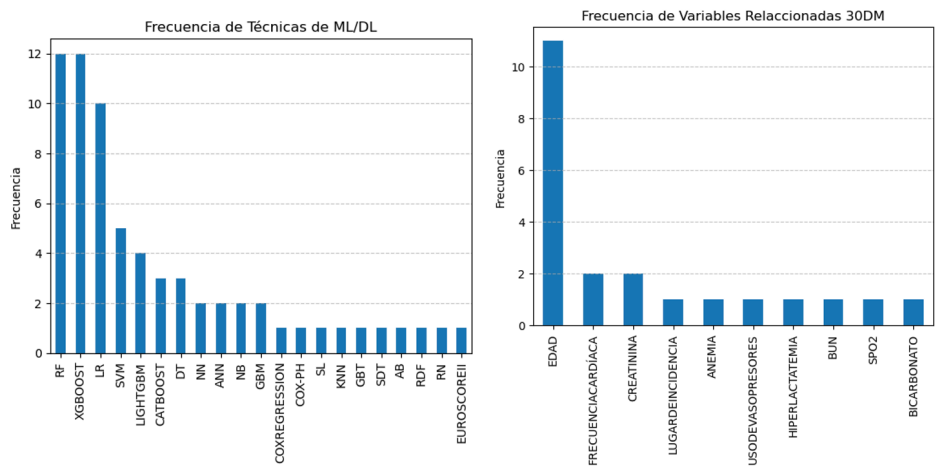


Ilustración 3. Características de los estudios (Elaboración Propia)

2.3.3. RIESGO DE SESGO EN LOS ESTUDIOS

Sobre los estudios analizados, se ha realizado una evaluación del sesgo de los resultados utilizando la metodología *QUADAS-2* [16], una herramienta desarrollada para valorar la calidad de los estudios de pruebas diagnósticas incluidos en una revisión sistemática.

Para esta evaluación se plantean las siguientes cuestiones relativas a cada estudio analizado:

- ¿La muestra es representativa? (Si/No)
- ¿Las variables están bien definidas? (Si/No)
- ¿Las métricas utilizadas son adecuadas? (Si/No)
- ¿Existe posible conflicto de interés? (Si/No)

En función de las respuestas obtenidas a las cuestiones anteriores definiremos el nivel de sesgo del estudio según la siguiente tabla:

¿La muestra es representativa?	¿Existe posible conflicto de interés?	Evaluación Sesgo Independiente
¿Las variables están bien definidas?		
¿Las métricas utilizadas son adecuadas?		

3 Síes	No	Bajo
0 Síes	Sí	Alto
1-2 Síes	Sí/No	Moderado

Tabla 7 Riesgo de sesgo en los estudio (Elaboración Propia)

Se eliminarán del análisis aquellos estudios que tengan un Sesgo Independiente Alto o Moderado, dejando únicamente los estudio con un Sesgo Independiente Bajo.

2.3.4. RESULTADOS DE LOS ESTUDIOS INDIVIDUALES

2.3.5. SÍNTESIS DE LOS RESULTADOS

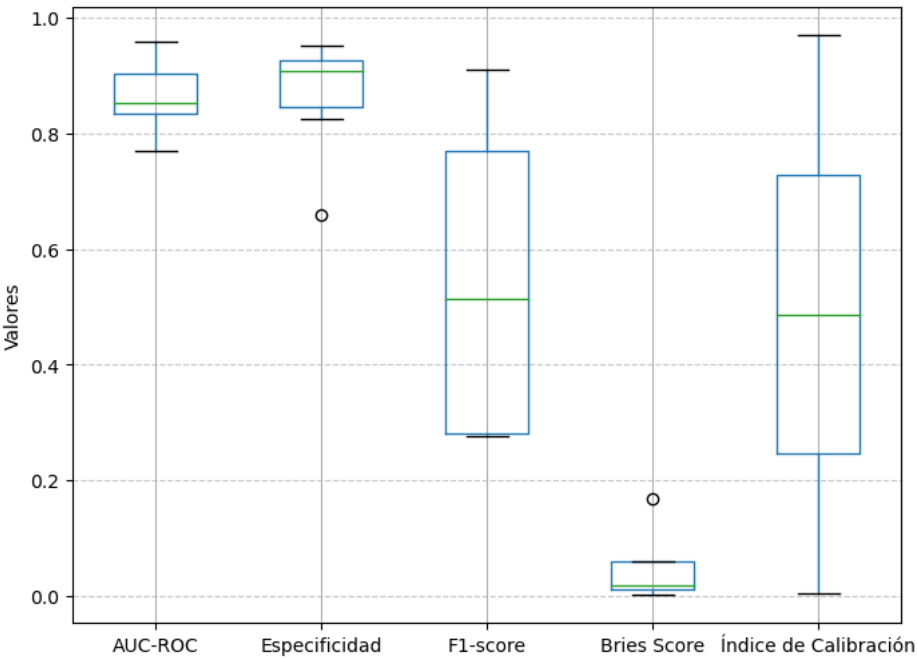


Ilustración 4. Boxplot métricas en los estudios individuales (Elaboración Propia)

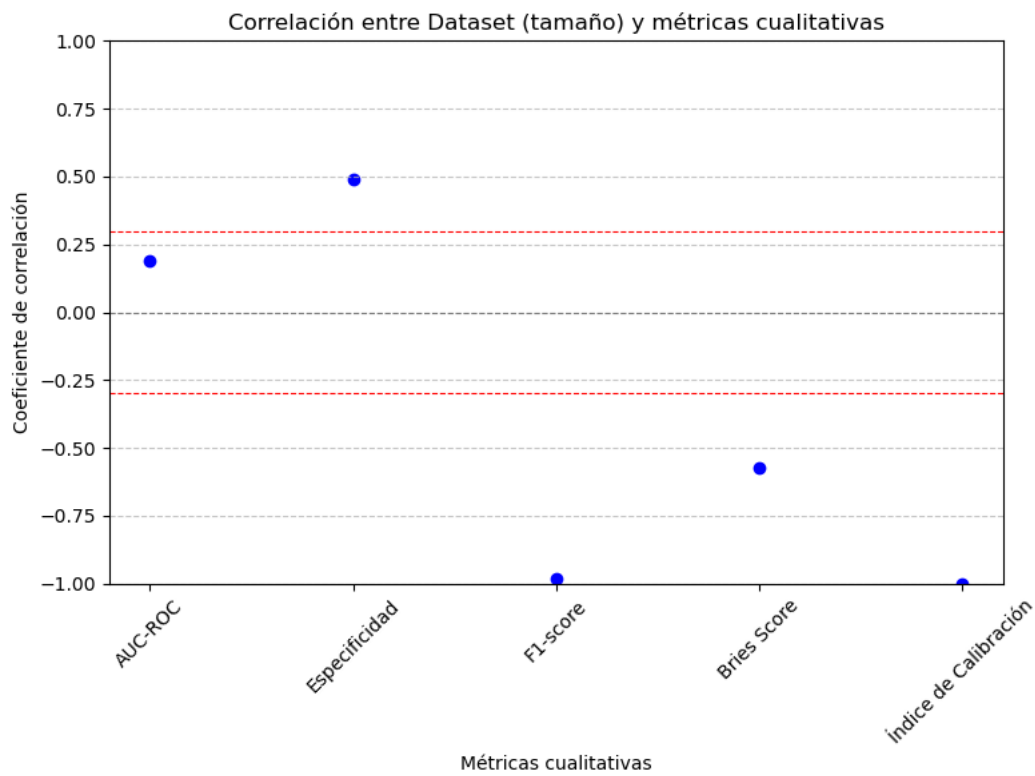


Ilustración 5. Correlación métricas de los estudios individuales (Elaboración Propia)

2.3.6. RIESGO DE SESGO EN LOS ESTUDIOS

INDIVIDUALES

Se presentan a continuación los estudios excluidos por tener un Sesgo Independiente entre Moderado y Alto

ESTUDIO	QUADAS-2				
	¿La muestra es representativa? (Si/No)	¿Las variables están bien definidas? (Si/No)	¿Las métricas utilizadas son adecuadas? (Si/No)	¿Existe posible conflicto de interés? (Sí/No)	Evaluación Sesgo Independiente
<i>ICURE: Intensive care unit (ICU) risk evaluation for 30-day mortality. Developing and evaluating a multivariable machine learning prediction model for patients admitted to the general [15]</i>	Si	Si	Si	Si	Moderado
<i>Machine learning-based prediction of 30-day major adverse cardiac and cerebrovascular events in non-cardiac surgery patients [17]</i>	Si	Si	Si	Si	Moderado
<i>Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using Xgboost [18]</i>	Si	Si	Si	Si	Moderado

Tabla 8. Riesgo de sesgo en los estudios individuales (Elaboración Propia)

2.3.7. MEDIDAS DE RESUMEN

Para evaluar el rendimiento de los modelos se aplicarán diferentes métricas de desempeño teniendo las siguientes consideraciones:

Métricas de discriminación

- AUC-ROC – Valores superiores a 0,8 considerados como buen rendimiento en la capacidad discriminativa del modelo
- Sensibilidad y Especificidad – Se analizarán ambos valores para evaluar la tasa de verdaderos positivos y la capacidad del modelo para evitar falsos negativos

Métricas de rendimiento del modelo

- Precision, recall – Clave en la evaluación de modelos de clasificación
- F1-score – Promedio ponderado entre precisión y recall, útil para evaluar el equilibrio entre ambos
- MSE (Error Cuadrático Medio) y RMSE (Raíz del Error Cuadrático Medio)
– Utilizado en modelos de regresión para medir el error de predicción

Métricas de calibración

- Índice de calibración – Comparación de la probabilidad estimada frente a los valores reales
- Brier Score - Evalúa la cantidad de las probabilidades estimadas por el modelo (valores más bajos indican mejor calibración)

2.3.8. SÍNTESIS DE RESULTADOS

Síntesis Cuantitativas

En los estudios que reporten resultados comparables (ROC-AUC, sensibilidad, especificidad, F1-score...) se realizará un análisis estadístico conjunto, calculando el rendimiento completo de los modelos de ML comparados (Modelo de efectos aleatorios). Además, se utilizará un estadístico I^2 para evaluar la heterogeneidad entre estudios, considerando valores superiores al 50% como indicativos de alta variabilidad.

Síntesis Cualitativa

Para los estudios con diferencias metodológicas significativas (uso de diferentes modelos de ML, fuentes de datos, métricas no estandarizadas...) se realizará una síntesis narrativa para identificar patrones generales en la aplicación de modelos predictivos, agrupando los estudios según:

- Tipo de modelo de ML utilizado
- Tamaño del conjunto de datos
- Fuente de los datos y variables empleadas

2.3.9. RIESGO DE SESGO ENTRE LOS ESTUDIOS

Para evitar el riesgo del sesgo entre los estudios, incluidos en la revisión, se analizarán diferentes fuentes de distorsión de la evidencia, como diferencias en el diseño, calidad de los datos y enfoques analíticos utilizados en la predicción de la 30DM

Estrategias Evaluación del Sesgo:

- Diagrama de embudo y prueba de *Egger*: Para detectar sesgo de publicación mediante la distribución de los estudios y la asimetría en los resultados
- Índice de heterogeneidad (I^2): Para medir la variabilidad entre estudios, valores superiores al 50% indican alta heterogeneidad
- Evaluación cualitativa: Se revisarán aspectos como la fuente de datos, validación de modelos y reporte de resultados negativos.

2.4. DISCUSIÓN

2.4.1. RESUMEN DE LA EVIDENCIA

ID	ESTUDIO	Modelos utilizados	Métricas de Evaluación	Validación con dataset externo
1	<i>Prediction of 30-day mortality in heart failure patients with hypoxic hepatitis: Development and external validation of an interpretable machine learning model [19]</i>	<i>CatBoost, XGBoost, LightGBM, RF, SVM, DT, LR</i>	<i>AUC, Accuracy, Recall, F1-score, MCC, Brier-score</i>	Si
2	<i>A machine learning algorithm-based risk prediction score for in-hospital/30 day mortality after adult cardiac surgery [20]</i>	<i>XGBoost</i>	<i>AUC-ROC, F1-score, Brier-score, DCA</i>	Si
3	<i>ICURE: Intensive care unit (ICU) risk evaluation for 30-day mortality. Developing and evaluating a multivariable machine learning prediction model for patients admitted to the general [21]</i>	<i>XGBoost, RF, RN, LR</i>	<i>AUC-ROC, Sensibilidad, Especificidad, Valor Predictivo Negativo, Valor Predictivo Positivo, Calibración</i>	No
4	<i>Interpretable Machine Learning Approach for Predicting 30-Day Mortality of Critical Ill Patients with Pulmonary Embolism and Heart Failure: A Retrospective Study [22]</i>	<i>SVM, RF, XGBost, LR, CatBoost, LightGBM</i>	<i>AUC-ROC, DCA, NRI, IDI, Sensibilidad, Especificidad, PPV, NPV</i>	No
5	<i>Machine learning models to predict 30-day mortality for critical patients with myocardial infarction: a retrospective analysis from MIMIC-IV database [23]</i>	<i>XGBoost, RDF</i>	<i>AUC-ROC, Calibración (Hosmer-Lemeshow), Decision Curve Analysis (DCA), Sensibilidad</i>	No
6	<i>A machine learning algorithm-based risk prediction score for in-hospital/30-day mortality after adult cardiac surgery [5]</i>	<i>XGBoost</i>	<i>AUC-ROC, F1-score, DCA, Brier score</i>	No

7	<i>Machine learning algorithm for predicting 30-day mortality in patients receiving rapid response system activation: A retrospective nationwide cohort study [24]</i>	<i>LightGBM, XGBoost, RF, NN</i>	<i>AUC-ROC, Sensibilidad, Presicion, Especificidad</i>	No
8	<i>A machine learning approach to predicting 30-day mortality following paediatric cardiac surgery: findings from the Australia New Zealand Congenital Outcomes Registry for Surgery (ANZCORS) [25]</i>	<i>LR, GBT, ANN</i>	<i>AUC-ROC, Calibración, Sensibilidad, Especificidad,</i>	No
9	<i>Machine learning-based prediction of 30-day major adverse cardiac and cerebrovascular events in non-cardiac surgery patients [17]</i>	<i>GBM, LR, RF, AB, DT</i>	<i>AUC-ROC, AUPRC</i>	
10	<i>Late Mortality After Myocardial Injury in Critical Care Non-Cardiac Surgery Patients Using Machine Learning Analysis [26]</i>	<i>SDT, Cox Regression</i>	<i>Hazard Ratio, Intervalo de Confianza, AUC-ROC</i>	No
11	<i>Can machine learning improve mortality prediction following cardiac surgery? [27]</i>	<i>ANN, RF, NB, LR</i>		No
12	<i>Predicting 30-day mortality in severely injured elderly patients with trauma in Korea using machine learning algorithms: a retrospective study [28]</i>	<i>LR, DT, RF, XGBoost</i>	<i>AUC-ROC, Accuracy, Precision, Recall, Especificidad, F1-score, SHAP</i>	
13	<i>Prediction of 30-Day Mortality for ICU Patients with Sepsis-3 [29]</i>	<i>LightGBM, Catboost, RF, KNN, SVM, XGBoost, LR</i>	<i>ROC-AUC</i>	No
14	<i>Optimizing cardiovascular disease mortality prediction: a super learner approach in the tehran lipid and glucose study [30]</i>	<i>SL, GBM, SVM, Cox-PH</i>	<i>TD-AUC, C-index, Brier Score, Prediction Error</i>	No

15	<i>Comparison of machine learning techniques in prediction of mortality following cardiac surgery: analysis of over 220 000 patients from a large national database [31]</i>	<i>EuroSCORE II, LR, RF, NN, XGBoost, SVM</i>	<i>ROC-AUC, F1-score, ECE, Brier-score, DCA</i>	No
16	<i>Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using Xgboost [18]</i>	<i>XGBoost</i>	<i>AUC-ROC</i>	No
17	<i>Prediction of 30-day Mortality for Heart Failure Patients with Cardiogenic Shock Using Random Forest Hyperparameter Tuning [4]</i>	<i>RF</i>	<i>Accuracy, Kappa, AUC-ROC</i>	
18	<i>Prediction Model for 30-Day Mortality after Non-Cardiac Surgery Using Machine-Learning Techniques Based on Preoperative Evaluation of Electronic Medical Records [32]</i>	<i>XGBoost, RF, LR, NB</i>	<i>AUC-ROC, Sensibilidad, Especificidad, PPV, NPV</i>	No

Tabla 9. Resumen de la evidencia de los estudios analizados (Elaboración Propia)

2.4.2. LIMITACIONES

Pese al rigor metodológico empleado en esta revisión, se identifican diversas limitaciones que deben de ser consideradas al interpretar los resultados. En primer lugar, se evidenció una heterogeneidad sustancial entre los estudios incluidos, particularmente en lo referente a las características de la población, fuente de datos y criterios de inclusión, lo cual limita la posibilidad de realizar comparaciones directas o meta-análisis más robustos.

Adicionalmente, una proporción considerable de los estudios carece de validación externa con conjuntos de datos independientes, lo que compromete la generalización de los modelos predictivos a otros entornos clínicos. Esta ausencia de replicabilidad representa una barrera significativa para la implementación clínica.

La calidad de los datos empleados también varía entre los estudios, y en algunos casos no se documenta adecuadamente la procedencia del dataset, el tratamiento de valores ausentes o las estrategias de preprocesamiento, lo cual puede introducir sesgos

no controlados. A ello se suma la diversidad en las métricas de evaluación utilizadas: si bien métricas como AUC-ROC, F1-score o Brier score fueron frecuentes, también se emplearon medidas no estandarizadas, dificultando la comparación transversal del rendimiento.

Otro aspecto crítico es la interpretabilidad de los modelos. Muchos de los enfoques más eficaces se basan en algoritmos de tipo “caja negra”, como los ensambladores o redes neuronales, cuya falta de transparencia impide una interpretación clínica directa y limita su aceptación por parte de los profesionales de la salud.

2.4.3. CONCLUSIONES

La presente revisión sistemática permite concluir que los modelos de ML constituyen una herramienta prometedora para la predicción de mortalidad a 30 días (30DM) en pacientes con enfermedades cardíacas. Algoritmos como *XGBoost*, *Random Forest* y *Logistic Regression*, demostraron desempeños superiores en términos de discriminación y calibración, especialmente cuando se utilizaron métricas robustas como AUC-ROC, sensibilidad, especificidad y Brier score.

Las variables clínicas más representativas, comunes en los estudios como la edad, frecuencia cardíaca, creatinina, lugar de incidencia, anemia o el uso de vasopresores, destacan la importancia de contar con registros médicos electrónicos estructurados y completos. No obstante, la ausencia de validación externa en la mayoría de los estudios limita la aplicabilidad de los modelos desarrollados a contextos reales y diversos.

Con todo esto, se puede concluir que el campo aún enfrenta desafíos significativos en cuanto a la generalización, transparencia y utilidad clínica de los modelos predictivos de 30DM. Es imprescindible fomentar estudios prospectivos, con diseño riguroso y validación externa, que permitan trasladar estos avances al ámbito clínico de forma efectiva y ética.

3. ANÁLISIS DE DATOS

El objetivo del presente apartado es realizar un análisis y evaluación de los datos, previa al entrenamiento del modelo de aprendizaje para la predicción de la mortalidad a 30 días en pacientes de cirugía cardíaca.

3.1. METODOLOGÍA Y CRITERIOS

La metodología utilizada para el análisis de los datos se resume en el siguiente diagrama:

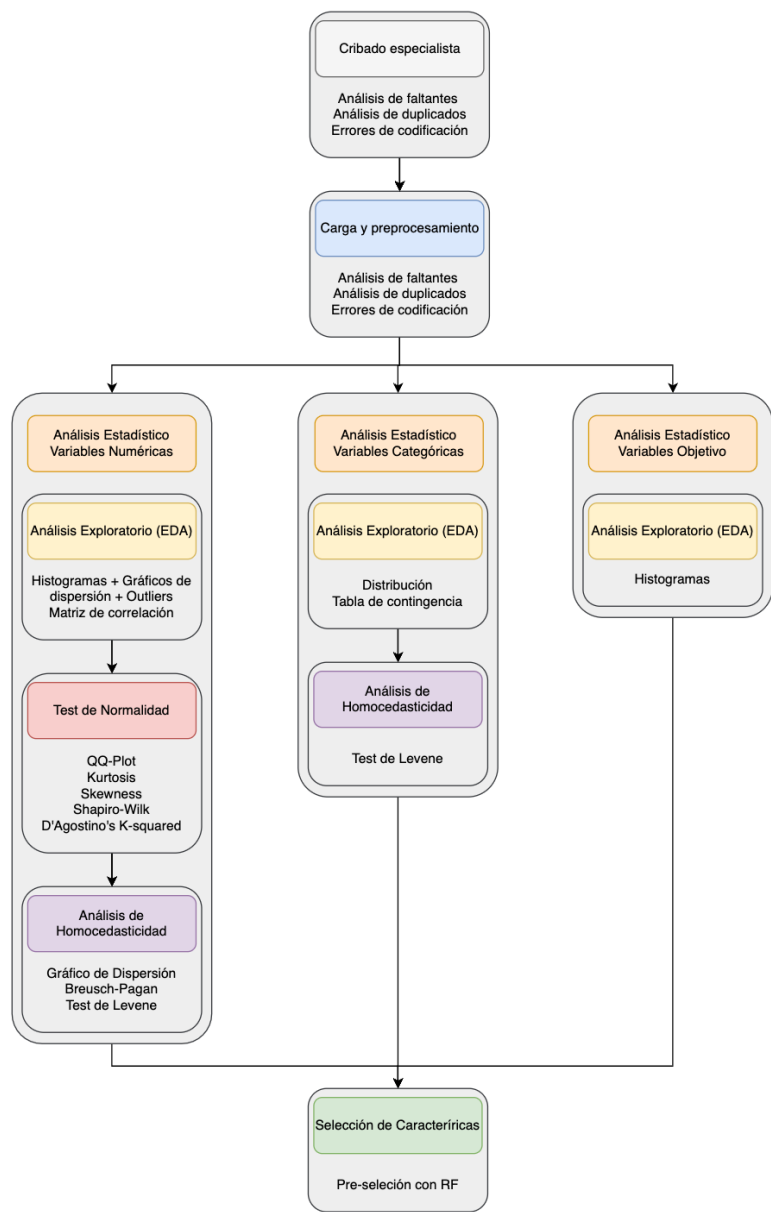


Ilustración 6. Metodología Análisis de Datos (Elaboración Propia)

3.1.1. CRIBADO ESPECIALISTA

Cribado de los datos en función de la importancia definida por el especialista clínico

Explicación detallada de cada variable ¿Qué es? ¿Qué representa? ¿En qué se mide? ¿Rango de valores estándar?

3.1.2. CARGA Y PREPROCESAMIENTO

Carga y preprocesamiento de los datos cribados

ANÁLISIS DE FALTANTES

Se estudiarán valores faltantes en los registros y se valorará la eliminación o la imputación de datos sobre los mismos

ANÁLISIS DE DUPLICADOS

Se analizará la presencia de registros duplicados en el dataset y se eliminarán los registros, en caso de que exista duplicidad

ERRORES DE CODIFICACIÓN

En el caso de las variables categóricas se estudiará si existen errores en la codificación de las categorías debidas:

- Errores de tipado
- Errores de mayúsculas, minúsculas
- Espacios en blanco

3.1.3. ANÁLISIS ESTADÍSTICO VARIABLES NUMÉRICAS

3.1.3.1 ANÁLISIS EXPLORATORIO (EDA)

HISTOGRAMAS

Representación de la frecuencia de aparición de los datos, indicando valores característicos como la media o la mediana

GRÁFICO DE DENSIDADES

Representación del KDE y de los diagramas de violín (*violin plot*) de cada variable, separada para sus posibles clases

El *KDE* (*Kernel Density Estimation*), es un gráfico que nos ayuda a estimar la probabilidad de la función de densidad de una variable

El Violin Plot, es una combinación de un boxplot y una representación de la densidad de probabilidad de los datos en esa área, básicamente consisten en una representación visual de la función de densidad de probabilidad (KDE) sobre el conjunto de datos

OUTLIERS

Se analizará, con el violín plot, la presencia o no de outliers y se contrastará con los “valores estándar” definidos para cada variable

MATRIZ DE CORRELACIÓN

La matriz de correlación es una tabla que muestra los coeficientes de correlación entre variables

El coeficiente de correlación es un valor que mide la “fuerza” y la “dirección” de la relación entre dos variables. Los valores de la correlación oscilan entre [-1; 1] indicando:

- -1; Correlación negativa fuerte
- 0; No hay correlación
- 1; Correlación positiva fuerte

El cálculo del coeficiente de correlación se hace según el estimador de *Pearson* (*r*) el cual se define como el producto de las desviaciones de la media de la variable 1 por las de la variable 2 divididas por el producto de las desviaciones estándar

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)S_x S_y}$$

En el estudio de correlación, se considerará que existe una correlación entre variables cuando el valor de r sea superior a 0,3 o inferior a -0,3

3.1.3.2 TEST DE NORMALIDAD

Evalúa si los datos siguen una distribución Gaussiana, para definir si se pueden utilizar, o no, test estadísticos paramétricos.

GRÁFICOS QQ

Los Gráficos QQ (*QQ-Plot*), o gráfico de cuantiles teóricos, comparan los cuantiles de la distribución observada con los cuantiles teóricos de una distribución normal con la misma media y desviación estándar que los datos

Cuanto más se aproximen los datos a una normal, más alineados están los puntos entorno a la recta

Interpretación de los resultados:

- Puntos siguen la línea diagonal => Normalidad en cada grupo
- Curvatura o desviaciones => No hay normalidad en ese grupo
- Extremos alejados => Posibles *outliers* en ese grupo

KURTOSIS

La *kurtosis* es una variable que indica la propensión de los datos a tener valores extremos, si tenemos un valor grande implica que existe un mayor número de valores de la variable muy dispersos y alejados del centro de esta

Interpretación de los resultados:

- Kurtosis = 3 - Distribución normal (mesocurtosis)
- Kurtosis > 3 - Distribución con colas más pesadas, es decir, la mayor parte de los datos aparecen en un rango vertical estrecho (leptocurtosis)
- Kurtosis < 3 - Distribución más plana (platicurtosis)
 - [-1; 1] – Ligera desviación
 - [-2; 2] – Evidente desviación (pero no extrema)

SKEWNESS

El *skewness*, o asimetría, es una variable que indica si los datos están sesgados hacia valores menores o mayores

Interpretación de los resultados:

- $Skewness = 0$ – Distribución normal
- $Skewness > 0$ – Sesgo a la derecha (cola larga hacia la derecha)
- $Skewness < 0$ – Sesgo a la izquierda (cola larga hacia la izquierda)

SHAPIRO-WILK

El Shapiro-Will, es una prueba estadística utilizada para contrastar la normalidad de un conjunto de datos, que plantea como hipótesis nula, que la muestra proviene de una población normalmente distribuida.

El estadístico de la prueba es:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Siendo la hipótesis nula que la población está distribuida normalmente, si el p-valor es menor al nivel de significancia establecido (0,05 para este estudio), la hipótesis nula será rechazada, es decir:

- $p\text{-valor} < 0,05$ – No hay normalidad
- $p\text{-valor} > 0,05$ – Si hay normalidad

D'AGOSTINO'S K-SQUARED

El D'Agostino's K^2 test, es una prueba estadística para determinar si un conjunto de datos está normalmente distribuido.

La prueba calcula la diferencia entre la función de distribución acumulativa (CDF) observada de los datos y el CDF esperado bajo el supuesto de normalidad. Esta diferencia se cuadra y se compara con un valor crítico de la distribución chi-cuadrada.

Si el valor excede el valor crítico ($<0,05$ para este estudio), se rechaza la hipótesis de nula normalidad, lo que indica que los datos no siguen una distribución normal

3.1.3.3 ANÁLISIS DE HOMOCEDASTICIDAD

La homocedasticidad nos indica cuando la varianza del error (residuales) es constante en todos los niveles de la variable independiente, es decir, mide si la varianza es constante en los datos

Para variables numéricas se define a través de los gráficos de dispersión, histogramas y tendencias y para las variables categóricas se hace a través de la Prueba de Levene

Evalúa si las variables de dos o más grupos son iguales (homocedasticidad) o si son diferentes (heterocedasticidad)

3.1.4. ANÁLISIS ESTADÍSTICO VARIABLES CATEGÓRICAS

3.1.4.1 ANÁLISIS EXPLORATORIO (EDA)

HISTOGRAMAS

Representación de la frecuencia de aparición de los datos, utilizado para observar la distribución de las diferentes categorías

TABLA DE CONTINGENCIA

Las Tablas de Contingencia son un registro del recuento entre dos o más variables categóricas, en esta tabla se representa un recuento de cada categoría de la variable

Para una mejor interpretación, se representarán las Tablas de Contingencia de cada variable categórica tanto en datos absolutos de recuento como en datos porcentuales, para poder hacer una evaluación en aquellas clases en las que exista un recuento superior al 80% y se pueda comparar contra el recuento de valores absolutos, definiendo de esta manera si se trata, o no, de un valor representativo

3.1.4.2 ANÁLISIS DE HOMOCEDASTICIDAD

TEST DE LEVENE

El test de *Levene* es una prueba estadística utilizada para evaluar la igualdad de varianzas (homocedasticidad) de dos o más grupos.

Si el valor p del test de Levene es menor que un nivel de significancia ($<0,05$ para este estudio), se rechaza la hipótesis nula, concluyéndose que existe heterocedasticidad, es decir, que las varianzas no son iguales entre grupos.

3.2. DATASET

Los datos se recogieron prospectivamente en el *HCUV*, en el marco de un registro clínico de pacientes sometidos a cirugía cardíaca, dispone de 325 registros y 141 variables diferentes. Se trata de un *dataset* pequeño para *machine learning*, por lo que será imprescindible aplicar la validación cruzada.

3.3. CRIBADO ESPECIALISTA

Tras un análisis previo del *dataset* realizado por un especialista clínico en cardiología, se definen las siguientes características como las más importantes de cara a su evaluación

VARIABLES NUMÉRICAS

VARIABLE	¿Qué es?	¿Qué representa?	¿En qué se mide?	Rango de valores estándar
EDAD ['Edad']	Edad del paciente	Años que tiene el paciente en el momento del estudio	Años	N/A
IMC ['IMC']	Índice de Masa Corporal	Relación entre el peso y la altura de una persona. Se usa para clasificar el estado nutricional.	kg/m ²	Bajo peso: < 18,5 kg/m ² Normal: 18,5-24,9 kg/m ² Sobrepeso: 25-29,9 kg/m ² Obesidad:>30 kg/m ²
HTOPRE ['Htopre']	Hematocrito preoperatorio	Porcentaje del volumen sanguíneo ocupado por los glóbulos rojos antes de la cirugía	%	Hombres: 40-52% Mujeres: 36-58%
LEUCOGRAMA ['Leucograma']	Conteo de leucocitos en sangre	Mide la cantidad de glóbulos blancos, un indicador de infecciones o inflamaciones	Células por microlitro de sangre cél/μL	4.000-11-000 cél/μL
CREATININAPRE ['Creatininapre']	Nivel de creatinina en sangre antes de la cirugía	Evalúa la función renal, ya que la creatinina es un producto de deshecho de los músculos, eliminado por los riñones	Miligramos por decilitro mg/dL	Hombres: 0,7-1,3 mg/dL Mujeres: 0,6-1,1 mg/dL
PROTEINASA TOTALES ['Proteinasatotales']	Nivel total de proteínas en sangre	Indica el estado nutricional y puede reflejar problemas hepáticos o renales	Gramos por decilitro gr/dL	6,4-8,3gr/dL
PCR ['PCR']	Proteína C Reactiva	Reactante de fase aguda inespecífico	Miligramos por litro [mg/L]	Normal: <2 mg/L Inflamación Moderada: 3-10mg/L

				Inflamación severa: >10mg/L
TEST FUERZA AGARRE [‘FuerzaTestAgarre’]	Prueba de fuerza de prensión manual	Evalúa la fuerza muscular del paciente. Usada como indicador del estado funcional y nutricional	Kilogramos de fuerza [Kg]	Hombres: 30-55Kg Mujeres: 20-35Kg
TEST 5 METROS [‘Test5metros ’]	Resultado de prueba de velocidad de la marcha	Evalúa el tiempo que emplea el paciente en recorrer 5 metros y refleja el estado funcional del paciente	Segundos [s]	Normal: <=6s Riesgo de fragilidad: >6s Alto riesgo de caídas: >8s
ESTANCIA HOSPITALARIA CALCULADA [‘Estanciahospitalari acalculada’]	Tiempo de hospitalización	El tiempo de hospitalización del paciente	Días	N/A

Tabla 10. Descripción Variables Numéricas (Elaboración Propia)

VARIABLES CATEGÓRICAS

VARIABLE	¿Qué es?	¿Qué representa?	¿En qué se mide?	Rango de valores estándar
GÉNERO [‘Género’]	Sexo del paciente	Diferencia entre masculino y femenino	Categórica (Binaria)	Hombre = 0 Mujer = 1
DM DICOTÓMICA [‘DMdicotomica’]	Diabetes mellitus (DM)	Presencia o ausencia de diabetes	Categórica (Binaria)	Diabetes SI = 1 Diabetes NO = 0
TABAQUISMO DICOT [‘TabaquismoDicot’]	Tabaquismo	Indica si el paciente es fumador o no	Categórica (Binaria)	Fumador = 1 No Fumador = 0
HTA [‘HTA’]	Hipertensión arterial	Diagnóstico de hipertensión	Categórica (Binaria)	Si = 1 No = 0
DL [‘DL’]	Dislipidemia	Niveles anormales de lípidos en sangre	Categórica (Binaria)	No = 0 Si = 1
Obesidad [‘Obesidad’]	Índice de masa corporal elevado	Paciente con obesidad según criterios médicos	Categórica (Binaria)	No = 0 Si = 1
EPOC Dicot [‘EPOCdicot’]	Enfermedad Pulmonar Obstructiva Crónica (EPOC)	Diagnóstico de EPOC	Categórica (Binaria)	No = 0 Si = 1

Insuficiente preoperatoria [‘Insuficienciapreoperatoria’]	Insuficiencia renal preoperatoria	Disfunción renal antes de la cirugía	Categórica (Binaria)	No = 0 Si = 1
Complicaciones Todas [‘ComplicacionesTODAS’]	Presencia de complicaciones	Si el paciente tuvo alguna complicación tras el procedimiento	Categórica (Binaria)	No = 0 Si = 1
Complicaciones MACE [‘ComplicacionesMACE’]	Eventos adversos cardiovasculares mayores	Complicaciones cardiovasculares graves tras el procedimiento	Categórica (Binaria)	No = 0 Si = 1
Fragilidad GrST [‘FragilidadGrST’]	Índice de fragilidad de grupos ST	Nivel de fragilidad según el <i>Test de Fuerza de Agarre (Grip Strength Test)</i>	Categórica (Binaria)	No = 0 Si = 1
Fragilidad GST [‘FragilidadGST’]	Índice de fragilidad GST	Clasificación de fragilidad según el <i>Test de Velocidad de la Marcha (Gait Speed Test)</i>	Categórica (Binaria)	No = 0 Si = 1
Barthel [‘Barthel’]	Índice de Barthel	Mide la independencia funcional en actividades diarias	Puntos [0-100]	Dependencia total: 0-20 Dependencia severa: 21-60 Dependencia moderada: 61-90 Dependencia escasa: 91-9 Independencia: 100
Katz [‘Katz’]	Escala de Katz	Evalúa la capacidad para realizar actividades básicas diarias	Puntos [0-6]	Independiente: 0-1 Dependencia moderada: 2-3 Dependencia Severa: 4-6
Frail [‘Frail’]	Escala de Frail	Evalúa la fragilidad en adultos mayores	Puntos [0-6]	>1: Fragilidad
Edmonton [‘Edmonton’]	Escala de fragilidad de Edmonton	Mide la fragilidad en adultos mayores con comorbilidades	Puntos [0-17]	No frágil: 0-4 Frag. Leve: 5-6 Frag. Moderada: 7-8 Frag. Severa: >9

Tabla 11. Descripción Variables Categóricas (Elaboración Propia)

VARIABLE OBJETIVO

VARIABLE	¿Qué es?	¿Qué representa?	¿En qué se mide?	Rango de valores estándar
MORTALIDAD 30 DÍAS ['Mortalidad30']	Mortalidad a 30 días	Indica si el paciente falleció dentro de los 30 días posteriores al procedimiento	Categórica (Binaria)	Vivo = 0 Muerto = 1

Tabla 12. Descripción Variable Objetivo (Elaboración Propia)

3.4. CARGA Y PREPROCESAMIENTO

3.4.1. ANÁLISIS DE FALTANTES

VARIABLES NUMÉRICAS

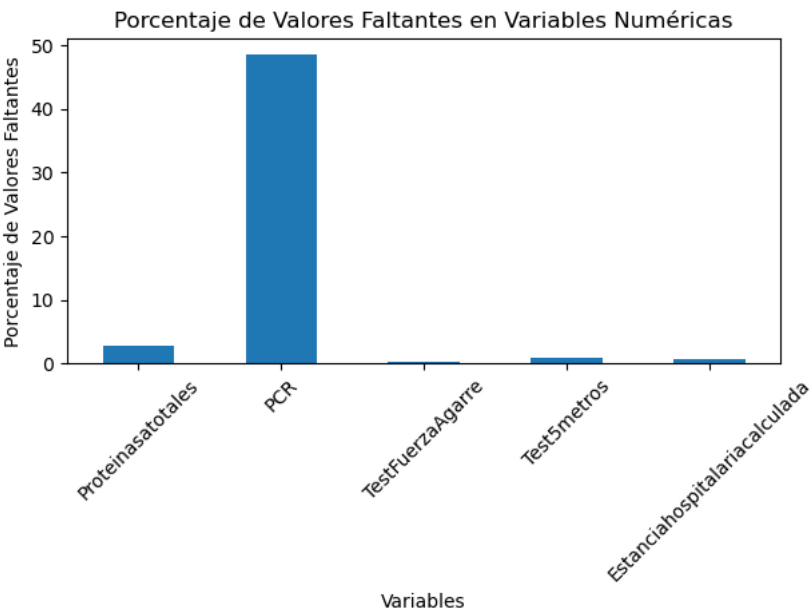


Ilustración 7. Análisis de faltantes (Elaboración Propia)

Para completar los valores faltantes de las variables numéricas se realiza una imputación de los datos faltantes utilizando los valores de mediana de la distribución de cada variables

VARIABLES CATEGÓRICAS

Respecto a las variables categóricas, no se observan valores faltantes en los registros

3.4.2. ANÁLISIS DE DUPLICADOS

No se observa ningún registro duplicado en el *dataset*

3.4.3. ERRORES DE CODIFICACIÓN

No se observa ningún error de codificación (errores mayúsculas/minúsculas, espacios en blanco, faltas ortográficas...) en las variables categóricas

3.5. ANÁLISIS ESTADÍSTICO VARIABLES NUMÉRICAS

3.5.1. ANÁLISIS EXPLORATORIO (EDA)

3.5.1.1 HISTOGRAMAS + GRÁFICOS DE DISPERSIÓN + OUTLIERS

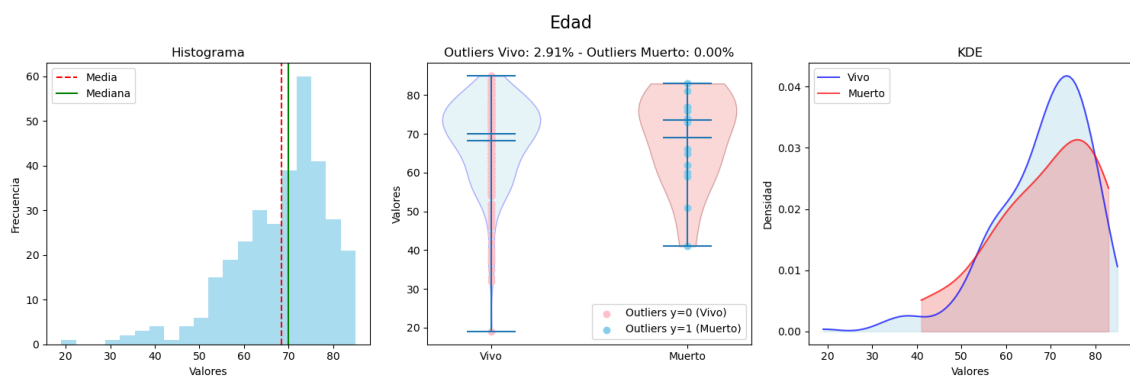


Ilustración 8. EDA Edad (Elaboración Propia)

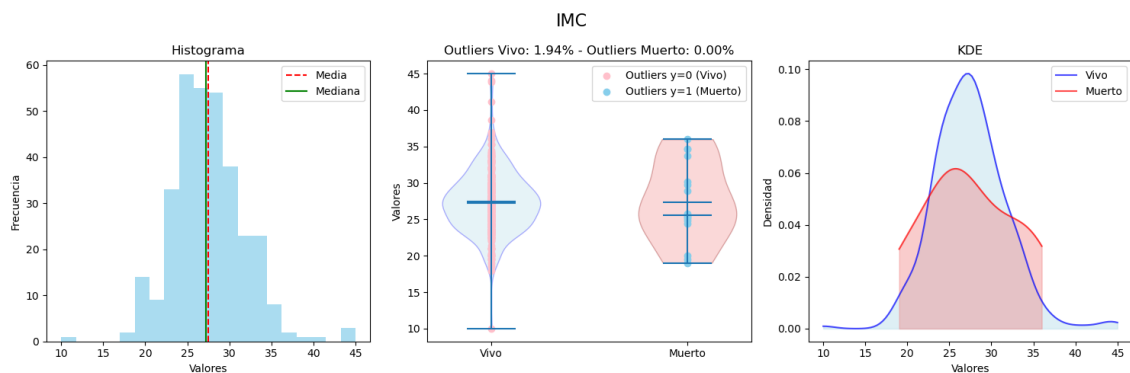


Ilustración 9. EDA IMC (Elaboración Propia)

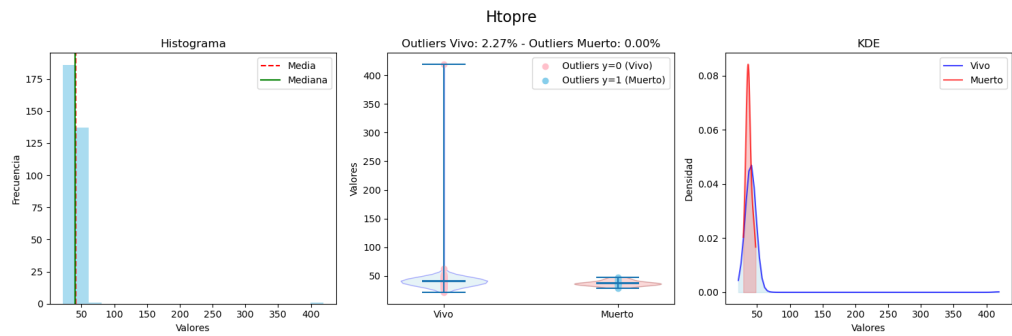


Ilustración 10. EDA *Htopre* (Elaboración Propia)

Se observa un registro con un valor de *Htopre* (Hematocrito preoperatorio) fuera del rango estándar definido, por lo que debería de eliminarse del dataset (perteneciente a la Clase “Vivo”)

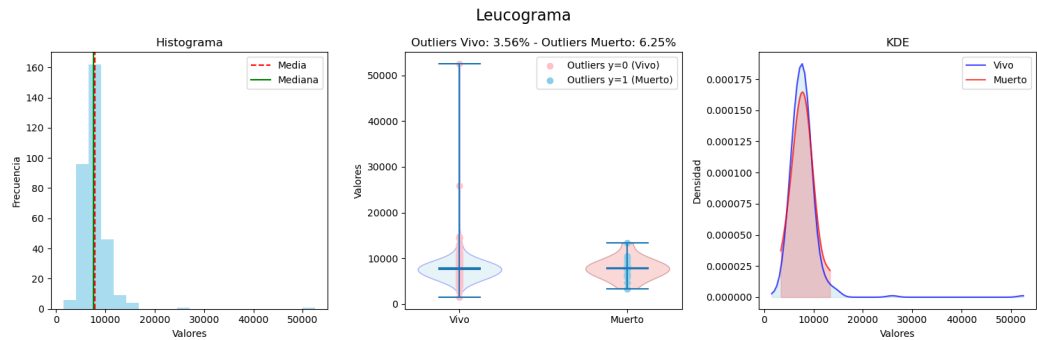


Ilustración 11. EDA *Leucograma* (Elaboración Propia)

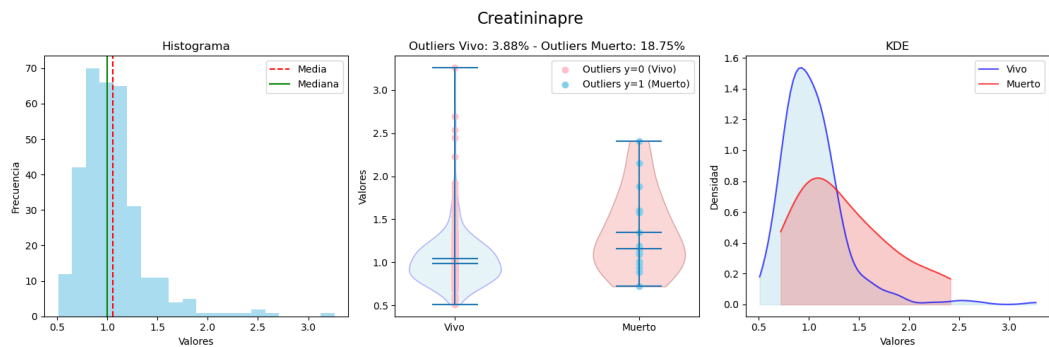


Ilustración 12. EDA *Creatininapre* (Elaboración Propia)

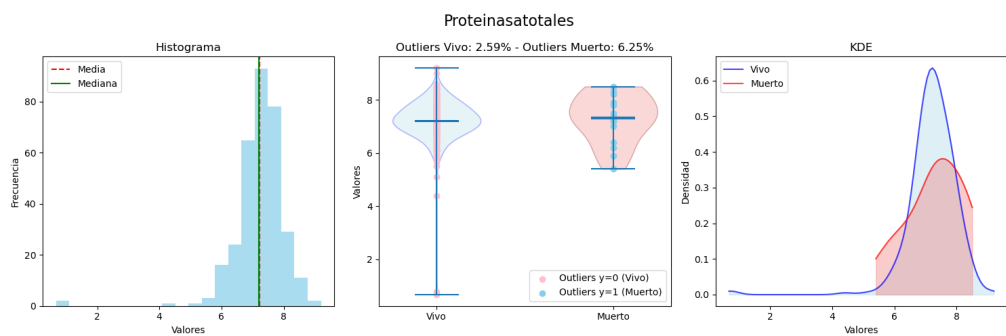


Ilustración 13. EDA *Proteinasatotales* (Elaboración Propia)

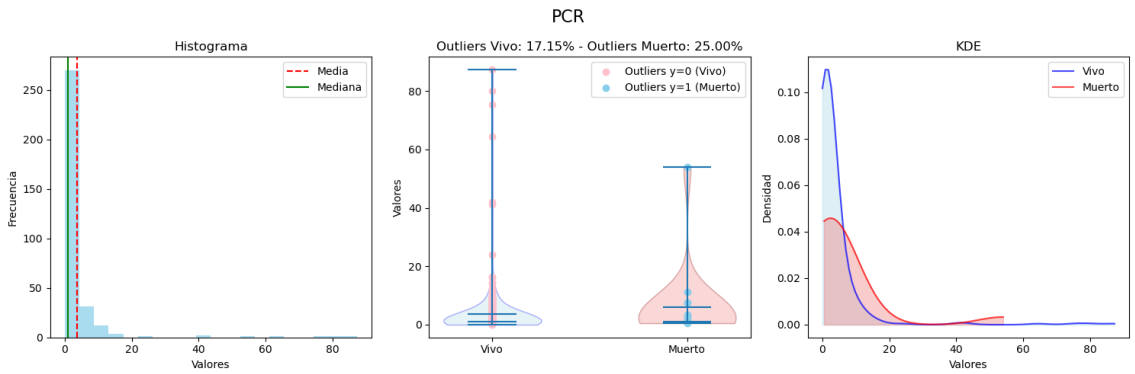


Ilustración 14. EDA PCR (Elaboración Propia)

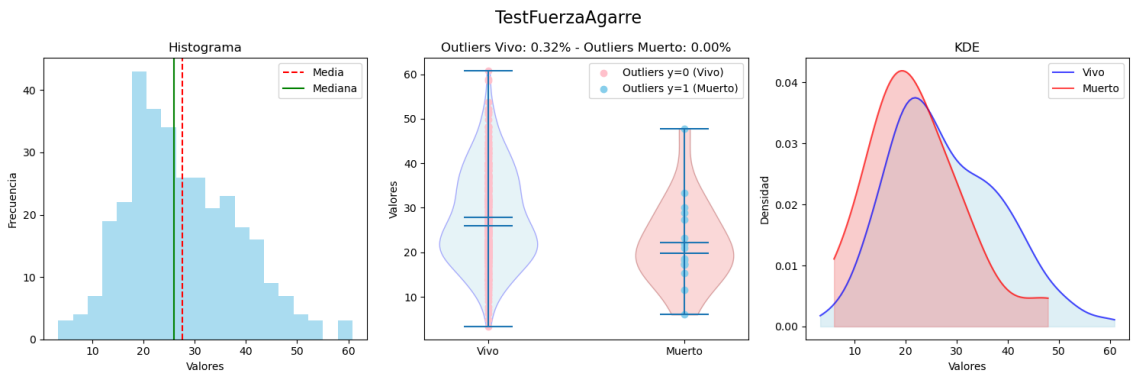


Ilustración 15. EDA Test Fuerza Agarre (Elaboración Propia)

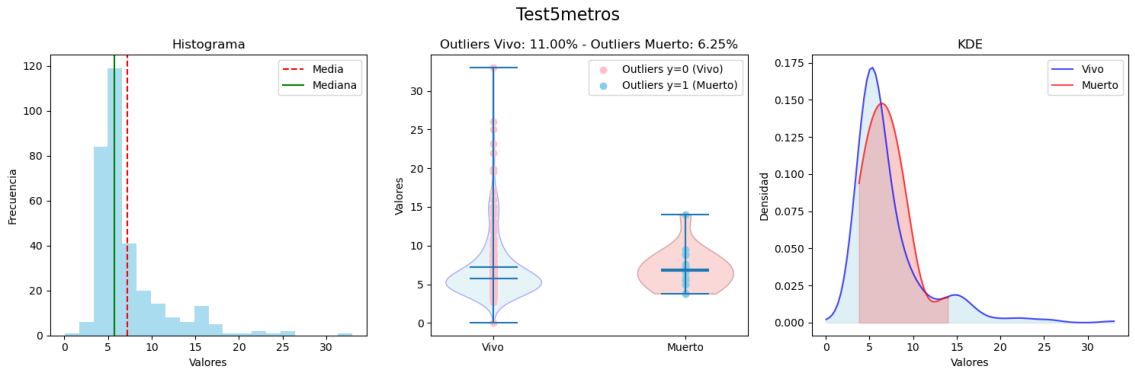


Ilustración 16. EDA Test5metros (Elaboración Propia)

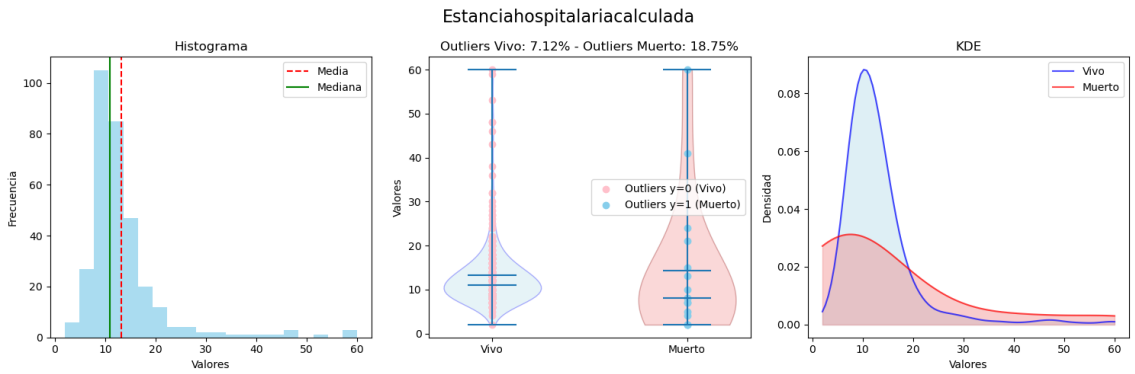


Ilustración 17. EDA Estanciahospitalariacalculada (Elaboración Propia)

3.5.1.2 MATRIZ DE CORRELACIÓN

Matriz de correlación de las variables numéricas

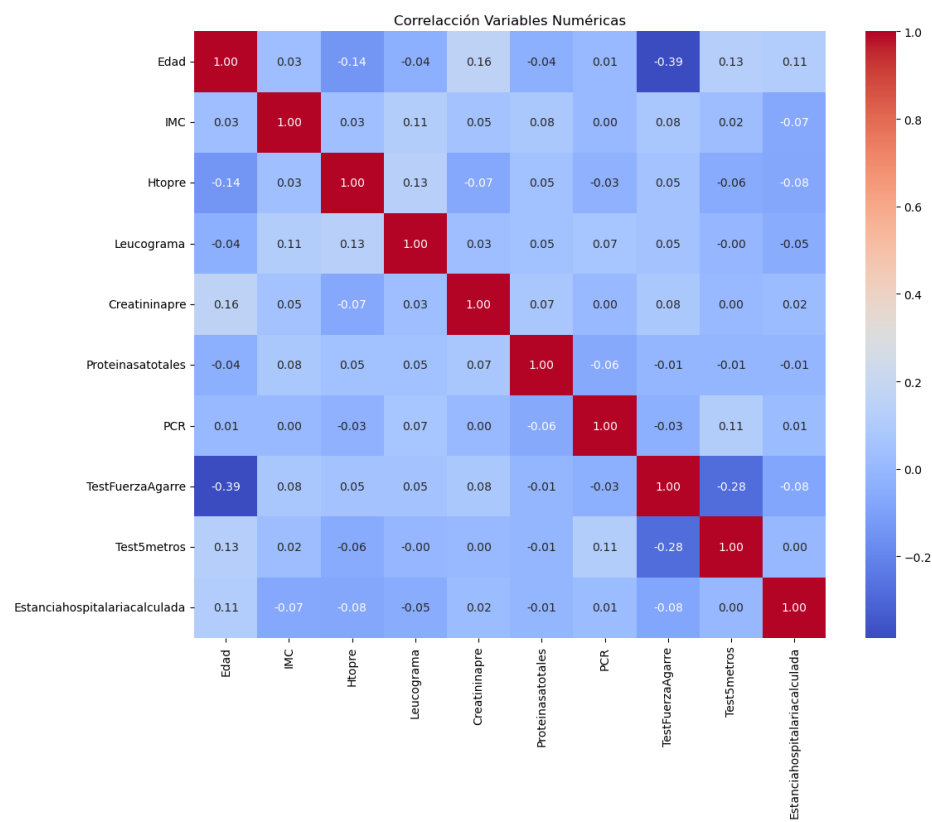


Ilustración 18. Matriz de correlación de las variables numéricas (Elaboración Propia)

Correlación de las variables numéricas con la variable objetivo

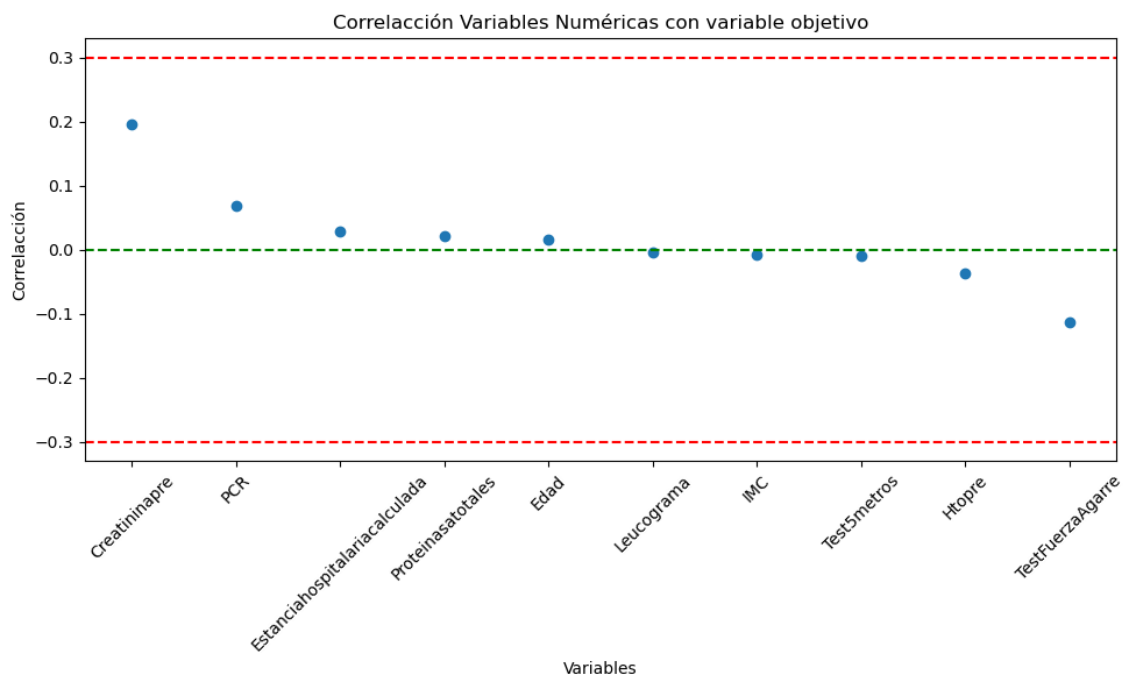


Ilustración 19. Correlación de las variables numéricas con la variable objetivo (Elaboración Propia)

3.5.2. TEST DE NORMALIDAD

3.5.2.1 DIAGRAMAS QQ

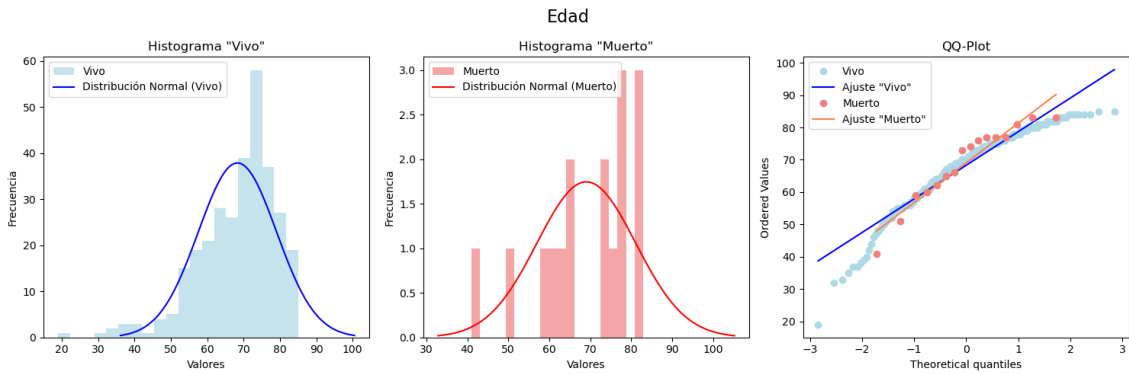


Ilustración 20. QQ Edad (Elaboración Propia)

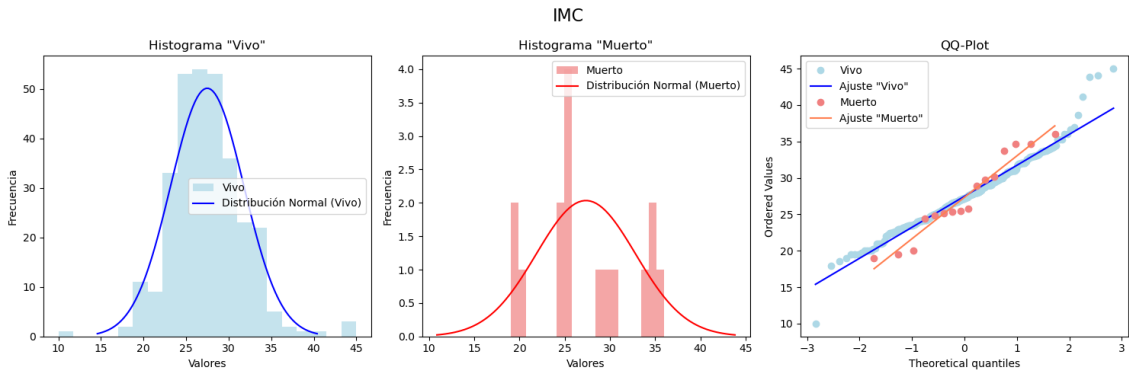


Ilustración 21. QQ IMC (Elaboración Propia)

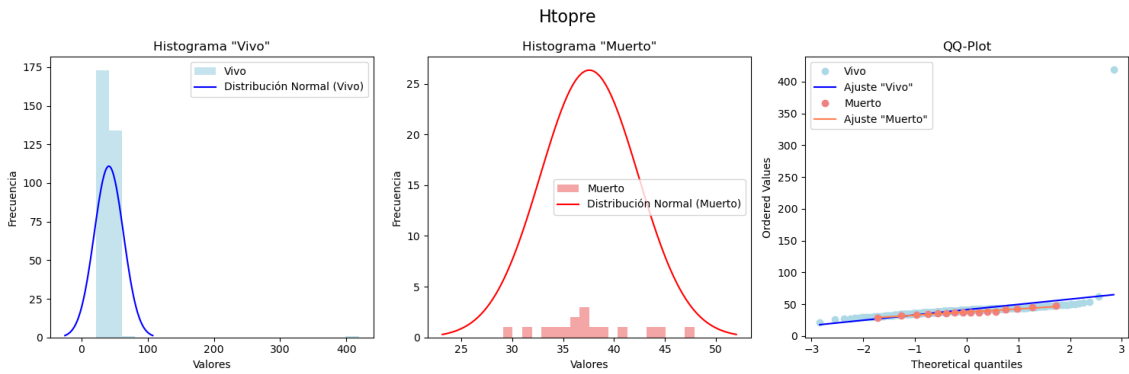


Ilustración 22. QQ Htopre

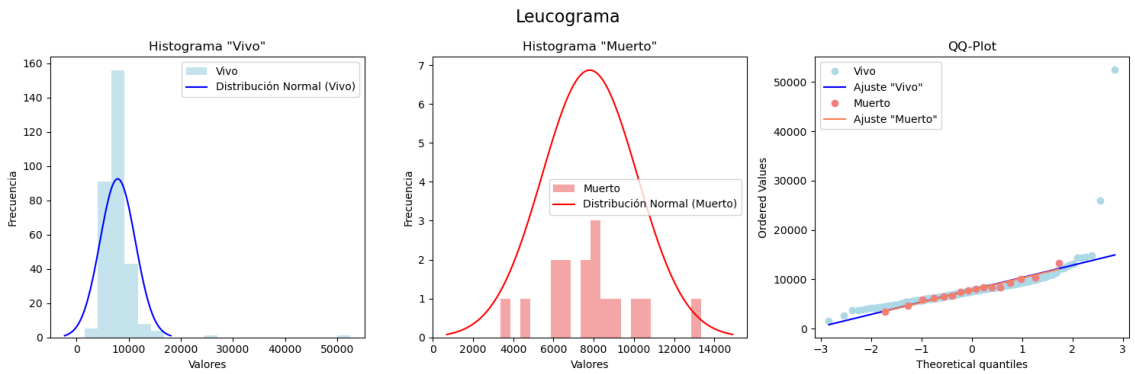


Ilustración 23. QQ Leucograma (Elaboración Propia)

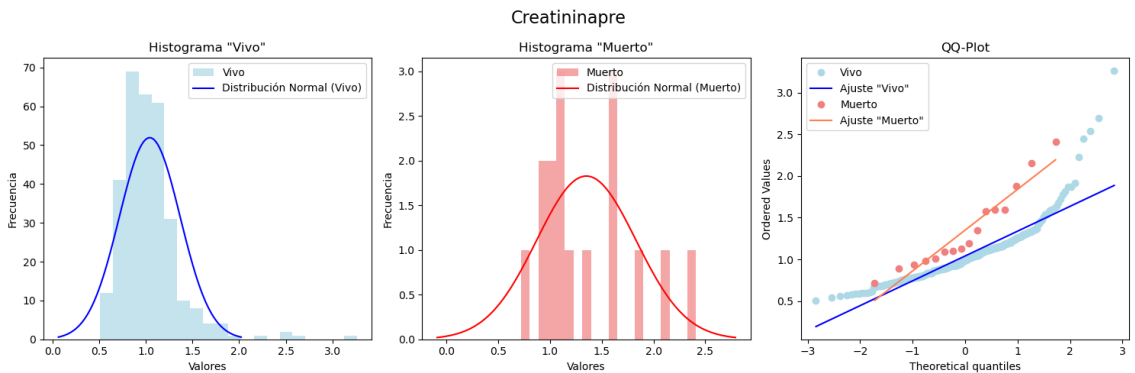


Ilustración 24. QQ Creatininapre (Elaboración Propia)

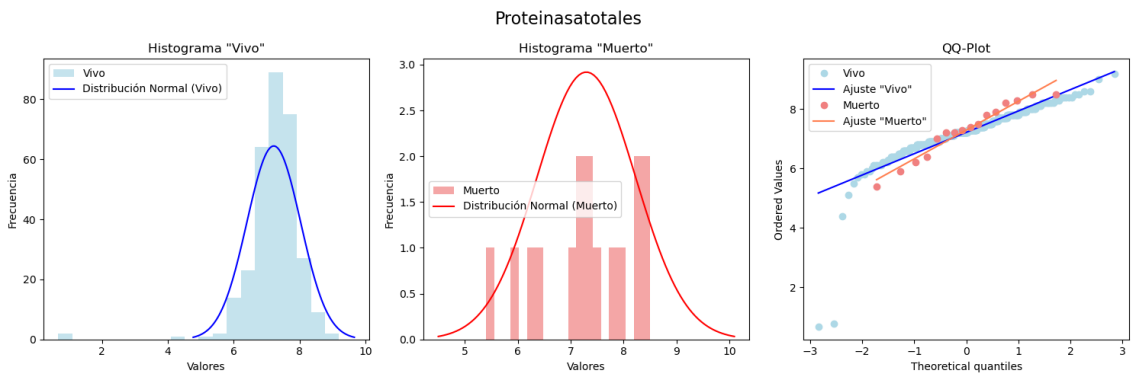


Ilustración 25. QQ Proteinasatotales (Elaboración Propia)

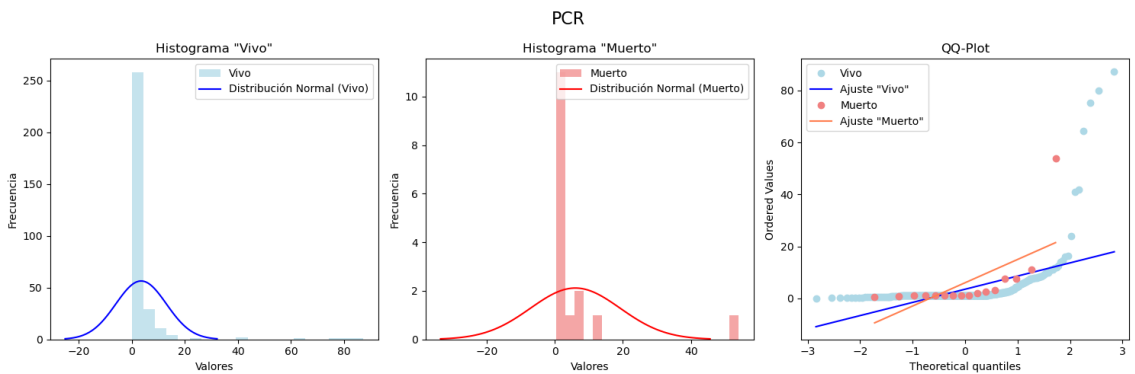


Ilustración 26. QQ PCR (Elaboración Propia)

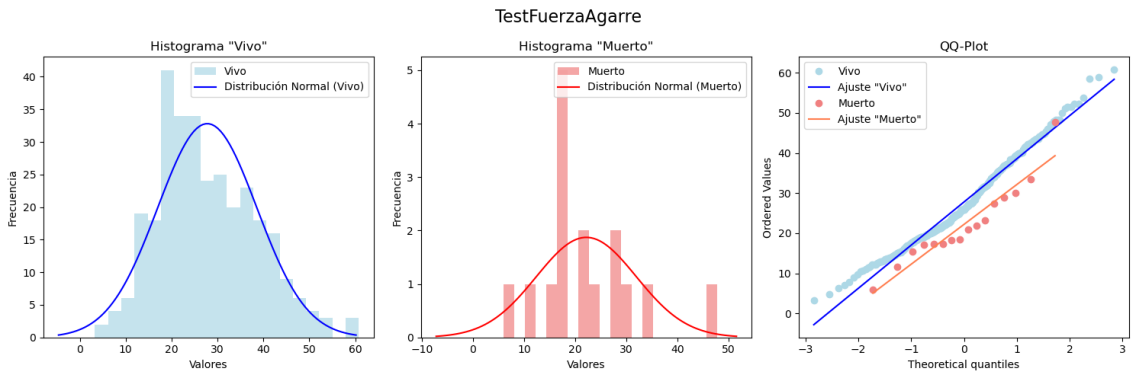


Ilustración 27. QQ TestFuerzaAgarre (Elaboración Propia)

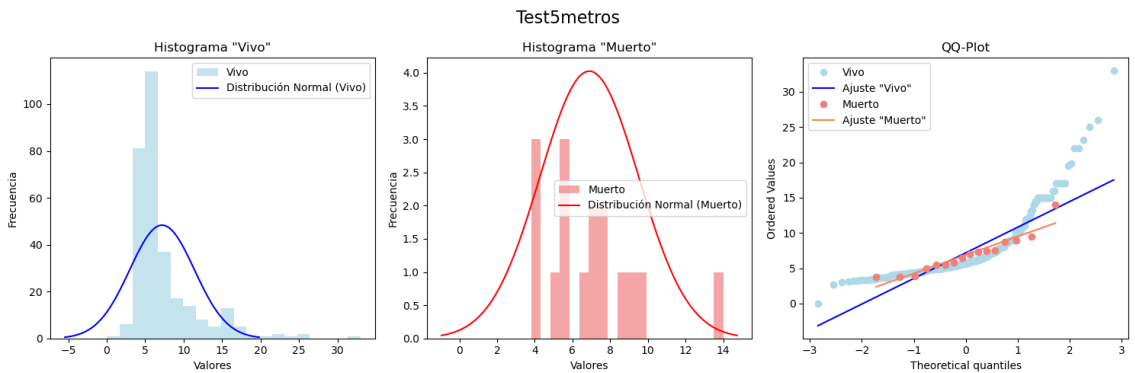


Ilustración 28. QQ Test5metros (Elaboración Propia)

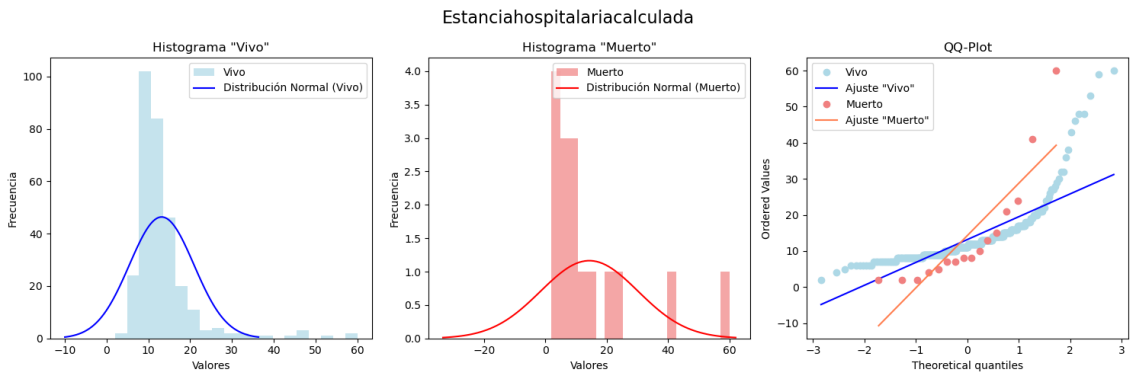


Ilustración 29. QQ Estanciahospitalariacalculada (Elaboración Propia)

3.5.2.2 KURTOSIS + SKEWNESS + SHAPIRO-WILK + D'AGOSTINO'S

VARIABLE	CLASE	KURTOSIS	SWEKNESS	SHAPIRO-WILK	D'AGOSTINO'S
Edad	Vivo	1,844	-1,143	$4,8 \cdot 10^{-11}$	$2,166 \cdot 10^{-14}$
	Muerto	0,331	-0,915	0,125	0,227
IMC	Vivo	2,227	0,495	$3,8 \cdot 10^{-6}$	$1,6 \cdot 10^{-7}$
	Muerto	-1,001	0,108	0,239	0,546
Htopre	Vivo	279,44	16,303	$5,4 \cdot 10^{-35}$	$6,36 \cdot 10^{-151}$
	Muerto	0,369	0,556	0,662	0,503
Leucograma	Vivo	99,165	7,981	$1,3 \cdot 10^{-27}$	$2,1 \cdot 10^{-105}$
	Muerto	1,086	0,407	0,916	0,408
Creatininapre	Vivo	10,765	2,438	$5,3 \cdot 10^{-18}$	$5,1 \cdot 10^{-42}$
	Muerto	0,101	0,915	0,134	0,249
Proteinasatotales	Vivo	25,074	-3,382	$3,4 \cdot 10^{-20}$	$2,8 \cdot 10^{-59}$
	Muerto	-0,437	-0,555	0,413	0,574
PCR	Vivo	48,341	6,626	$8,1 \cdot 10^{-33}$	$2,2 \cdot 10^{-91}$
	Muerto	13,826	3,638	$7,3 \cdot 10^{-7}$	$3,7 \cdot 10^{-9}$
TestFuerzaAgarre	Vivo	-0,213	0,451	$2,7 \cdot 10^{-4}$	$5,1 \cdot 10^{-4}$
	Muerto	2,104	1,044	0,239	0,045
Test5metros	Vivo	7,676	2,428	$1,2 \cdot 10^{-21}$	$3,6 \cdot 10^{-39}$
	Muerto	2,506	1,245	0,081	0,017
Estancia_hospitalaria_calculada	Vivo	13,833	3,287	$4,7 \cdot 10^{-24}$	$2,6 \cdot 10^{-43}$
	Muerto	4,036	2,021	$5,8 \cdot 10^{-4}$	$3,4 \cdot 10^{-4}$

Tabla 13. Métricas Test de Normalidad (Elaboración Propia)

Observamos que las variables “Edad”, “IMC”, “Htopre”, “Leucograma”, “Proteinasa”, “PCR”, “Test5metros” y “Estanciahospitalariacalculada” no sigue una distribución normal en al menos uno de sus grupos, generalmente en la clase de “Vivos”. Esto justifica el uso de test no paramétricos para comparar las variables entre “Vivos” y “Muertos”. También implica que para modelos predictivo, puede ser necesario realizar transformaciones o utilizar modelos robustos o no paramétricos.

3.5.3. ANÁLISIS DE HOMOCEDASTICIDAD

3.5.3.1 DIAGRAMAS DE DISPERSIÓN

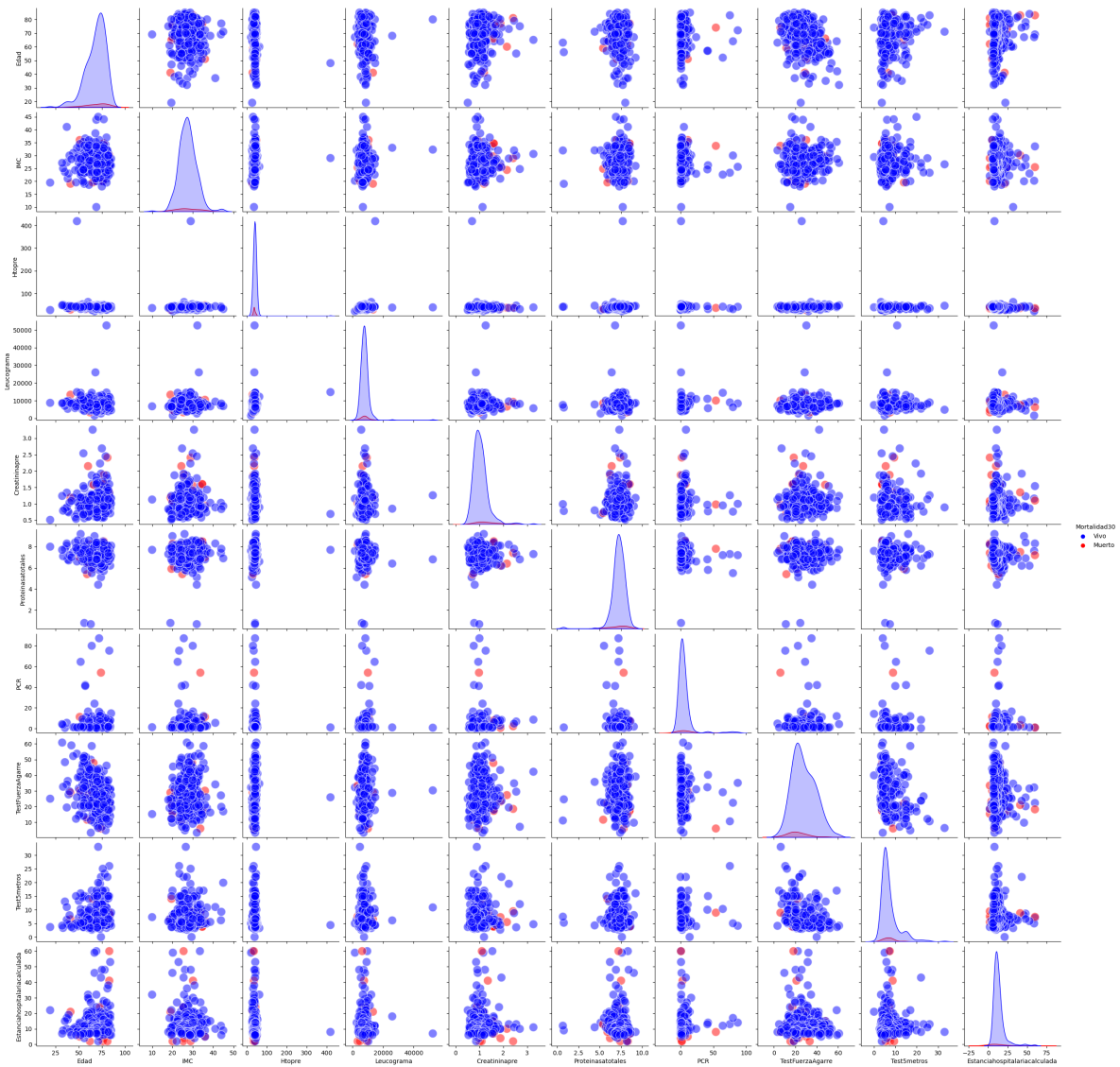


Tabla 14. Diagramas de dispersión (Elaboración Propia)

De los diagramas de dispersión podemos observar:

- Distribuciones no simétricas, muchos histogramas diagonales muestran asimetría (*skewness*), lo que respalda los resultados del test de normalidad, los datos no son normalmente distribuidos para muchas variables.
- Patrones de dispersión (*off-diagonal*), en varios diagramas de dispersión se observan formas de embudo o de expansión de la nube de puntos, lo cual es un indicio de heterocedasticidad.
- Separación por clases, las clases “Vivo” y “Muerto” no están claramente separadas en la mayoría de los pares de variables, esto sugiere que los modelos lineales simples pueden no tener gran capacidad discriminativa sin selección de variables.

3.5.3.2 TEST DE BREUSCH- PAGAN

p-valor = 0,006678 < 0,05

Rechazamos la hipótesis nula

No hay homocedasticidad, hay heterocedasticidad

3.5.3.3 TEST DE LEVENE

Si $p\text{-valor} > 0,05$, no se rechaza la hipótesis nula \Rightarrow Hay Homocedasticidad

Si $p\text{-valor} < 0,05$, se rechaza la hipótesis nulas \Rightarrow Hay Heterocedasticidad

Variable Numérica	$p\text{-valor}$	Resultado
<i>Edad</i>	0,2138 > 0,05	Homocedasticidad
<i>IMC</i>	0,2184 > 0,05	Homocedasticidad
<i>Htopre</i>	0,7171 > 0,05	Homocedasticidad
<i>Leucograma</i>	0,7611 > 0,05	Homocedasticidad
<i>Creatininapre</i>	0,1695 > 0,05	Homocedasticidad
<i>Proteinasatotales</i>	0,6390 > 0,05	Homocedasticidad
<i>PCR</i>	0,4603 > 0,05	Homocedasticidad
<i>TestFuerzaAgarre</i>	0,030 < 0,05	Heterocedasticidad
<i>Test5metros</i>	0,516 > 0,05	Homocedasticidad
<i>Estanciahospitalaiacalculada</i>	0,011 < 0,05	Heterocedasticidad

Tabla 15. Test de Levene Variables Numéricas (Elaboración Propia)

3.6. ANÁLISIS ESTADÍSTICO VARIABLES CATEGÓRICAS

3.6.1. ANÁLISIS EXPLORATORIO (EDA)

3.6.1.1 HISTOGRAMAS + TABLAS DE CONTINGENCIA

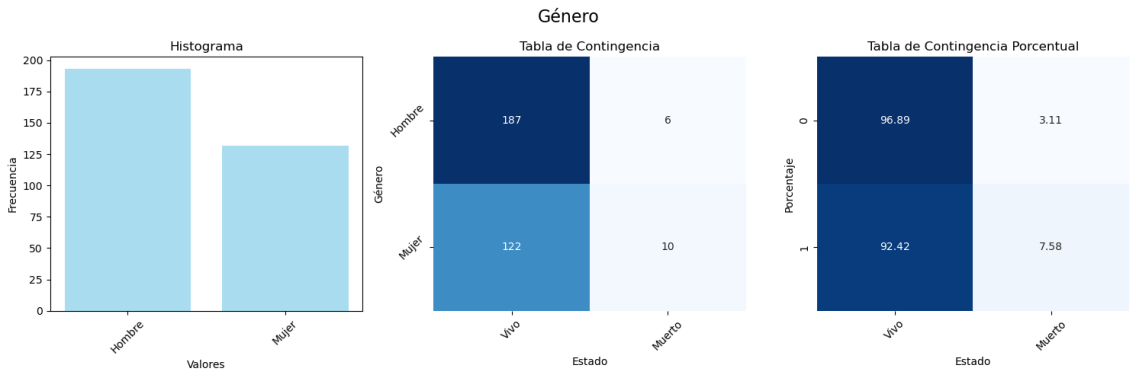


Ilustración 30. EDA Género (Elaboración Propia)

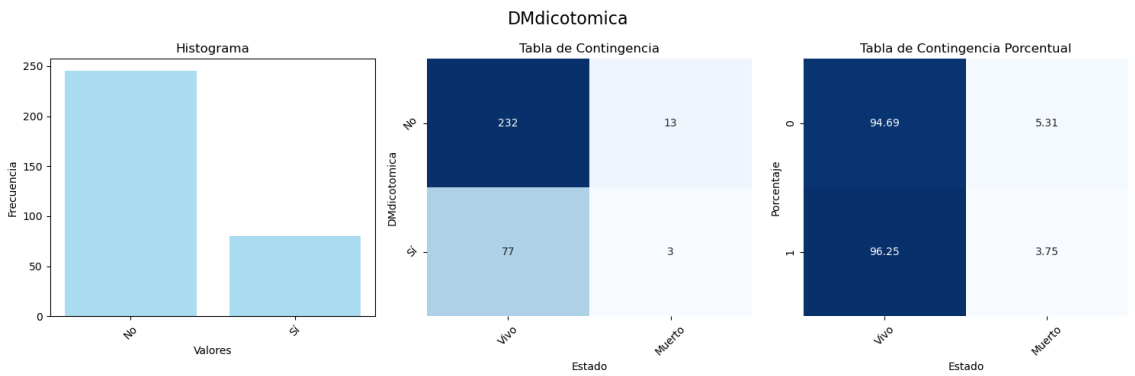


Ilustración 31. EDA DMdicotomica (Elaboración Propia)

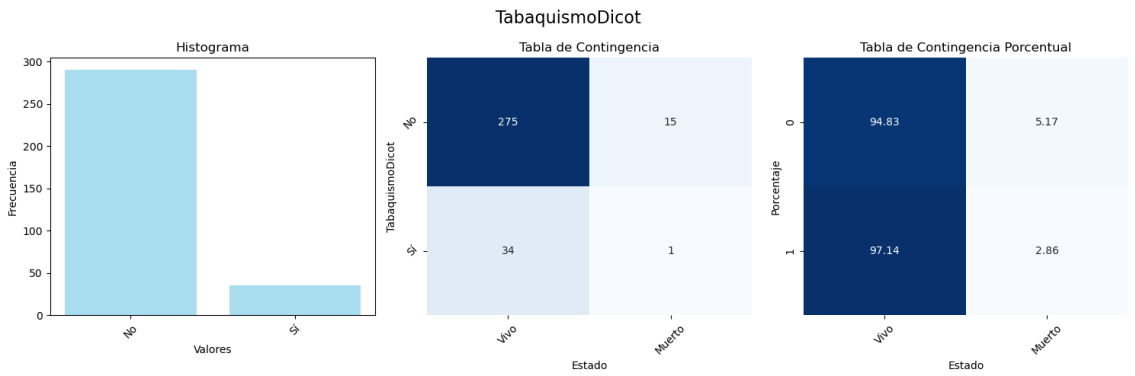
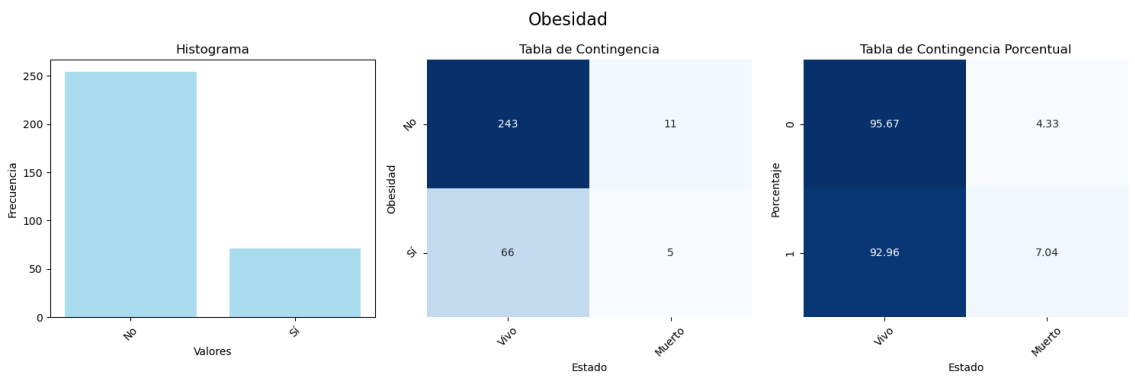
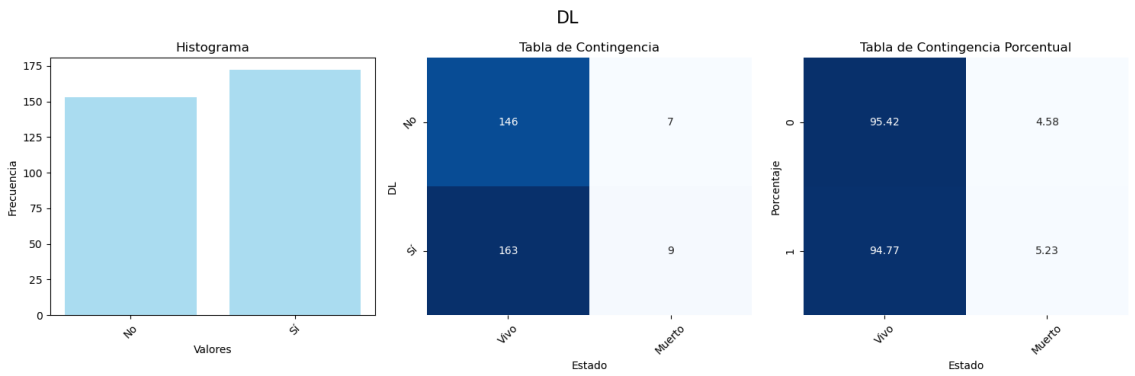
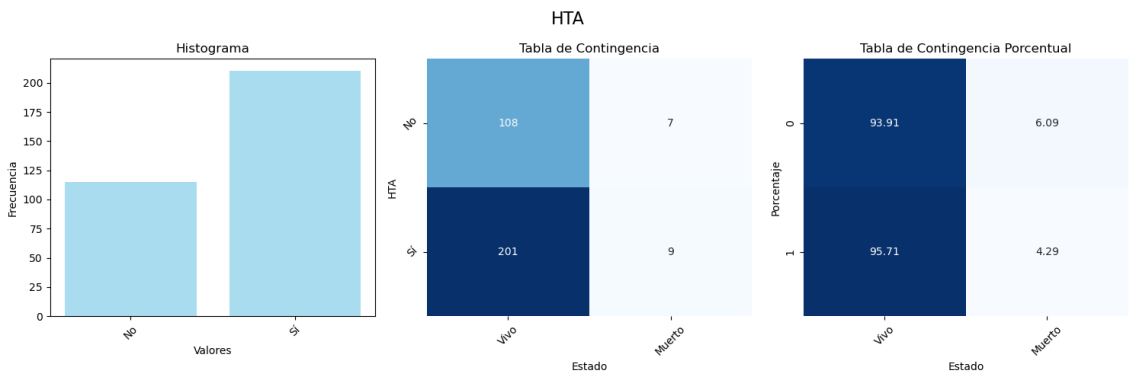


Ilustración 32. EDA TabaquismoDicot (Elaboración Propia)



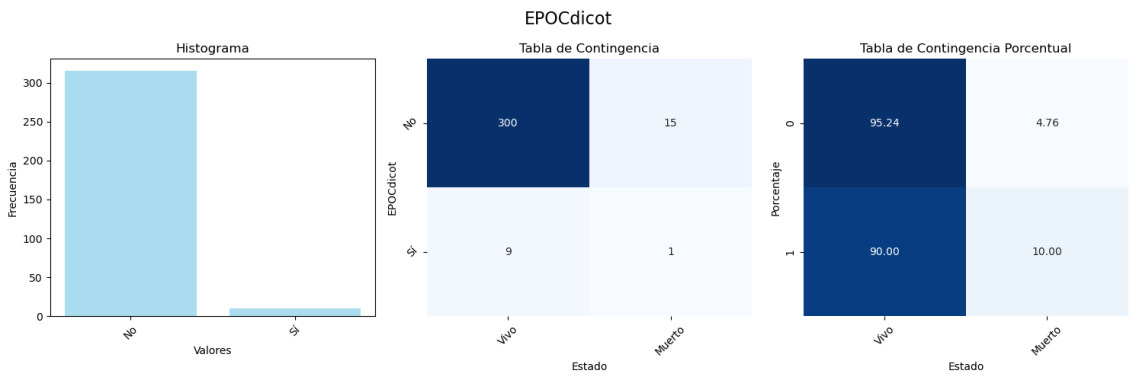


Ilustración 36. EDA EPOCdicot (Elaboración Propia)

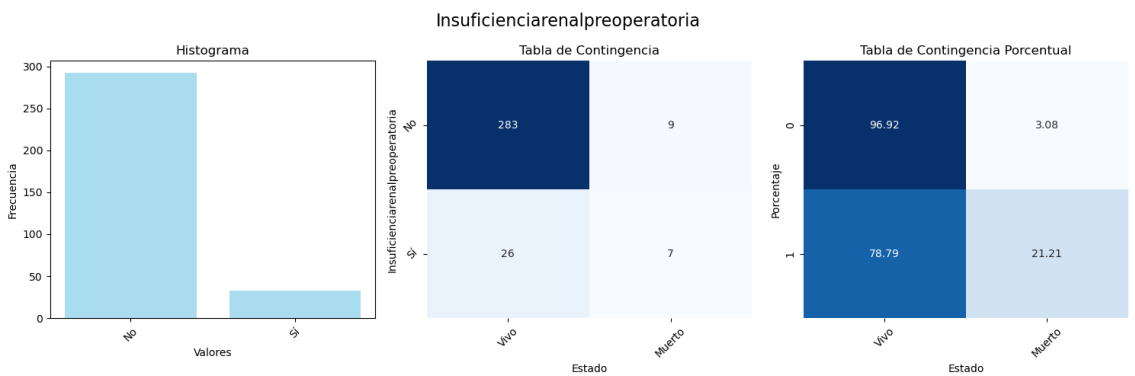


Ilustración 37. EDA Insuficienciarenalpreoperatoria (Elaboración Propia)

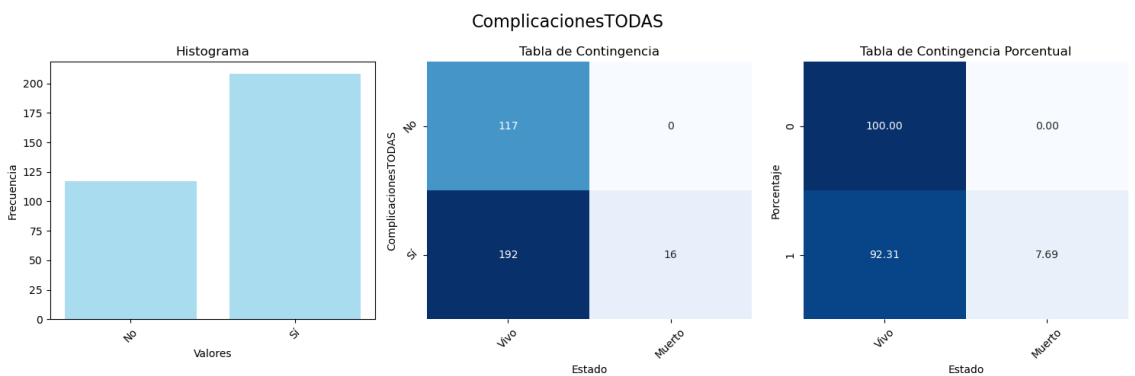


Ilustración 38. EDA ComplicacionesTODAS (Elaboración Propia)

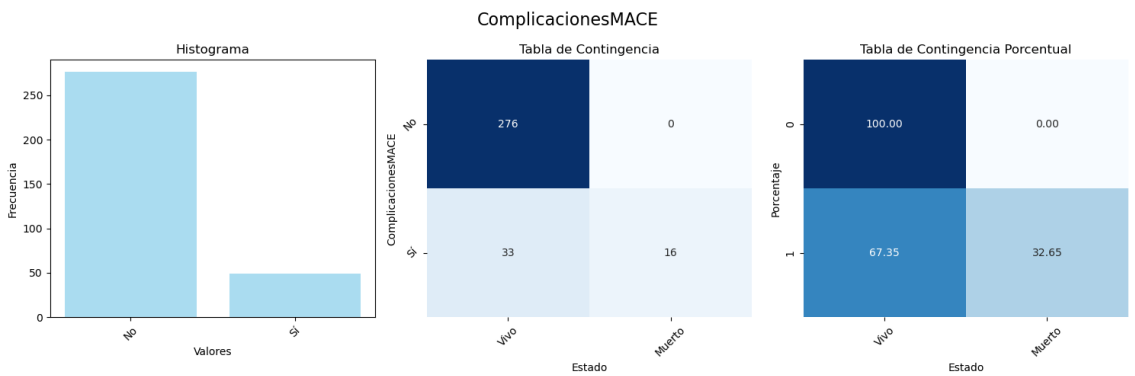


Ilustración 39. EDA ComplicacionesMACE (Elaboración Propia)

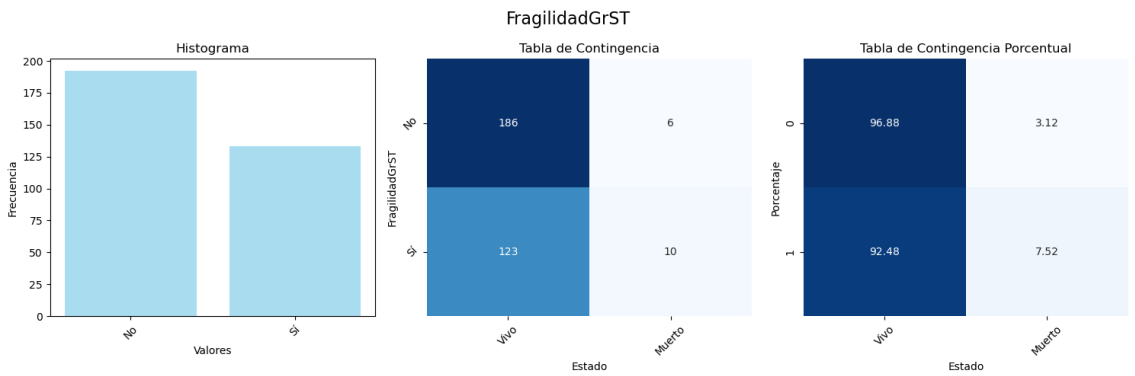


Ilustración 40. EDA FragilidadGrST (Elaboración Propia)

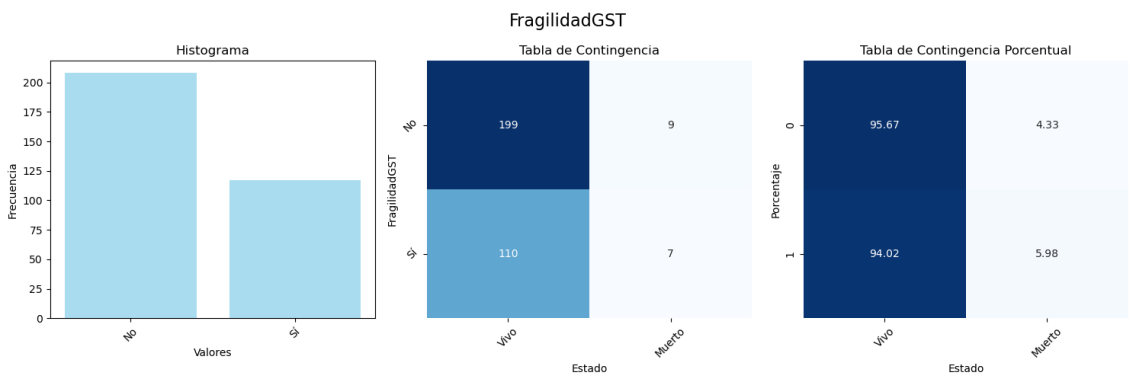


Ilustración 41. FragilidadGST (Elaboración Propia)

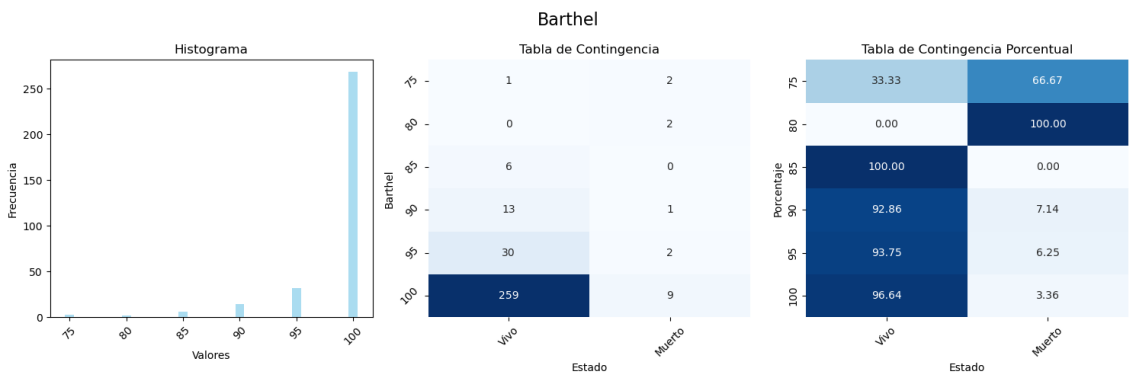


Ilustración 42. EDA Barthel (Elaboración Propia)

Se observa un mayor índice (porcentual) de “Muerto” en clases del índice “Barthel” inferiores o iguales a 80 (valor no representativo por el bajo número de muestras en dicha clase)

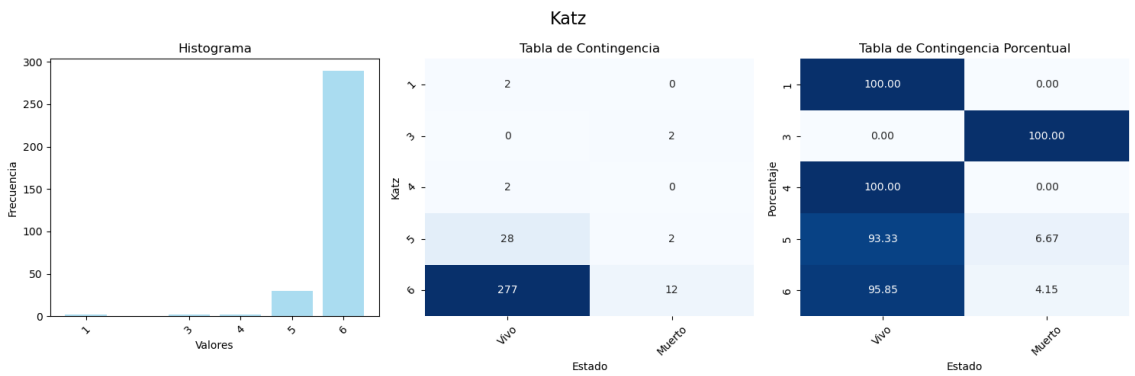


Ilustración 43. EDA Katz (Elaboración Propia)

Se observa un mayor índice (porcentual) de “Muerto” en clases del índice “Katz” igual a 3 (valor no representativo por el bajo número de muestras en dicha clase)

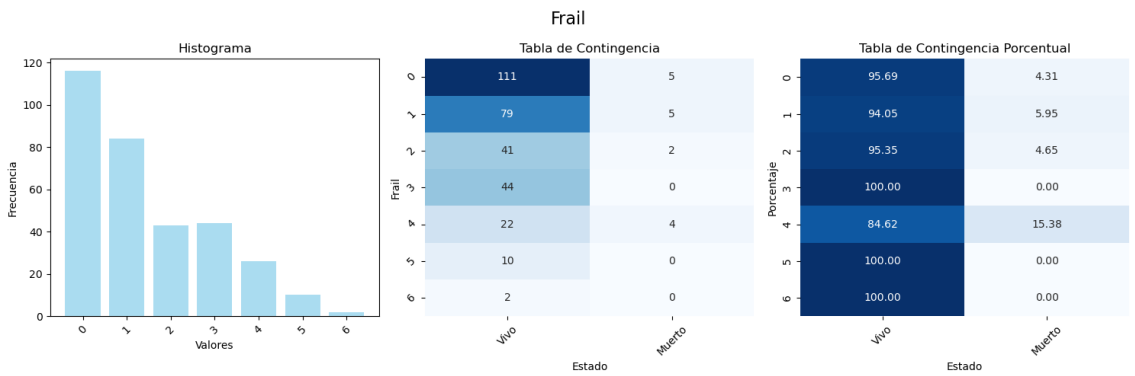


Ilustración 44. EDA Frail (Elaboración Propia)

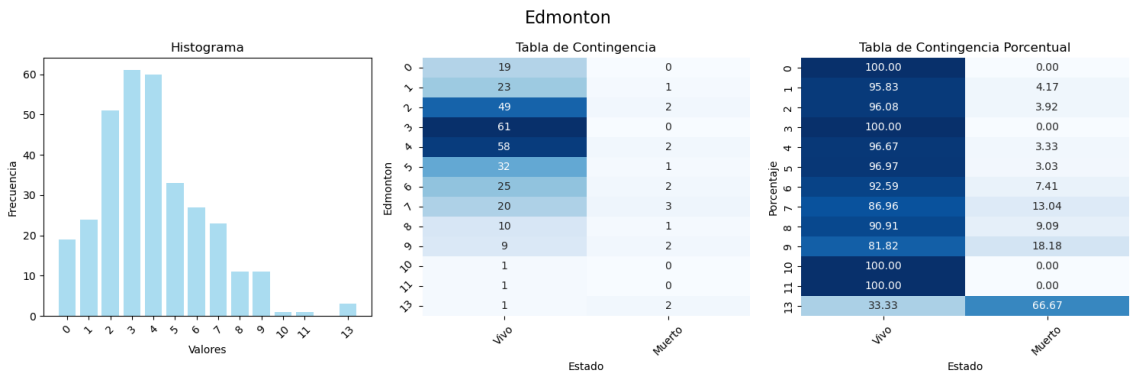


Ilustración 45. EDA Edmonton (Elaboración Propia)

Se observa un mayor índice (porcentual) de “Muerto” en clases del índice “Endmonton” igual a 13 (valor no representativo por el bajo número de muestras en dicha clase)

CONCLUSIONES

- Observamos que todas las variables categóricas son binarias excepto “Barthel”, “Katz”, “Frail” y “Endmonton”, algo que era de esperar ya que estas son las únicas variables relativas a índices médicos
- En todas las variables donde aparecen variables con un mayor porcentaje de “Muerto” en ciertas clases, no se pueden considerar como hipótesis de comportamiento de los datos, ya que el número absoluto de muestras es muy pequeño (<3)

3.6.2. ANÁLISIS DE HOMOCEDASTICIDAD

3.6.2.1 TEST DE LEVENE

Si $p\text{-valor} > 0,05$, no se rechaza la hipótesis nula \Rightarrow Hay Homocedasticidad

Si $p\text{-valor} < 0,05$, se rechaza la hipótesis nulas \Rightarrow Hay Heterocedasticidad

Variable Categórica	$p\text{-valor}$	Resultado
<i>Género</i>	$0,8747 > 0,05$	Homocedasticidad
<i>DMdicotomica</i>	$0,5778 > 0,05$	Homocedasticidad
<i>TabaquismoDicto</i>	$0,5512 > 0,05$	Homocedasticidad
<i>HTA</i>	$0,4744 > 0,05$	Homocedasticidad
<i>DL</i>	$0,7853 > 0,05$	Homocedasticidad
<i>Obesidad</i>	$0,3520 > 0,05$	Homocedasticidad
<i>EPOCdicto</i>	$0,4525 > 0,05$	Homocedasticidad
<i>Insuficientciarenalpreoperatoria</i>	$0,000003821 < 0,05$	Heterocedasticidad
<i>ComplicacionesTODAS</i>	$0,00201 < 0,05$	Heterocedasticidad
<i>ComplicacionesMACE</i>	$0,1688 > 0,05$	Homocedasticidad
<i>FragilidadGST</i>	$0,8547 > 0,05$	Homocedasticidad
<i>FragilidadGrST</i>	$0,5092 > 0,05$	Homocedasticidad
<i>Barthel</i>	$0,00000001049 < 0,05$	Heterocedasticidad
<i>Katz</i>	$0,0619 > 0,05$	Homocedasticidad
<i>Frail</i>	$0,8198 > 0,05$	Homocedasticidad
<i>Edmonton</i>	$0,067 > 0,05$	Homocedasticidad

Tabla 16. Test de Levene Variables Categóricas (Elaboración Propia)

3.7. ANÁLISIS ESTADÍSTICO VARIABLE

OBJETIVO

3.7.1. ANÁLISIS EXPLORATORIO (EDA)

3.7.1.1 HISTOGRAMA

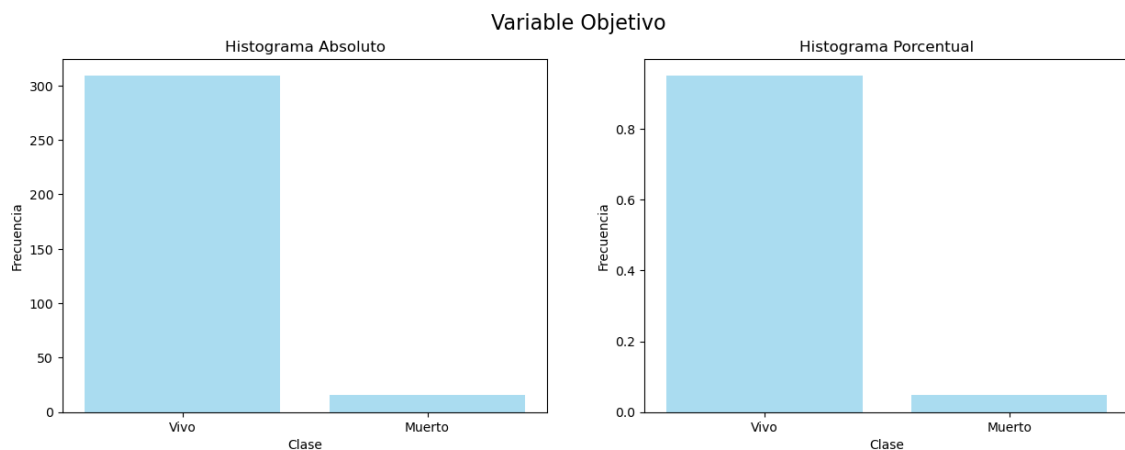


Ilustración 46. EDA Variable Objetivo (Elaboración Propia)

Se observa un desbalanceo muy alto en las clases de la variable a predecir, 95%-5%, lo que indica la necesidad de implementar técnicas de *over-sampling* o *under-sampling* para equilibrar la variable.

Debido al pequeño número de muestras del dataset (325 registros) se descarta la utilización de técnicas de *under-sampling* y se apuesta por la implementación de *over-sampling*, técnicas estadísticas para el sobre muestreo de minorías.

Como técnica de *over-sampling*, se utilizará SMOTE (Synthetic Minority Over-sampling Technique), una técnica avanzada de over-sampling que toma el conjunto de datos como una entrada, pero solo aumenta el porcentaje de los casos minoritarios a través de los datos de sus “vecinos más próximos”.

3.8. SELECCIÓN DE CARACTERÍSTICAS

3.8.1. PRE-SELECCIÓN CON RANDOM FOREST

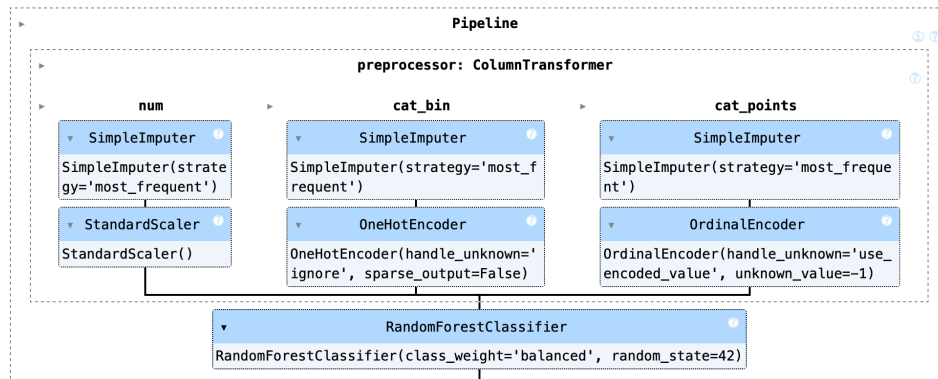


Ilustración 47. Pipeline RF pre-selección (Elaboración Propia)

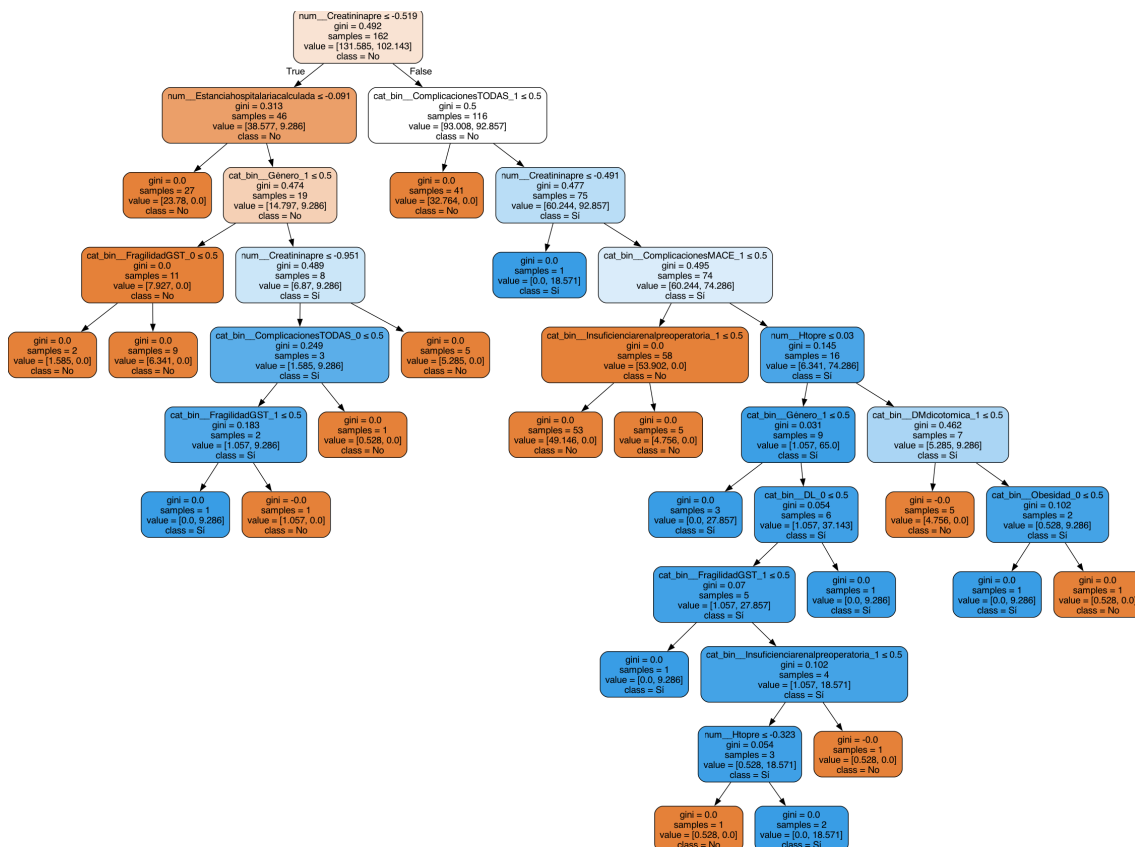


Ilustración 48. RF pre-selección (Elaboración Propia)

3.8.2. SELECCIÓN DEFINITIVA

Se seleccionan aquellas variables cuya importancia acumulada supera el 85% del total

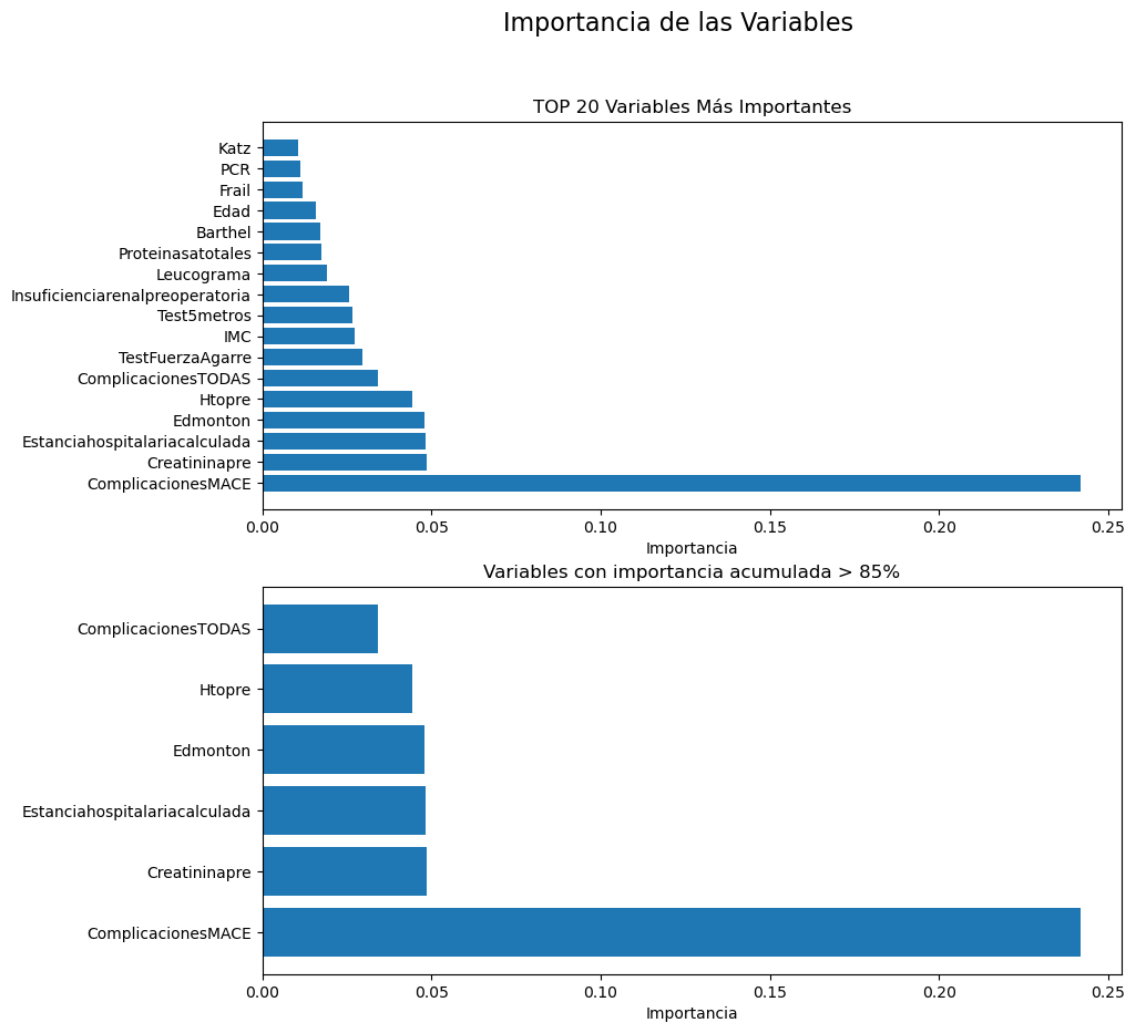


Ilustración 49. Comparativa importancia de variables (Elaboración Propia)

4. SELECCIÓN DEL MODELO

4.1. INTRODUCCIÓN

El objetivo del apartado anexo es el de seleccionar y optimizar el modelo de ML óptimo para las variables previamente definidas del *dataset* de estudio

4.2. CONCEPOS PREVIOS

4.2.1. SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) es una técnica para balancear datasets desbalanceados en problemas de clasificación. En lugar de duplicar ejemplos minoritarios, sintetiza nuevos datos creando puntos interpolados entre muestras reales.

Algoritmo

Para cada punto minoritario x :

1º - Selecciona un vecino cercano x_{nn}

2º - Crea un nuevo punto $x_{new} = x + \lambda \cdot (x_{nn} - x)$ donde $\lambda \sim U(0,1)$

4.2.2. MODELOS

Se definen a continuación los modelos de ML empleados en el presente anexo

4.2.2.1 SVM

SVM (Support Vector Machine), o Máquinas de Soporte Vectorial, son un algoritmo de clasificación que busca el hiperplano óptimo que maximiza el margen entre clases. El hiperplano óptimo es el que aumenta más el margen entre las dos clases, es decir, la anchura máxima de la franja paralela al hiperplano, que no contiene puntos de datos en su interior.

Es ideal para la clasificación binaria con buenos márgenes de separación.

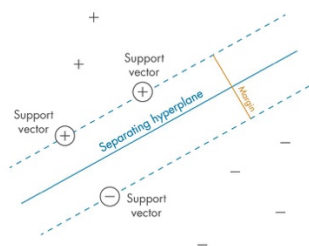


Ilustración 50. Modelo SVM (Fuente: Matlab MathWorks)

4.2.2.2 XGBOOST

XGBoost (Extreme Gradient Boost) es un algoritmo de ML basado en árboles de decisión que construye un conjunto de modelos débiles (normalmente árboles pequeños) de forma secuencial. En cada iteración, el nuevo árbol intenta corregir los errores cometidos por los anteriores, minimizando una función de pérdida a través de gradiente descendiente.

Lo hace optimizando una función objetivo que combina la pérdida del modelo y un término de regularización que penaliza la complejidad del modelo, lo que lo hace más robusto frente al sobreajuste.

Cada nuevo árbol se ajusta a los residuos negativos del gradiente, y se suma al modelo final. *XGBoost* incluye mejoras de rendimiento como poda de árboles, paralelismo, manejo de valores perdidos y regularización L1/L2, lo que lo hace muy eficiente y preciso en tareas de clasificación y regresión.

4.2.2.3 RANDOM FOREST

Random Forest es un algoritmo de *ensembles* basado en árboles de decisión. Construye múltiples árboles de forma aleatoria y combina sus predicciones para mejorar la precisión y reducir el sobreajuste. Cada árbol se entrena con un subconjunto aleatorio de características en lugar de usar todas.

La predicción final se obtiene por votación mayoritaria en clasificación o promedio en regresión. Este enfoque introduce diversidad entre árboles, lo que reduce la varianza del modelo y mejora su capacidad de generalización. Es robusto, fácil de implementar y eficaz en muchos tipos de datos.

4.2.2.4 CATBOOST

CatBoost (Categorical Boosting) es un algoritmo de *gradient boosting* optimizado para manejar variables categóricas de forma nativa, sin necesidad de transformarlas manualmente (como con *one-hot encoding*)

Utiliza árboles de decisión simétricos y aplica una técnica llamada ordenamiento por permutación, que convierte variables categóricas en números basándose en estadísticas calculadas de forma ordenada para evitar el *target leakage*.

Al igual que otros métodos de *boosting*, construye modelos secuenciales donde cada nuevo árbol corrige los errores del anterior, minimizando una función de pérdida.

Es eficiente, resistente al sobreajuste y especialmente útil cuando se trabaja con datos tabulares con muchas columnas categóricas.

4.2.2.5 KNN

KNN (K-Nearest Neighbors) es un algoritmo de clasificación y regresión basado en instancias. No entrena un modelo explícito, sino que almacena los datos de entrenamiento y toma decisiones comparando nuevas muestras con las ya vistas.

Algoritmo

1º - Calcula la distancia euclídea a todos los puntos de entrenamiento

$$d(x_i, x_j) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2}$$

2º - Selecciona los k vecinos más cercanos

3º - Asigna la clase más común entre esos k vecinos (votación)

4.2.2.6 MLP

MLP (Multilayer Perceptron) es una red neuronal artificial compuesta por capas de neuronas conectadas de forma totalmente densa. Cada neurona realiza una transformación lineal seguida de una función no lineal:

$$z = W \cdot x + b \quad a = \phi(z)$$

Donde ϕ es una función de activación como *ReLU* o sigmoide.

El entrenamiento se realiza mediante *backpropagation* y descenso de gradiente, minimizando una función de pérdida como entropía cruzada o MSE.

El MLP es capaz de aprender relaciones no lineales complejas, aunque requiere de un buen ajuste de hiperparámetros y normalización de los datos. Es la base de redes más profundas como las *Deep Neuronal Networks*.

4.2.3. VOTING CLASSIFIER

Voting Classifier es un método de ensamblado que combina las predicciones de varios modelos base para mejorar la precisión. Funciona de dos maneras:

- *Voting Hard* (Votación Dura), la clase predicha es la que recibe más votos (mayoría) entre los clasificadores.
- *Voting Soft* (Votación Suave), combina las probabilidades por cada modelo y elige la clase con mayor probabilidad promedio.

Se utiliza para aprovechar la diversidad de modelos y reducir errores, especialmente cuando los modelos base son diferentes

4.3. EQUILIBRADO DE LAS CLASES

Para equilibrar las clases de la variable objetivo, en todos los modelos seleccionados tras los escalados y codificaciones correspondientes, se aplicará la técnica de SMOTE para el equilibrado de las clases

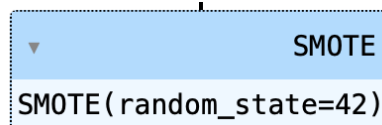


Ilustración 51. Pipeline SMOTE (Elaboración Propia)

4.4. PRE-SELECCIÓN DEL MODELO

Para hacer una preselección se prueban con valores de configuración estándar una serie de modelos (o mezcla de ellos) para su comparativa.

4.4.1. SVM

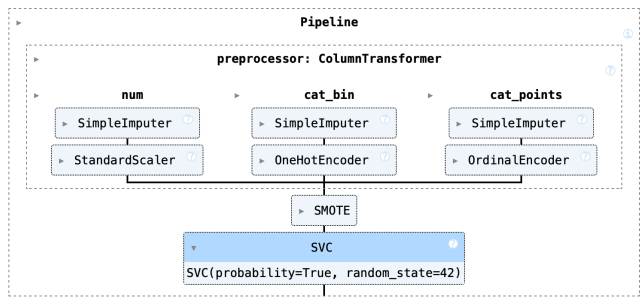


Ilustración 52. Pipeline SVM (Elaboración Propia)

*F1 Score: 0.523 ± 0.145
ROC AUC: 0.949 ± 0.015
PR AUC: 0.569 ± 0.148
Accuracy: 0.918 ± 0.008
Recall: 0.817 ± 0.185
Precision: 0.422 ± 0.163
MCC: 0.532 ± 0.12
F-beta Score: 0.455 ± 0.157
Brier Score: 0.057 ± 0.010
Log Loss: 0.222 ± 0.090
Hamming Loss: 0.082 ± 0.008*

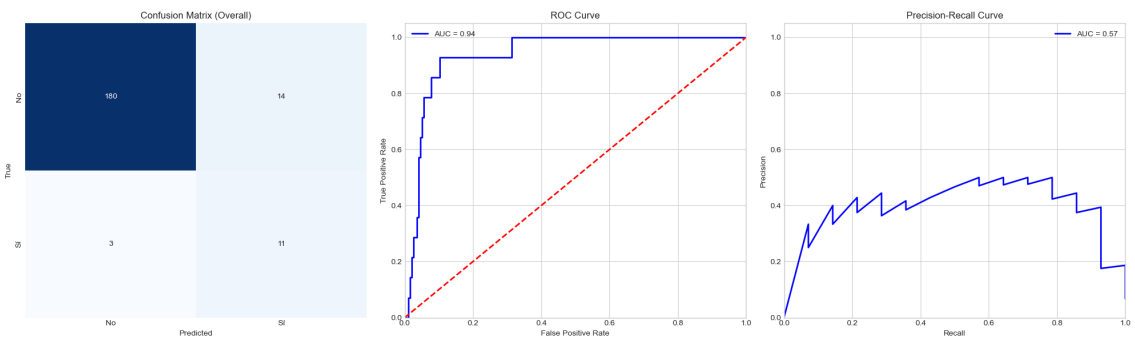


Ilustración 53. Rendimiento SVM

4.4.2. XGBOOST

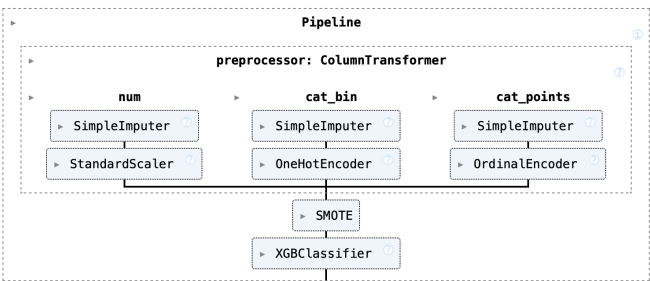


Ilustración 54. Pipeline XGBoost

*F1 Score: 0.501 ± 0.070
ROC AUC: 0.956 ± 0.035
PR AUC: 0.844 ± 0.260
Accuracy: 0.918 ± 0.025
Recall: 0.717 ± 0.218
Precision: 0.417 ± 0.083
MCC: 0.493 ± 0.08
F-beta Score: 0.443 ± 0.069
Brier Score: 0.059 ± 0.018
Log Loss: 0.204 ± 0.066
Hamming Loss: 0.082 ± 0.025*

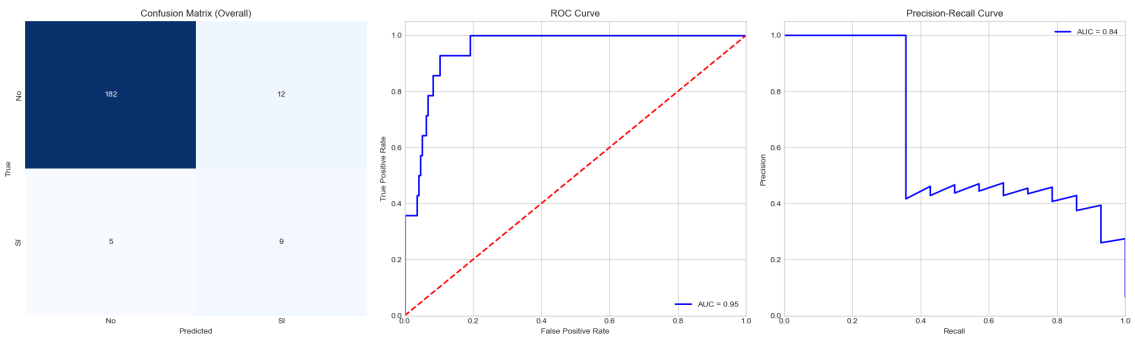


Ilustración 55. Rendimiento XGBoost (Elaboración Propia)

4.4.3. RANDOM FOREST

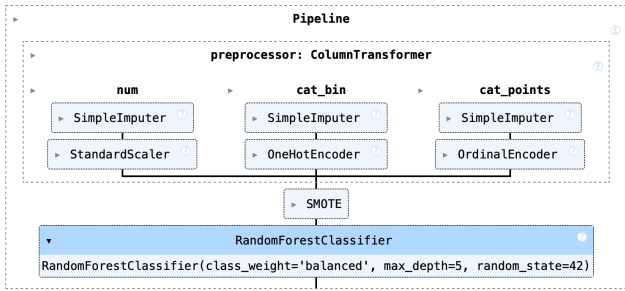


Ilustración 56. Pipeline RF (Elaboración Propia)

F1 Score: 0.425 ± 0.043
ROC AUC: 0.953 ± 0.030
PR AUC: 0.564 ± 0.240
Accuracy: 0.899 ± 0.025
Recall: 0.667 ± 0.216
Precision: 0.341 ± 0.075
MCC: 0.413 ± 0.06
F-beta Score: 0.368 ± 0.063
Brier Score: 0.061 ± 0.011
Log Loss: 0.178 ± 0.031
Hamming Loss: 0.101 ± 0.025

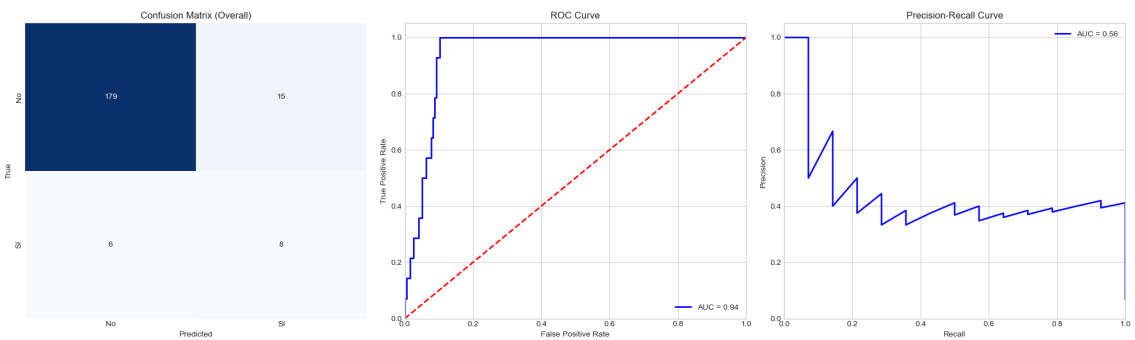


Ilustración 57. Rendimiento RF (Elaboración Propia)

4.4.4. CATBOOST

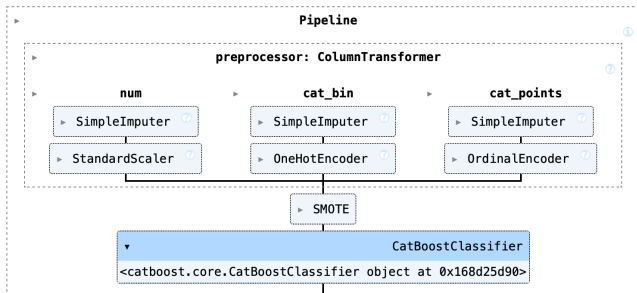


Ilustración 58. Pipeline Catboost

F1 Score: 0.497 ± 0.079
ROC AUC: 0.965 ± 0.024
PR AUC: 0.811 ± 0.223
Accuracy: 0.918 ± 0.016
Recall: 0.717 ± 0.166
Precision: 0.421 ± 0.137
MCC: 0.491 ± 0.05
F-beta Score: 0.445 ± 0.120
Brier Score: 0.063 ± 0.015
Log Loss: 0.248 ± 0.052
Hamming Loss: 0.082 ± 0.016

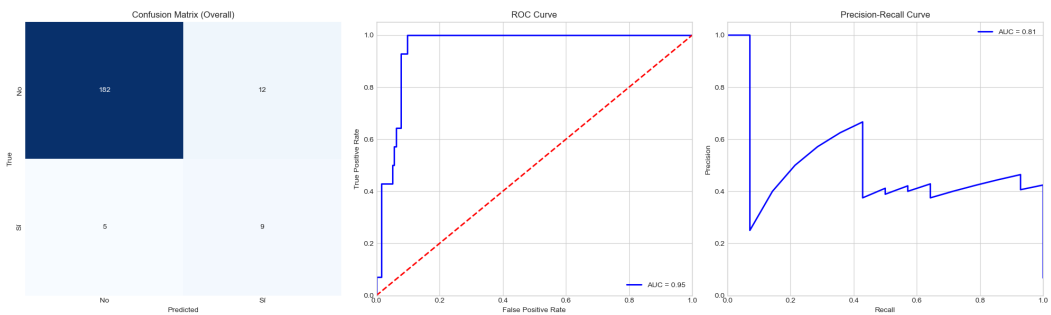


Ilustración 59. Rendimiento Catboost

4.4.5. KNN

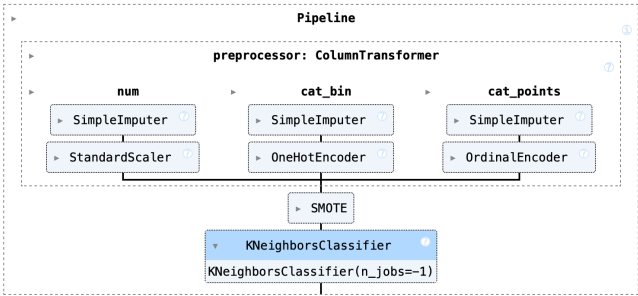


Ilustración 60. Pipeline KNN

F1 Score: 0.371 ± 0.089
ROC AUC: 0.909 ± 0.060
PR AUC: 0.474 ± 0.195
Accuracy: 0.889 ± 0.032
Recall: 0.583 ± 0.260
Precision: 0.298 ± 0.095
MCC: 0.348 ± 0.12
F-beta Score: 0.321 ± 0.090
Brier Score: 0.071 ± 0.015
Log Loss: 0.832 ± 0.461
Hamming Loss: 0.111 ± 0.032

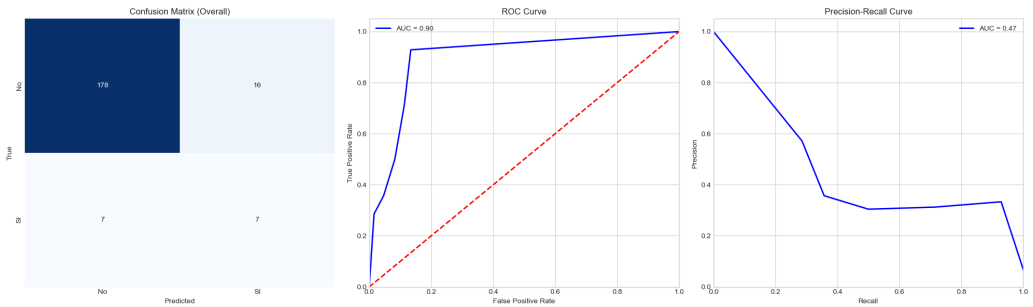


Ilustración 61. Rendimiento KNN

4.4.6. MLP

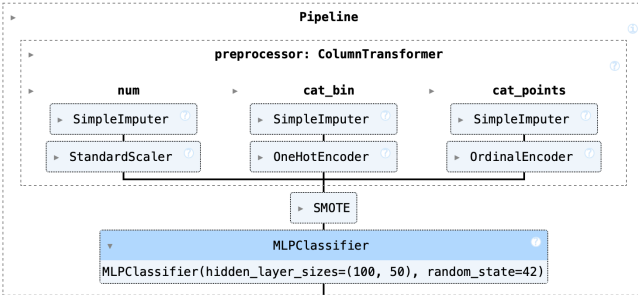


Ilustración 62. Pipeline MLP

F1 Score: 0.491 ± 0.203
ROC AUC: 0.949 ± 0.055
PR AUC: 0.810 ± 0.259
Accuracy: 0.923 ± 0.024
Recall: 0.683 ± 0.247
Precision: 0.450 ± 0.234
MCC: 0.488 ± 0.18
F-beta Score: 0.463 ± 0.225
Brier Score: 0.055 ± 0.012
Log Loss: 0.286 ± 0.117
Hamming Loss: 0.077 ± 0.024

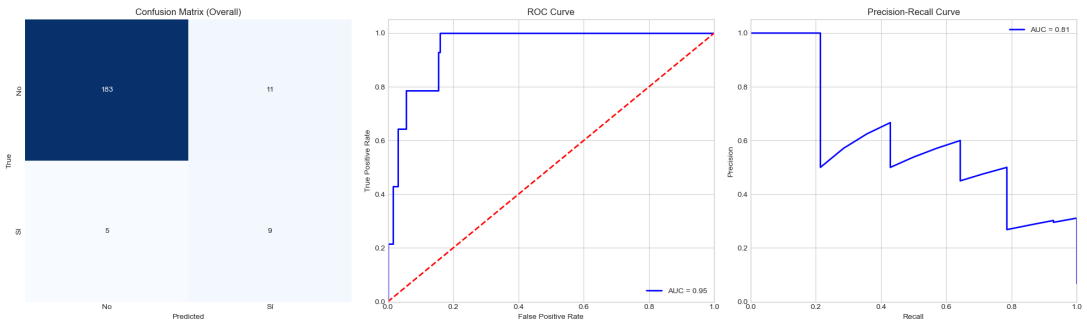


Ilustración 63. Rendimiento MLP

4.4.7. XGBOOST + CATBOOST (Voting Classifier)

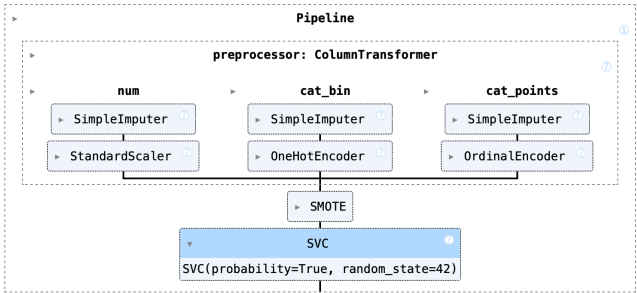


Ilustración 64. Pipeline XGBoost + Catboost (Elaboración Propia)

F1 Score: 0.447 ± 0.035
ROC AUC: 0.965 ± 0.024
PR AUC: 0.811 ± 0.230
Accuracy: 0.909 ± 0.021
Recall: 0.667 ± 0.216
Precision: 0.378 ± 0.095
MCC: 0.438 ± 0.03
F-beta Score: 0.399 ± 0.072
Brier Score: 0.059 ± 0.015
Log Loss: 0.196 ± 0.047
Hamming Loss: 0.091 ± 0.021

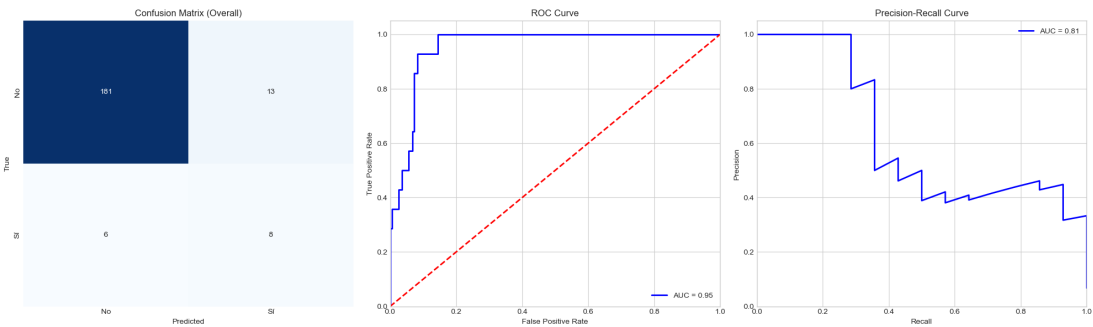


Ilustración 65. Rendimiento XGBoost + Catboost (Elaboración Propia)

4.4.8. RANDOM FOREST + CATBOOST (Voting Classifier)

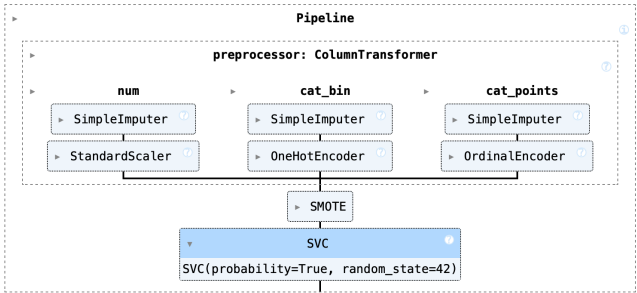


Ilustración 66. Pipeline RF + Catboost (Elaboración Propia)

F1 Score: 0.459 ± 0.053
ROC AUC: 0.965 ± 0.021
PR AUC: 0.761 ± 0.196
Accuracy: 0.913 ± 0.017
Recall: 0.667 ± 0.216
Precision: 0.396 ± 0.108
MCC: 0.451 ± 0.04
F-beta Score: 0.414 ± 0.088
Brier Score: 0.055 ± 0.012
Log Loss: 0.172 ± 0.041
Hamming Loss: 0.087 ± 0.017

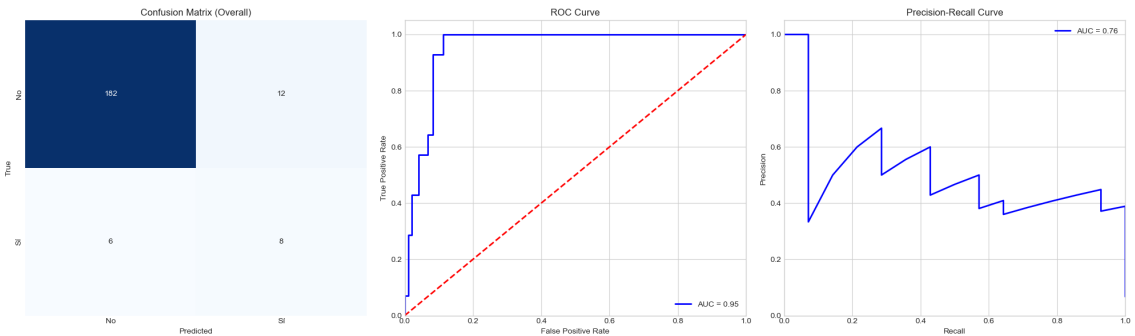


Ilustración 67. Rendimiento RF + Catboost

4.4.9. COMPARATIVA

MODELO	ROC AUC	PR AUC	ACCURACY	RECALL	PRECISION
<i>SVC</i>	0,94	0,57	0,918	0,817	0,422
<i>XGBoost</i>	0,95	0,84	0,918	0,717	0,417
<i>Random Forest</i>	0,94	0,56	0,899	0,667	0,341
<i>CatBoost</i>	0,95	0,81	0,918	0,717	0,421
<i>KNN</i>	0,90	0,47	0,889	0,583	0,298
<i>MLP</i>	0,95	0,81	0,923	0,683	0,450
<i>XGBoost + CatBoost</i>	0,95	0,81	0,909	0,667	0,378
<i>RF + Catboost</i>	0,95	0,76	0,913	0,667	0,396

Tabla 17. Comparativa modelos (Elaboración Propia)

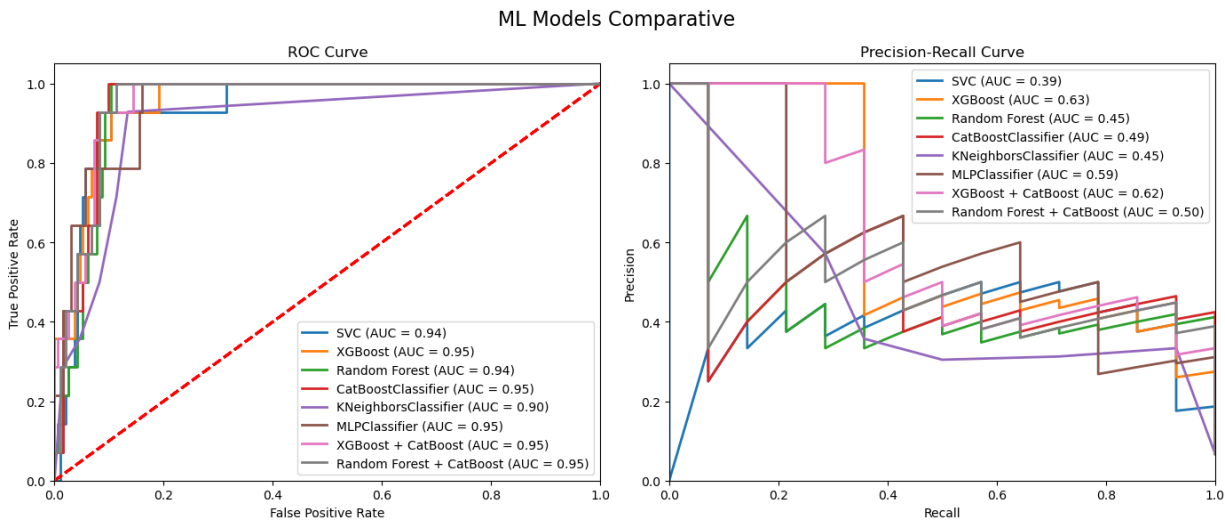


Ilustración 68. Comparativa Modelos(Elaboración Propia)

MODELO	F1	MCC	F-beta	BRIER	LOG LOSS	HAMMING LOSS
<i>SVC</i>	0,523	0,532	0,455	0,057	0,222	0,082
<i>XGBoost</i>	0,501	0,493	0,443	0,059	0,204	0,082
<i>Random Forest</i>	0,425	0,413	0,368	0,061	0,178	0,101
<i>CatBoost</i>	0,497	0,491	0,445	0,063	0,248	0,082
<i>KNN</i>	0,371	0,348	0,321	0,071	0,832	0,111
<i>MLP</i>	0,491	0,488	0,463	0,055	0,286	0,077
<i>XGBoost + CatBoost</i>	0,447	0,438	0,399	0,059	0,196	0,091
<i>RF + Catboost</i>	0,459	0,451	0,414	0,055	0,172	0,087

Tabla 18. Métricas modelos (Elaboración Propia)

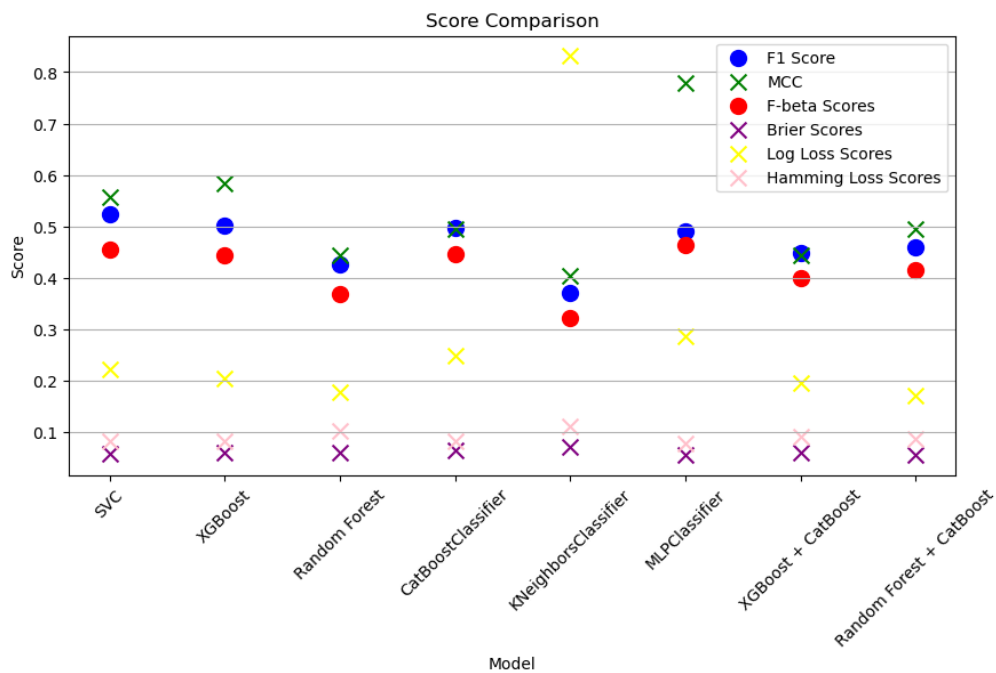


Ilustración 69. Métricas modelos (Elaboración Propia)

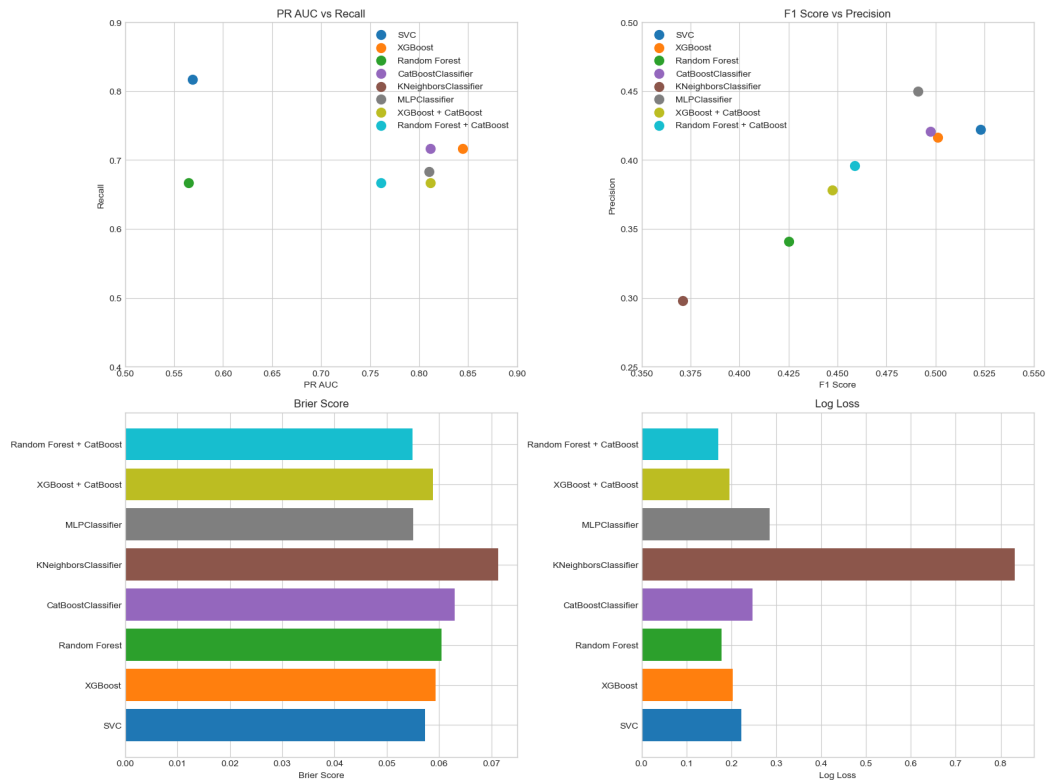


Ilustración 70. Comparativa métricas modelos (Elaboración Propia)

PR-AUC vs Recall

XGBoost, CatBoost y MLP logran un gran equilibrio entre *recall* (sensibilidad) y *PR-AUC*

SVC logra el recall más alto (0,817), lo cual clínicamente es importante para minimizar falsos negativos, aunque con un PR-AUC moderado

F1 Score vs Precision

SCV destaca con un F1 superior, combinado con un buen precisión-*recall*.

KNN muestra el peor desempeño, con baja precisión y F1, lo que sugiere una alta tasas de falsos positivos.

Comparación Brier Score

MLP y RF + Catboost presentan los *Brier Score* más bajos, lo que indica una buena calibración de probabilidad, algo fundamental en decisiones clínicas basadas en información precisa y relevante sobre un paciente.

De nuevo, KNN queda muy por detrás, con un *Brier Score* elevado.

Comparación Log Loss

RF + Catboost y RF tienen los mejores log los, los más bajos, lo que indica alta confiabilidad probabilística.

KNN destaca por su mal rendimiento, con el log los más alto (0,832) desaconsejando por completo su uso.

4.4.10. SELECCIÓN

Para seleccionar el modelo óptimo para la predicción de la 30DM se plantean dos premisas que condicionan, junto con las métricas obtenidas la determinación del modelo óptimo:

- Evitar falsos negativos, es decir, no identificar a un paciente que sí está en riesgo de morir, ya que es el caso más crítico
- Es aceptable tener más falsos positivos (pacientes identificados con riesgo de muerte que luego no lo hacen), permitiendo así una atención preventiva sin consecuencias graves

Con estas premisas, observamos que lo prioritario es maximizar el *recall* y el PR-AUC (reducir falsos positivos)

Con ello concluimos que **la mejor opción es el *XGBoost***, debido a que:

- *PR-AUC* más alto, de 0,84
- Buen *recall* (0,71), por encima de la media de los modelos analizados y un valor suficientemente alto para el modelo
- *Brier Score* y *Log Loss* bajos, lo que indica una buena calibración de probabilidad
- No tiene el F1 más alto, pero en este caso se prioriza el *recall*

4.5. OPTIMIZACIÓN MODELO SELECCIONADO

Para seleccionar los hiperparámetros óptimos del modelo seleccionado deberemos de optimizar el modelo contra diferentes métricas en función del resultado buscado.

- Optimización del *recall*, si lo más importante es detectar casi todos los pacientes en riesgo (buscar *recall* alto), algo mejor para intervenciones clínicas urgentes
- Optimización de PR-AUC, si se busca un mejor equilibrio entre *recall* y precisión en general, es decir, un modelo más estable

Por ello se opta por hacer un modelo ponderado por importancia que asigne un 60% al *recall* y un 40% al PR-AUC obteniendo los siguientes resultados

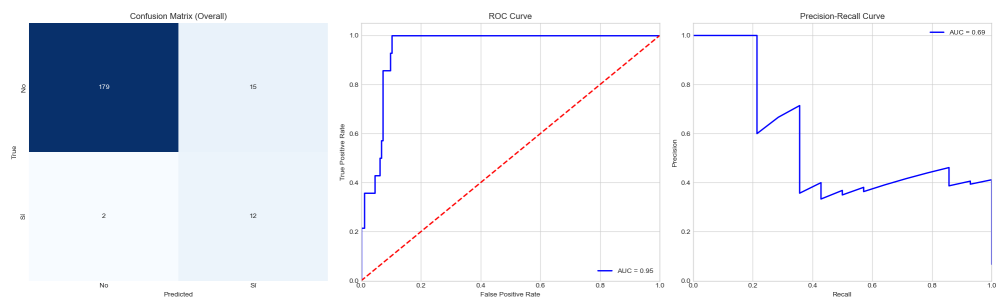


Ilustración 71. Rendimiento XGBoost (Elaboración Propia)

MÉTRICA	VALOR	INTERPRETACIÓN
ROC AUC	0.964 ± 0.021	Muy buena separabilidad del modelo
PR AUC	0,694 ± 0,196	Buen área bajo la curva precisión-recall, mejorando respecto a otros modelos.
ACCURACY	0,918 ± 0,008	Alto, pero menos importante en desequilibrio
RECALL	0,900 ±0,173	Muy alto, detecta el 90% de los pacientes que morirán, la prioridad clínica
PRECISION	0,421 ± 0,135	Aceptable y coherente con la premisa de optimización del <i>recall</i>
F1 SCORE	0,548 ±0,138	Balance entre precisión y <i>recall</i>
F-beta	0,462 ± 0,137	Penaliza más falsos positivos
BRIER SCORE	0,082 ±0,007	Buena calibración de probabilidades
LOG LOSS	0,315 ± 0,016	Bajo, buena confianza en las predicciones
HAMMING LOS	0,082 ± 0,008	Baja tasa de error

Tabla 19. Métricas XGBoost

5. DESARROLLO APLICACIÓN

5.1. INTRODUCCIÓN

PreMoCir es una herramienta de ayuda a la decisión clínica que estima la probabilidad de mortalidad de un paciente sometido a cirugía cardíaca, basándose en variables clínicas preoperatorias y datos de evolución hospitalaria utilizando modelos de ML.

En este apartado se hará una revisión a los requerimientos de dicha aplicación, así como a la forma en la que se ha desarrollado y desplegado la misma.

5.2. REQUERIMIENTOS

Los requerimientos para el desarrollo de la aplicación se separan en requerimientos funcionales, no funcionales y éticos-legales. Para cada requerimiento se definirá:

- ID – Identificación única para cada requisito
- Descripción – Definición del requisito
- Importancia – Impacto del requisito en el sistema
- Tipo – Indicando si es requisito obligatorio (*Must*) o recomendado (*Should*)
- Prioridad – Urgente/Alta/Media/Baja, para planificar el desarrollo

5.2.1. REQUERIMIENTOS FUNCIONALES

RF1 - Registro de entrada de datos clínicos	
Descripción	Permitir al usuario ingresar datos relevantes del paciente para la predicción
Importancia	Crítica
Tipo	<i>Must</i>
Prioridad	Alta
RF2 – Validación de los datos de entrada	
Descripción	Verificar que los datos introducidos sean correctos y completos
Importancia	Alta
Tipo	<i>Must</i>
Prioridad	Alta

RF3 – Procesamiento y predicción

Descripción	Aplicar el modelo de ML para calcular la probabilidad de mortalidad
Importancia	Crítica
Tipo	<i>Must</i>
Prioridad	Alta

RF4 – Visualización de los resultados

Descripción	Mostrar al usuario la predicción de manera clara y comprensible
Importancia	Alta
Tipo	<i>Must</i>
Prioridad	Alta

RF5 – Explicabilidad del modelo

Descripción	Mostrar al usuario que variables influyeron en la predicción
Importancia	Crítica
Tipo	<i>Should</i>
Prioridad	Alta

RF6 – Interfaz de usuario

Descripción	Proporcionar una interfaz accesible e intuitiva desde el navegador
Importancia	Crítica
Tipo	<i>Must</i>
Prioridad	Alta

RF7 – Exportación de resultados

Descripción	Permitir descargar la predicción y sus detalles en formato PDF o CSV
Importancia	Media
Tipo	<i>Should</i>
Prioridad	Media

RF8 – Módulo de ayuda/guía

Descripción	Incluir una guía para explicar el uso e interpretación de la aplicación
Importancia	Media
Tipo	<i>Should</i>
Prioridad	Baja

5.2.2. REQUERIMIENTOS NO FUNCIONALES

RNF1 – Disponibilidad

Descripción	El sistema debe estar disponible al menos el 95% del tiempo
Importancia	Alta
Tipo	<i>Must</i>
Prioridad	Alta

RNF2 – Rendimiento

Descripción	El tiempo de respuesta del sistema no debe superar los 3 segundos
--------------------	---

Importancia	Alta
--------------------	------

Tipo	<i>Must</i>
-------------	-------------

Prioridad	Alta
------------------	------

RNF3 – Escalabilidad

Descripción	La aplicación debe ser fácilmente ampliable con nuevos modelos o variables
--------------------	--

Importancia	Media
--------------------	-------

Tipo	<i>Should</i>
-------------	---------------

Prioridad	Media
------------------	-------

RNF4 – Seguridad

Descripción	Los datos deben de estar protegidos, con protocolos seguros y sin almacenamiento indebido
--------------------	---

Importancia	Crítica
--------------------	---------

Tipo	<i>Must</i>
-------------	-------------

Prioridad	Urgente
------------------	---------

RNF5 – Usabilidad

Descripción	La interfaz debe ser intuitiva y comprensible para usuarios no técnicos
--------------------	---

Importancia	Alta
--------------------	------

Tipo	<i>Must</i>
-------------	-------------

Prioridad	Alta
------------------	------

RNF6 – Portabilidad

Descripción	Debe de funcionar correctamente en distintos sistemas operativos y navegadores
--------------------	--

Importancia	Media
--------------------	-------

Tipo	<i>Should</i>
-------------	---------------

Prioridad	Media
------------------	-------

RNF7 – Mantenibilidad

Descripción	El código debe de estar documentado y organizado para facilitar su mantenimiento
--------------------	--

Importancia	Alta
--------------------	------

Tipo	<i>Must</i>
-------------	-------------

Prioridad	Media
------------------	-------

RNF8 – Trazabilidad

Descripción	Se debe poder reconstruir cada predicción a partir de las entradas, sin almacenar datos sensibles
--------------------	---

Importancia	Media
--------------------	-------

Tipo	<i>Should</i>
-------------	---------------

Prioridad	Baja
------------------	------

5.2.3. REQUERIMIENTOS ÉTICO-LEGALES

REL1 – Anonimización de datos

Descripción	No se debe permitir el ingreso o almacenamiento de datos personales identificables
Importancia	Crítica
Tipo	<i>Must</i>
Prioridad	Urgente

REL2 – Cumplimiento legal

Descripción	Cumplir con la normativa de protección de datos (RGPD)
Importancia	Crítica
Tipo	<i>Must</i>
Prioridad	Urgente

REL3 – Advertencia de uso

Descripción	Mostrar un aviso claro de que la herramienta es de apoyo y no sustituye la decisión médica
Importancia	Alta
Tipo	<i>Must</i>
Prioridad	Alta

REL4 – Auditoría de decisiones

Descripción	Permitir interpretar y justificar cómo se ha generado cada predicción
Importancia	Alta
Tipo	<i>Should</i>
Prioridad	Media

5.3. TECNOLOGÍAS

Para el desarrollo de la aplicación se ha hecho uso de una serie de herramientas tanto para el desarrollo como para el despliegue

5.3.1. PYTHON

Python es un lenguaje de programación de alto nivel de programación interpretado con una filosofía que hace gran hincapié en la legibilidad del código. Soporta orientación a objetos, programación imperativa y programación funcional



Ilustración 72. Logo Python (Fuente Python Organization)

Se ha utilizado tanto como lenguaje de programación para el análisis de los datos, para la evaluación y selección del modelo así como en el desarrollo de la propia aplicación.

5.3.2. STREAMLIT

Streamlit es un *framework* de *Python* de código abierto que facilita la creación y despliegue de aplicaciones web interactivas para proyectos de ciencia de datos y ML. Permite transformar análisis de datos y modelos de aprendizaje en aplicaciones web de una forma rápida y sencilla, sin necesidad de conocimientos avanzados de desarrollo web.



Ilustración 73. Logo Streamlit (Fuente Streamlit Oficial)

Se ha utilizado tanto para el desarrollo de la aplicación web como para el despliegue de esta a través de su servicio de *cloud*.

5.3.3. SCIKIT-LEARN

Scikit-learn es una biblioteca de código abierto para *Python* utilizada en ML que proporciona una amplia gama de algoritmos y herramientas para preprocesamiento de datos, análisis predictivo, reducción de dimensionalidad...



Ilustración 74. Logo Scikit-learn (Fuente SKLearn)

Se ha utilizado principalmente para la selección, comparativa, evaluación y optimización del modelo de ML empleado en la aplicación.

5.4. MÓDULOS Y ESTRUCTURA

La aplicación está estructurada en de la siguiente manera:

- *main.py* – Punto de entrada de la aplicación
- *sidebar.py* – Lógica del menú lateral
- *footer.py* – Pie de página
- *model_utils.py* – Funciones asociadas a la carga del modelo y funciones auxiliares
- *prediction.py* – Lógica de la predicción

5.5. DESPLIEGUE

El despliegue de la aplicación se ha hecho a través de la solución *cloud* propia de *Streamlit* y puede accederse a la misma través del siguiente enlace o *QR*.

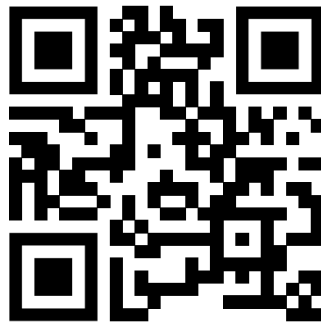


Ilustración 75. QR PreMoCir (Elaboración Propia)

<https://premocir.streamlit.app>

5.6. REPOSITORIO

El código completo de la aplicación puede encontrarse de manera pública en el siguiente repositorio de GitHub.

<https://github.com/telmomm/PreMoCir>

5.7. MANUAL DE USUARIO

El Manual de Usuario está integrado dentro de la propia aplicación y es accesible a través del menú lateral de la misma, en el que se indica al usuario como realizar de manera correcta las predicción en dos únicos pasos.

1º - Introducción de datos

Introduce los siguientes valores en los campos visibles:

- Hematocrito preoperatorio [%]
- Creatinina preoperatoria [mg/dL]
- Fragilidad Edmonton [0-17]
- Fecha de ingreso hospitalario
- Complicaciones MACE (Si el paciente ha tenido complicaciones cardiovasculares graves)
- Complicaciones TODAS (Si el paciente ha tenido cualquier tipo de complicación médica relevante)

2º - Realizar predicción

Pulsar el botón “Realizar predicción”

La aplicación mostrará la probabilidad de mortalidad estimada en porcentaje para los datos suministrados.

The screenshot displays the 'Predicción de la Mortalidad en Cirugía Cardíaca' interface. On the left is a sidebar menu with options: '¿Qué es PreMoCir?', 'Manual de Usuario', 'Advertencias', and 'GitHub'. The main area contains input fields for 'Hematocrito preoperatorio (%)' (40,00), 'Creatinina preoperatoria (mg/dL)' (1,00), and a slider for 'Fragilidad Edmonton' (set to 7). There are also fields for 'Fecha de ingreso' (2025/05/15) and 'Días hospitalizado: 10'. Two checkboxes are present: 'Complicaciones MACE' (checked) and 'Complicaciones Todas' (unchecked). A red 'Realizar predicción' button is at the bottom left of the form. Below the button, a green box displays 'Probabilidad Mortalidad: 79.7887%'. At the bottom of the interface, a footer reads 'Desarrollado por Telmo Miguel Medina | © 2025'.

Ilustración 76. PreMoCir (Elaboración Propia)

5.8. LICENCIA

La aplicación desarrollada se distribuye bajo Licencia *MIT* (*Massachusetts Institute of Technology*) la cual es una licencia de software de código abierto que permite a los usuarios:

- Libertad de Uso – Pueden usar, copiar, modificar, fusionar, distribuir, sublicencias y vender el software.
- Sin Garantía – El *software* se proporciona “tal cual” sin garantía ni responsabilidad por parte del titular de los derechos de autor o contribuyentes.
- Inclusión de Aviso de Derechos de Autor: Cualquier redistribución del software o de su parte debe de incluir un aviso de derechos de autor y la licencia original.
- Sin Restricciones Adicionales – No impone restricciones adicionales sobre el software distribuido basado en la licencia *MIT*
- Compatibilidad – Es compatible con torso proyectos y permite la combinación con software bajo diferentes licencias.
- Simplicidad – Es una de las licencias más simples y permisivas, facilitando su comprensión y aplicación.

En la aplicación desarrollada se puede consultar toda la información relativa a la licencia de esta, en el apartado “Licencia” de la documentación.

5.9. ADVERTENCIAS Y CONDICIONES DE USO

La aplicación desarrollada se trata de una herramienta de apoyo y no sustituye el criterio clínico profesional.

Los datos procesados no se almacenan ni se transmiten, garantizando privacidad y anonimato.

El modelo es probabilístico y puede estar sujeto a errores inherentes al ML.

Los resultados deben de interpretarse en contexto clínico.

6. CONCLUSIONES

6.1. CONCLUSIONES

Las principales conclusiones del estudio y del desarrollo son:

- **Viabilidad del ML en el ámbito clínico**

El uso de algoritmos de machine learning ha demostrado ser altamente eficaz en la predicción de mortalidad a corto plazo en pacientes sometidos a cirugía cardíaca.

Modelos como *XGBoost* superan ampliamente a técnicas estadísticas tradicionales, ofreciendo un mejor ajuste, mayor sensibilidad y mejor capacidad de generalización (dentro de la muestra utilizada).

- **Importancia de la calidad y estructura de los datos**

La base de datos clínica utilizada (aunque limitada en tamaño) ha sido clave en el desarrollo del modelo. El cribado clínico previo, el tratamiento de valores atípicos, los test estadísticos y la validación cruzada fueron fundamentales para asegurar la fiabilidad del modelo.

- **Métricas y validación**

La combinación de métricas de discriminación (*AUC-ROC*), calibración (*Brier score*, *Log loss*) y balance (*F1-score*, *MCC*) ha permitido una evaluación integral del rendimiento del modelo, orientada no solo a la exactitud, sino también a la utilidad clínica.

- **Aplicación funcional y accesible**

La implementación de una aplicación web (*Streamlit*) demuestra la aplicabilidad del modelo más allá del entorno académico, permitiendo el uso práctico en tiempo real.

La interfaz es sencilla y orientada al usuario clínico, facilitando la interacción con la herramienta predictiva.

- **Limitaciones actuales**

El tamaño del dataset es reducido (325 registros), lo que limita la generalización.

No se realizó una validación externa multicéntrica, lo cual es esencial para garantizar la aplicabilidad en diferentes hospitales o regiones.

- **Relevancia para futuras investigaciones**

Este trabajo sienta una base sólida para líneas futuras, incluyendo validación en otros entornos, expansión de la base de datos, exploración de otros algoritmos, e integración en sistemas hospitalarios.

Además, abre la puerta al desarrollo de herramientas de predicción para otros eventos clínicos, como reingresos o complicaciones postoperatorias.

6.2. LINEAS FUTURAS

Se proponen las siguientes líneas de trabajo para mejorar el rendimiento de la aplicación de predicción de la mortalidad.

- **Validación externa multicéntrica**, aplicar el modelo en otros hospitales y poblaciones para garantizar su generalización
- **Aumento del tamaño del *dataset***, incorporar nuevos registros para mejorar la robustez estadística del modelo
- **Desarrollo de modelos explicables**, profundizar en métodos interpretables como SHAP o LIME, para tener una mayor transparencia clínica
- **Automatización en tiempo real**, integración con sistemas hospitalarios (HIS) para predicción continua durante el ingreso.
- **Predicción en otros eventos adversos**, ampliar el alcance del modelo para incluir complicaciones postoperatorias o reingresos.

7. REFERENCIAS

7.1. REFERENCIAS BIBLIOGRÁFICAS

- [1] Organización Mundial de la Salud, «Las diez causas principales de defunción», 7 de agosto de 2024. [En línea]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [2] P. Kamath, C. Nayak, Dr. N. Pai D, D. Noronha, y N. S S, «Utility of Global Registry of Acute Coronary Event Score (GRACE) and Thrombolysis in Myocardial Infarction Risk Score (TIMI Score) in predicting 30-day Mortality in patients undergoing primary angioplasty at a Tertiary care center – A record-based study», *F1000Research*, vol. 13, p. 272, abr. 2024, doi: 10.12688/f1000research.145283.1.
- [3] D. Schellings *et al.*, «Predictive value of NT-proBNP for 30-day mortality in patients with non-ST-elevation acute coronary syndromes: a comparison with the GRACE and TIMI risk scores», *Vasc. Health Risk Manag.*, vol. Volume 12, pp. 471-476, nov. 2016, doi: 10.2147/VHRM.S117204.
- [4] S. Kanani, I. Mahdavi, N. Ziaie, y B. Rahimpour Cami, «Prediction of 30 -day Mortality for Heart Failure Patients with Cardiogenic Shock Using Random Forest Hyperparameter Tuning», *J. Health Biomed. Inform.*, vol. 11, n.º 1, pp. 83-95, jun. 2024, doi: 10.34172/jhbmi.2024.15.
- [5] S. Sinha, T. Dong, A. Dimagli, A. Judge, y G. D. Angelini, «A machine learning algorithm-based risk prediction score for in-hospital/30-day mortality after adult cardiac surgery», *Eur. J. Cardiothorac. Surg.*, vol. 66, n.º 4, p. ezae368, oct. 2024, doi: 10.1093/ejcts/ezae368.
- [6] M. Á. Acosta-Benito y I. Martín-Lesende, «Fragilidad en atención primaria: diagnóstico y manejo multidisciplinar», *Aten. Primaria*, vol. 54, n.º 9, p. 102395, sep. 2022, doi: 10.1016/j.aprim.2022.102395.
- [7] S. A. M. Nashef *et al.*, «EuroSCORE II», *Eur. J. Cardiothorac. Surg.*, vol. 41, n.º 4, pp. 734-745, abr. 2012, doi: 10.1093/ejcts/ezs043.
- [8] C. L. B. Solís, S. G. Arriola, y A. O. Manzano, «Índice de Barthel (IB): Un instrumento esencial para la evaluación funcional y la rehabilitación».

- [9] S. Shah, F. Vanclay, y B. Cooper, «Improving the sensitivity of the Barthel Index for stroke rehabilitation», *J. Clin. Epidemiol.*, vol. 42, n.º 8, pp. 703-709, ene. 1989, doi: 10.1016/0895-4356(89)90065-6.
- [10] D. R. Petty, A. House, P. Knapp, T. Raynor, y A. Zernansky, «Validity and reliability of the Edmonton Frail Scale».
- [11] L. J. Gleason, E. A. Benton, M. L. Alvarez-Nebreda, M. J. Weaver, M. B. Harris, y H. Javedan, «FRAIL Questionnaire Screening Tool and Short-Term Outcomes in Geriatric Fracture Patients», *J. Am. Med. Dir. Assoc.*, vol. 18, n.º 12, pp. 1082-1086, dic. 2017, doi: 10.1016/j.jamda.2017.07.005.
- [12] S. Katz, «Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function», *JAMA*, vol. 185, n.º 12, p. 914, sep. 1963, doi: 10.1001/jama.1963.03060120024016.
- [13] A. Liberati *et al.*, «The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration», *J. Clin. Epidemiol.*, vol. 62, n.º 10, pp. e1-e34, oct. 2009, doi: 10.1016/j.jclinepi.2009.06.006.
- [14] M. J. Page *et al.*, «Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas», *Rev. Esp. Cardiol.*, vol. 74, n.º 9, pp. 790-799, sep. 2021, doi: 10.1016/j.recesp.2021.06.016.
- [15] G. Urrútia y X. Bonfill, «Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis», *Med. Clínica*, vol. 135, n.º 11, pp. 507-511, oct. 2010, doi: 10.1016/j.medcli.2010.01.015.
- [16] A. Ciapponi, «QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica», *Evid. Actual. En Práctica Ambulatoria*, vol. 18, n.º 1, abr. 2015, doi: 10.51987/evidencia.v18i1.6341.
- [17] J. S. Kwun, H. Ahn, J. W. Suh, y S. Y. Yoo, «Machine learning-based prediction of 30-day major adverse cardiac and cerebrovascular events in non-cardiac surgery patients», *Eur. Heart J.*, vol. 44, n.º Supplement_2, p. ehad655.2433, nov. 2023, doi: 10.1093/eurheartj/ehad655.2433.
- [18] N. Hou *et al.*, «Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost», *J. Transl. Med.*, vol. 18, n.º 1, p. 462, dic. 2020, doi: 10.1186/s12967-020-02620-5.

- [19] R. Sun *et al.*, «Prediction of 30-day mortality in heart failure patients with hypoxic hepatitis: Development and external validation of an interpretable machine learning model», *Front. Cardiovasc. Med.*, vol. 9, p. 1035675, oct. 2022, doi: 10.3389/fcvm.2022.1035675.
- [20] S. Sinha, T. Dong, A. Dimagli, A. Judge, y G. D. Angelini, «A machine learning algorithm-based risk prediction score for in-hospital/30-day mortality after adult cardiac surgery», *Eur. J. Cardiothorac. Surg.*, vol. 66, n.º 4, p. ezae368, oct. 2024, doi: 10.1093/ejcts/ezae368.
- [21] T. Siöland *et al.*, «ICURE: Intensive care unit (ICU) risk evaluation for 30-day mortality. Developing and evaluating a multivariable machine learning prediction model for patients admitted to the general ICU in Sweden», *Acta Anaesthesiol. Scand.*, vol. 68, n.º 10, pp. 1379-1389, nov. 2024, doi: 10.1111/aas.14501.
- [22] J. Liu *et al.*, «Interpretable Machine Learning Approach for Predicting 30-Day Mortality of Critical Ill Patients with Pulmonary Embolism and Heart Failure: A Retrospective Study», *Clin. Appl. Thromb.*, vol. 30, p. 10760296241304764, ene. 2024, doi: 10.1177/10760296241304764.
- [23] X. Lin *et al.*, «Machine learning models to predict 30-day mortality for critical patients with myocardial infarction: a retrospective analysis from MIMIC-IV database», *Front. Cardiovasc. Med.*, vol. 11, p. 1368022, sep. 2024, doi: 10.3389/fcvm.2024.1368022.
- [24] T. Kurita *et al.*, «Machine learning algorithm for predicting 30-day mortality in patients receiving rapid response system activation: A retrospective nationwide cohort study», *Heliyon*, vol. 10, n.º 11, p. e32655, jun. 2024, doi: 10.1016/j.heliyon.2024.e32655.
- [25] K. S. Betts *et al.*, «A machine learning approach to predicting 30-day mortality following paediatric cardiac surgery: findings from the Australia New Zealand Congenital Outcomes Registry for Surgery (ANZCORS)», *Eur. J. Cardiothorac. Surg.*, vol. 64, n.º 2, p. ezad160, ago. 2023, doi: 10.1093/ejcts/ezad160.
- [26] B. F. De Oliveira Gomes *et al.*, «Late Mortality After Myocardial Injury in Critical Care Non-Cardiac Surgery Patients Using Machine Learning Analysis», *Am. J. Cardiol.*, vol. 204, pp. 70-76, oct. 2023, doi: 10.1016/j.amjcard.2023.07.044.

- [27] U. Benedetto *et al.*, «Can machine learning improve mortality prediction following cardiac surgery?», *Eur. J. Cardiothorac. Surg.*, vol. 58, n.º 6, pp. 1130-1136, dic. 2020, doi: 10.1093/ejcts/ezaa229.
- [28] J. Han *et al.*, «Predicting 30-day mortality in severely injured elderly patients with trauma in Korea using machine learning algorithms: a retrospective study», *J. Trauma Inj.*, vol. 37, n.º 3, pp. 201-208, sep. 2024, doi: 10.20408/jti.2024.0024.
- [29] M. Pishgar, H. Li, y Z. Yu, «Prediction of 30-Day Mortality for ICU Patients with Sepsis-3», 25 de junio de 2024, *In Review*. doi: 10.21203/rs.3.rs-4491750/v1.
- [30] P. Darabi, S. Gharibzadeh, D. Khalili, M. Bagherpour-Kalo, y L. Janani, «Optimizing cardiovascular disease mortality prediction: a super learner approach in the tehran lipid and glucose study», *BMC Med. Inform. Decis. Mak.*, vol. 24, n.º 1, p. 97, abr. 2024, doi: 10.1186/s12911-024-02489-0.
- [31] S. Sinha *et al.*, «Comparison of machine learning techniques in prediction of mortality following cardiac surgery: analysis of over 220 000 patients from a large national database», *Eur. J. Cardiothorac. Surg.*, vol. 63, n.º 6, p. ezad183, jun. 2023, doi: 10.1093/ejcts/ezad183.
- [32] B. Choi *et al.*, «Prediction Model for 30-Day Mortality after Non-Cardiac Surgery Using Machine-Learning Techniques Based on Preoperative Evaluation of Electronic Medical Records», *J. Clin. Med.*, vol. 11, n.º 21, p. 6487, nov. 2022, doi: 10.3390/jcm11216487.

7.2. ÍNDICE DE ILUSTRACIONES

Ilustración 1. Metodología (Elaboración Propia)	11
Ilustración 2. Selección de Estudios (Elaboración Propia).....	22
Ilustración 3. Características de los estudios (Elaboración Propia).....	23
Ilustración 4. Boxplot métricas en los estudios individuales (Elaboración Propia)	24
Ilustración 5. Correlación métricas de los estudios individuales (Elaboración Propia) .	25
Ilustración 6. Metodología Análisis de Datos (Elaboración Propia)	33
Ilustración 7. Análisis de faltantes (Elaboración Propia)	43
Ilustración 8. EDA Edad (Elaboración Propia)	44
Ilustración 9. EDA IMC (Elaboración Propia)	44
Ilustración 10. EDA Htopre (Elaboración Propia)	45
Ilustración 11. EDA Leucograma (Elaboración Propia)	45
Ilustración 12. EDA Creatininapre (Elaboración Propia).....	45
Ilustración 13. EDA Proteinasatotales (Elaboración Propia)	45
Ilustración 14. EDA PCR (Elaboración Propia).....	46
Ilustración 15. EDA Test Fuerza Agarre (Elaboración Propia).....	46
Ilustración 16. EDA Test5metros (Elaboración Propia).....	46
Ilustración 17. EDA Estanciahospitalariacalculada (Elaboración Propia)	46
Ilustración 18. Matriz de correlación de las variables numéricas (Elaboración Propia)	47
Ilustración 19. Correlación de las variables numéricas con la variable objetivo (Elaboración Propia).....	47
Ilustración 20. QQ Edad (Elaboración Propia).....	48
Ilustración 21. QQ IMC (Elaboración Propia)	48
Ilustración 22. QQ Htopre	48
Ilustración 23. QQ Leucograma (Elaboración Propia).....	49
Ilustración 24. QQ Creatininapre (Elaboración Propia)	49
Ilustración 25. QQ Proteinasatotales (Elaboración Propia).....	49
Ilustración 26. QQ PCR (Elaboración Propia)	49
Ilustración 27. QQ TestFuerzaAgarre (Elaboración Propia).....	50
Ilustración 28. QQ Test5metros (Elaboración Propia)	50
Ilustración 29. QQ Estanciahospitalariacalculada (Elaboración Propia).....	50
Ilustración 30. EDA Género (Elaboración Propia).....	55
Ilustración 31. EDA DMdicotomica (Elaboración Propia)	55
Ilustración 32. EDA TabaquismoDicot (Elaboración Propia).....	55
Ilustración 33. EDA HTA (Elaboración Propia)	56
Ilustración 34. EDA DL (Elaboración Propia)	56
Ilustración 35. EDA Obesidad (Elaboración Propia)	56
Ilustración 36. EDA EPOCdicot (Elaboración Propia)	57
Ilustración 37. EDA Insuficientciarenalpreoperatoria (Elaboración Propia)	57
Ilustración 38. EDA ComplicacionesTODAS (Elaboración Propia)	57
Ilustración 39. EDA ComplicacionesMACE (Elaboración Propia)	57
Ilustración 40. EDA FragilidadGrST (Elaboración Propia)	58

Ilustración 41. FragilidadGST (Elaboración Propia).....	58
Ilustración 42. EDA Barthel (Elaboración Propia).....	58
Ilustración 43. EDA Katz (Elaboración Propia).....	59
Ilustración 44. EDA Frail (Elaboración Propia).....	59
Ilustración 45. EDA Edmonton (Elaboración Propia).....	59
Ilustración 46. EDA Variable Objetivo (Elaboración Propia).....	62
Ilustración 47. Pipeline RF pre-selección (Elaboración Propia)	63
Ilustración 48. RF pre-selección (Elaboración Propia)	63
Ilustración 49. Comparativa importancia de variables (Elaboración Propia).....	64
Ilustración 50. Modelo SVM (Fuente: Matlab MathWorks).....	65
Ilustración 51. Pipeline SMOTE (Elaboración Propia).....	68
Ilustración 52. Pipeline SVM (Elaboración Propia).....	69
Ilustración 53. Rendimiento SVM.....	69
Ilustración 54. Pipeline XGBoost.....	69
Ilustración 55. Rendimiento XGBoost (Elaboración Propia).....	69
Ilustración 56. Pipeline RF (Elaboración Propia).....	70
Ilustración 57. Rendimiento RF (Elaboración Propia)	70
Ilustración 58. Pipeline Catboost.....	70
Ilustración 59. Rendimiento Catboost	70
Ilustración 60. Pipeline KNN	71
Ilustración 61. Rendimiento KNN.....	71
Ilustración 62. Pipeline MLP.....	71
Ilustración 63. Rendimiento MLP	71
Ilustración 64. Pipeline XGBoost + Catboost (Elaboración Propia).....	72
Ilustración 65. Rendimiento XGBoost + Catboost (Elaboración Propia)	72
Ilustración 66. Pipeline RF + Catboost (Elaboración Propia)	72
Ilustración 67. Rendimiento RF + Catboost.....	72
Ilustración 68. Comparativa Modelos(Elaboración Propia)	73
Ilustración 69. Métricas modelos (Elaboración Propia)	74
Ilustración 70. Comparativa métricas modelos (Elaboración Propia).....	75
Ilustración 71. Rendimiento XGBoost (Elaboración Propia).....	77
Ilustración 72. Logo Python (Fuente Python Organization).....	81
Ilustración 73. Logo Streamlit (Fuente Streamlit Oficial).....	82
Ilustración 74. Logo Scikit-learn (Fuente SKLearn).....	82
Ilustración 75. QR PreMoCir (Elaboración Propia)	83
Ilustración 76. PreMoCir (Elaboración Propia).....	84

7.3. ÍNDICE DE TABLAS

Tabla 1. Puntuaciones del Índice de Barthel (Elaboración Propia).....	15
Tabla 2. Clasificación Índice de Barthel	16
Tabla 3. Puntuación Índice de Edmonton (Elaboración Propia)	17
Tabla 4. Clasificación Índice de Edmonton (Elaboración Propia)	17
Tabla 5. Puntuación Índice de Frail.....	18
Tabla 6. Puntuación Índice de Katz.....	19
Tabla 7 Riesgo de sesgo en los estudio (Elaboración Propia).....	24
Tabla 8. Riesgo de sesgo en los estudios individuales (Elaboración Propia).....	26
Tabla 9. Resumen de la evidencia de los estudios analizados (Elaboración Propia)	31
Tabla 10. Descripción Variables Numéricas (Elaboración Propia).....	41
Tabla 11. Descripción Variables Categóricas (Elaboración Propia)	42
Tabla 12. Descripción Variable Objetivo (Elaboración Propia).....	43
Tabla 13. Métricas Test de Normalidad (Elaboración Propia).....	51
Tabla 14. Diagramas de dispersión (Elaboración Propia)	52
Tabla 15. Test de Levene Variables Numéricas (Elaboración Propia).....	54
Tabla 16. Test de Levene Variables Categóricas (Elaboración Propia)	61
Tabla 17. Comparativa modelos (Elaboración Propia)	73
Tabla 18. Métricas modelos (Elaboración Propia)	74
Tabla 19. Métricas XGBoost	77