

Bioingeniería informática aplicada a la predicción de enfermedades cardiológicas y su implementación en el Hospital Delicia Concepción Masvernati de la ciudad de Concordia, provincia de Entre Ríos

María E. Silva Layes, Marcelo G. Benedetto, Duval H. Benitez, Jesús F. Frola da Silva, Ana L. Carabio, Juan J. Aguirre, Elio D. Costen

Autoras/es: Facultad de Ciencias de la Administración. Universidad Nacional de Entre Ríos. Monseñor Tavella 1424. Concordia. CP (3200), Entre Ríos. Argentina.

Contacto: elizabeth.silva@uner.edu.ar

ARK: <https://id.caicyt.gov.ar/ark:/s22504559/ymlca5txf>

Resumen

Las enfermedades cardiovasculares constituyen una de las principales causas de mortalidad en el mundo, lo que plantea el desafío de contar con herramientas innovadoras que permitan anticipar riesgos, y fortalecer el diagnóstico temprano y la implementación de medidas preventivas. En este marco, el proyecto se orienta al desarrollo de un sistema predictivo de riesgo cardiovascular mediante la aplicación de técnicas de *machine learning*. El propósito central es generar un componente de software accesible, que permita a los usuarios médicos registrar variables relevantes y recibir información preventiva personalizada. Para alcanzar dicho objetivo se propone, en primer lugar, identificar variables clínicas relevantes, y el diseño de un esquema de datos que contemple calidad, consistencia e interpretabilidad de la información. En segundo lugar, se trabajará en el preprocesamiento de los registros clínicos, incluyendo limpieza, normalización e integración, acompañado de un análisis comparativo de herramientas específicas para esta etapa. En tercer lugar, se contemplará la selección de las alternativas algorítmicas adecuadas para la construcción de modelos predictivos robustos y aplicables al contexto clínico real. Con este enfoque, el proyecto busca sentar bases sólidas para incorporar la inteligencia artificial en el ámbito sanitario, contribuyendo al fortalecimiento de la toma de decisiones clínicas de manera proactiva.

Palabras clave: Minería de datos, Aprendizaje automático, Soporte a Decisiones Clínicas, Enfermedades cardiovasculares, aplicación predictiva, Medicina preventiva.

Objetivos propuestos y cumplidos

Desarrollo de un componente de software capaz de generar, con aprendizaje automatizado, un modelo con capacidades predictivas sobre enfermedades cardiológicas; que permitan un mejor soporte a decisiones de diagnóstico clínico y un avance significativo en la medicina preventiva para actos asistenciales de pacientes.

Objetivos cumplidos:

Los objetivos propuestos en general han sido cumplidos. No pudiéndose incorporar el componente desarrollado por la distribución desigual y escasez de registros de patologías con la que contamos, lo que limita la capacidad del modelo para aprender patrones clínicos más amplios.

Marco teórico (síntesis)

Actualmente es inconmensurable la información con la que cuentan las instituciones de salud sobre sus pacientes, tanto a nivel administrativo como asistencial. Los datos referidos al área clínica son utilizados por el médico para obtener información del paciente, ya sea para efectuar diagnósticos, tratamientos, medicación; así como también por los propios sistemas informáticos para generar múltiples alertas (como pueden ser interacciones medicamentosas, control sobre medicamentos contraindicados para una patología específica, resultados anormales en estudios, entre otros). Existe, en el sector sanitario la marcada necesidad de encontrar maneras de analizar e interpretar estos grandes volúmenes de datos, con el objetivo de encontrar patrones de comportamiento; que puedan servir de apoyo en la toma de decisiones médicas.

El análisis de este gran cúmulo de datos nos brinda una oportunidad inmejorable para [1]:

- Realización de estudios epidemiológicos (análisis de rendimientos de campañas de información, prevención, sustitución de fármacos, entre otros),
- Cálculo de expectativas de vida,
- Identificación de terapias médicas satisfactorias para diferentes enfermedades,
- Asociación de síntomas y clasificación diferencial de patologías,
- Estudio de factores de riesgo para la salud en distintas patologías,
- Segmentación de pacientes para una atención más inteligente y
- Identificación de terapias médicas y tratamientos erróneos para determinadas enfermedades.

De esta manera, aparecen nuevas oportunidades para mejorar, no sólo la medicina personalizada sino también el rendimiento y resultado de los sistemas sanitarios. Esto conlleva a identificar necesidades referidas a nuevas maneras de administrar, integrar, analizar e interpretar tales conjuntos de datos; que posibiliten encontrar patrones de comportamiento de utilidad en la toma de decisiones médicas. En este aspecto, la minería de datos es una herramienta fundamental que resulta imprescindible en el análisis de dichos datos.

Si bien, la minería de datos ha existido por más de tres décadas, su potencial se comienza a vislumbrar a mediados de la segunda década del siglo XXI. En particular, en el área de la salud es a fines de los 90's cuando comienzan los primeros intentos de rea-

lizar estudios que involucran la minería de datos asociada a investigaciones puntuales y que hoy ocupan una posición relevante en el área [3], [4], [5].

La esperanza de vida humana promedio está en aumento a lo largo de la población mundial, lo que plantea nuevos desafíos a los métodos de tratamiento actuales. Si bien los modelos de tratamiento y prácticas clínicas han cambiado, muchos de estos cambios se deben principalmente a los datos. Los médicos, quieren comprender con el mayor grado de detalle y lo más temprano posible acerca de sus pacientes, para detectar signos de advertencia de una enfermedad grave a medida que los mismos surgen. Para esto, es crucial que el gran cúmulo de datos clínicos que actualmente poseen las instituciones sanitarias pueda transformarse rápidamente en conocimiento.

Actualmente, la utilización de la minería de datos en el sector sanitario [2] es aplicado en diferentes instancias de los actos clínicos o de gestión. Entre los sistemas predictivos se encuentran aquellos utilizados para gestionar los recursos humanos en los hospitales, siendo un ejemplo de esto el que surge del análisis de ingresos de los pacientes a las emergencias. Los mismos, han permitido a los investigadores identificar patrones relevantes en las tasas de admisión; información que luego ha podido ser utilizada aplicando aprendizaje automático para encontrar los algoritmos más precisos que predijeron las futuras tendencias de ingresos de pacientes [6]. Esto permite la toma de decisiones respecto a la necesidad de incorporar personal adicional cuando se espera un gran número de consultas; lo que lleva a tiempos de espera reducidos para los pacientes y una mejor calidad de atención.

Otra aplicación en la medicina es la incorporación de alertas en tiempo real en circunstancias en las que el paciente no se encuentra en el hospital, recopilando información de manera continua de los datos de salud de los pacientes y enviándolos a la nube. Esto, permite analizar el estado de salud del público en general, procurando a los médicos comparar estos datos en el contexto socioeconómico y de ser necesario modificar las estrategias de entrega.

Igualmente, el uso de *Big Data* en la asistencia sanitaria permite la planificación estratégica al tener una mejor comprensión de las motivaciones de las personas a un tratamiento particular; permitiendo a los gestores de salud analizar los resultados verificados entre personas de diferentes grupos demográficos e identificar qué factores desalientan a las personas a tomar dicho tratamiento. En este sentido, cabe destacar que la Universidad de Florida utilizó *Google Maps* y datos gratuitos de salud pública para preparar mapas de calor dirigidos a múltiples problemas (crecimiento de la población y enfermedades crónicas), comparando posteriormente estos datos con la disponibilidad de servicios médicos en la mayoría de las áreas climatizadas. Estas acciones, han permitido reevaluar la estrategia de abordaje a la población y el agregado de unidades de atención a las áreas consideradas más problemáticas.

Si bien, en la actualidad existen investigaciones que conllevan a desarrollos informáticos que sirven de apoyo en tratamientos de distintas enfermedades, las implementaciones que permitan incorporar el apoyo a la toma de decisiones clínicas en el marco de la atención del paciente en tiempo real son escasas. Teniendo en cuenta que en nuestro país una de las principales causas de muerte son las enfermedades cardiovasculares, se considera necesario el desarrollo de sistemas fiables que sirvan de apoyo a las decisiones clínicas en dos momentos fundamentales (la respuesta del sistema cuando el paciente concurre a una visita médica y una respuesta proactiva que permita brindar apoyo a través de la aplicación de medicina preventiva); logrando

reducir el tiempo del diagnóstico y aumentar la precisión del mismo en este tipo de enfermedades [7].

Metodología utilizada para lograr los objetivos planteados:

1. Se realizó la revisión bibliográfica en torno a la aplicación de técnicas de inteligencia artificial y *machine learning* en el ámbito de la salud, particularmente en la predicción de enfermedades cardiovasculares. Se analizaron diferentes estudios donde se analizaba la utilización de diferentes variables clínicas y demográficas como predictores de riesgo, así como enfoques metodológicos de preprocesamiento de datos clínicos. Esta revisión sirvió como base para definir las variables iniciales del estudio, seleccionar estrategias de imputación de datos faltantes, y evaluar distintas herramientas de preprocesamiento que garantizaran la calidad y confiabilidad del *dataset* utilizado. Esta actividad se realizó transversalmente durante todo el proyecto debido a la actualización permanente del área.
2. Se definió un conjunto amplio de variables clínicas, demográficas y de hábitos de vida vinculadas al riesgo cardiovascular, seleccionadas con el apoyo de especialistas en cardiología. Esta etapa permitió establecer un marco robusto de indicadores que sirvió de base para el modelado posterior. La inclusión de variables adicionales sugeridas por los expertos aportó mayor riqueza al conjunto de datos, aun cuando algunas presentaron niveles de información faltante que debieron resolverse mediante técnicas de imputación.
3. Se obtuvieron y depuraron los conjuntos de datos anonimizados del Hospital "Delicia Concepción Masvernat" (Concordia, Entre Ríos). El proceso de recolección se vio retrasado por la pandemia de COVID-19, lo cual impactó en la disponibilidad y calidad inicial de los registros obtenidos. Sobre estos, se implementaron procedimientos de limpieza, normalización, reducción y transformación, que derivaron en un *dataset* estructurado y utilizable para las etapas analíticas posteriores.
4. Se realizó la evaluación de herramientas de preprocesamiento a través de un análisis comparativo de distintas plataformas de software (KNIME, RapidMiner, WEKA, Talend Open Studio, Data Civilizer y Data Cleaner), tomando en cuenta criterios de performance, funcionalidad, usabilidad, soporte de tareas específicas, documentación y actualización. Este estudio permitió seleccionar las herramientas más adecuadas para garantizar eficiencia y robustez en el tratamiento de grandes volúmenes de datos clínicos heterogéneos.
5. Se definió un esquema estandarizado en formato CSV a partir de los procesos de preprocesamiento, que contempló las principales variables clínicas y de laboratorio. Se establecieron criterios específicos, como considerar las últimas tres mediciones (en casos de más de 3 ocurrencias) de parámetros relevantes (PCR, creatinina, proteinuria, hemoglobina y colesterol), mejorando la confiabilidad de la información de entrada a los modelos predictivos.
6. Se desarrolló un componente de software, así como también una app que permite, ingresando los valores para las variables clínicas seleccionadas, obtener la predicción de riesgo de padecer una enfermedad cardiológica.

Resultados

Se utilizaron en primera instancia las variables de riesgo de acuerdo a análisis efectuados bajo las siguientes investigaciones: [8], [9], [11], [12], [13]. También fueron incluidas nuevas variables teniendo presente las sugerencias realizadas por el experto en cardiología. Los sets de datos con los que se trabajó corresponden a registros médicos de pacientes provenientes del Hospital “Delicia Concepción MASVERNAT” de la ciudad de Concordia, Provincia de Entre Ríos, entre los años 2018 a 2022 respectivamente. Se destaca que para algunas variables del conjunto de datos brindados hubieron algunos niveles de datos faltantes. Para ellos se consideró la mediana, enfoque utilizado en *machine learning* por diferentes algoritmos para tratar con la ausencia de valores en el aprendizaje automático [14].

Estructura de los datos formateados en la etapa de preprocesamiento

El preprocesamiento de datos involucró diferentes pasos a tener en cuenta: la limpieza, integración, reducción y transformación de datos. En cuanto a la reducción de datos, esta tuvo como objetivo obtener un menor volumen de datos, en la que los resultados analíticos fueran los mismos o muy cercanos a los del volumen inicial. A ella se aplicaron técnicas de compresión, selección de atributos, así como la construcción de nuevos atributos; teniendo en cuenta también la reducción de numerosidad, es decir, el reemplazo de datos, utilizando modelos paramétricos para representaciones más pequeñas.

Análisis de herramientas de preprocesamiento de datos. Criterios y características involucradas

Cada una de las herramientas seleccionadas fueron analizadas teniendo como base estudios previos realizados por expertos sobre herramientas utilizadas para realizar minería de datos [10] [11]. Con esta información se plantearon cuatro categorías principales a evaluar: performance, funcionalidad, usabilidad, soporte de determinadas tareas, a las que se le sumó la evaluación de otras dos categorías, tales como documentación y actualización.

Si bien el estudio estuvo basado mayoritariamente en considerar el análisis de los criterios planteados en [10] y [11] para las diferentes categorías, algunos de ellos fueron descartados para esta etapa, y se consideraron otros referidos específicamente a la etapa de preprocesamiento de datos.

Selección y descripción de los algoritmos

Primariamente se consideraron para su evaluación 6 (seis) herramientas de software que permiten preprocesamiento de datos: *KNIME*, *RapidMiner*, *WEKA*, *Talend Open Studio*, *Data Civilizer*, *Data Cleaner*

La preselección de estas herramientas se realizó considerando principalmente que las mismas fueran del tipo “open source”. Esta decisión estuvo basada en la disponibilidad del acceso necesario a dichas herramientas durante el período en que se lleva a cabo el proyecto, considerando que muchas de ellas, de tipo propietarias, limitan su acceso a un período de tiempo, a conjuntos de datos de determinado tamaño y/o restringen el uso de determinadas funcionalidades, entre otras dificultades.

Se destaca que luego de efectuar un análisis preliminar de cada una de las herramientas, se decidió desestimar las herramientas *Data Civilizer* y *Data Cleaner*, teniendo en cuenta que por un lado *Data Civilizer* no registra actualizaciones recientes siendo la última de hace unos años (2019), sí soporte técnico razonable, y escasa documentación, incompatible con otras herramientas KDD, posee bajo desempeño y un alto grado de dificultad al momento de comenzar a adquirir conocimientos en este software, por otro lado, referido a *Data Cleaner*, respecto a la documentación relevada de la herramienta y su historial de versiones, se pudo determinar que este paquete aún se encuentra en una etapa temprana de desarrollo.

Evaluación de las Herramientas

Para la realización de la evaluación de las diferentes herramientas para algunos criterios de *performance*, se tomó como base un *dataset* proporcionado por el Hospital “Delicia Concepción MASVERNAT.

Para desarrollar las pruebas de preprocesamiento con los criterios de *performance* de cantidad de datos y diversidad de fuentes de datos que maneja, robustez y que tan bien funciona el software con *n dataset* en gran escala; se trabajó con un primer conjunto de datos, en el cual se fueron duplicando los mismos para obtener diferentes cantidades de registro y efectuar el análisis comparativo.

Tabla 1. Tabla de ponderación de herramientas

Característica	Criterio	Peso de Referencia	RAPIDMINER	KNIME	WEKA	TALEND OPEN STUDIO
Performance (0.30)	Plataformas en las que corre	0,05	0,05	0,05	0,05	0,05
	Requerimientos del sistema para ejecutarse	0,1	0,06	0,1	0,1	0,1
	Arquitectura del software	0,05	0,025	0,025	0,025	0,025
	Acceso a datos heterogéneos	0,1	0,1	0,1	0,08	0,1
	Tamaño de los datos	0,2	0,2	0,2	0,066	0,2
	Eficiencia	0,15	0,1	0,08	0,15	0,12
	Interoperabilidad	0,05	0,05	0,05	0,05	0,01
	Robustez	0,3	0,2	0,2	0,3	0,3
Total			0,2355	0,2415	0,2463	0,2715

Tabla 1, continuación

Funcionalidad (0.20)	Técnicas para pre-procesamiento de datos	0,25	0,2	0,15	0,25	0,2
	Flexibilidad en el tipo de datos	0,15	0,15	0,12	0,12	0,15
	Exportación de datos	0,15	0,15	0,12	0,12	0,15
	Problemas que permite resolver	0,25				
	Modelado	0,2	0,2	0,2	0,2	0,2
Total			0,14	0,118	0,138	0,14
Usabilidad (0.30)	Interface de usuario	0,25	0,25	0,2	0,25	0,25
	Curva de aprendizaje	0,2	0,2	0,16	0,2	0,16
	Visualización de los datos	0,25	0,25	0,2	0,2	0,15
	Tipo de usuarios	0,15	0,12	0,12	0,15	0,12
	Reporte de errores	0,15	0,15	0,12	0,09	0,15
Total		0,3	0,291	0,2328	0,207192	0,17196936
Soporte de determinadas tareas (0.10)	Modificación de valores falsos	0,25	0,25	0,25	0,25	0,25
	Sustitución de datos	0,2	0,2	0,2	0,2	0,2
	Filtrado de datos	0,25	0,25	0,25	0,25	0,25
	Eliminación de datos	0,1	0,1	0,1	0,1	0,1
	Atributos derivados	0,2	0,2	0,2	0,2	0,2
Total		0,1	0,1	0,1	0,1	0,1
Documentación (0.05)	Calidad, cantidad e idiomas	0,35	0,35	0,28	0,35	0,28
	Revisiones	0,25	0,25	0,2	0,25	0,25
	Comunidad	0,4	0,4	0,4	0,32	0,4
Total			0,05	0,044	0,046	0,0465
Actualización (0.05)	Frecuencia	0,4	0,4	0,4	0,32	0,32
	Soporte técnico	0,35	0,28	0,21	0,35	0,21
	Bug tracking	0,25	0,2	0,15	0,2	0,2
Total			0,044	0,038	0,0435	0,0365

De una evaluación general, contemplando todas las características, pruebas de *performance* de los diferentes sets de datos y en base a las ponderaciones establecidas con los criterios precedentes, no es posible establecer que exista una herramienta que contemple todo y de la mejor manera. Lo que sí se puede determinar son las bondades de cada una de ellas y combinar su utilización de acuerdo a las necesidades que plantea el proyecto en cuestión en la etapa de preprocesamiento.

Se destaca que a pesar de no haberse identificado una herramienta que, para cada una de las variables analizadas supere al resto, sí se identificó que *RAPIDMINER* se destaca con una puntuación general total de 0.8605, siendo superior a todas las herramientas evaluadas; concluyendo así que para esta etapa del proyecto que se encuentra en desarrollo, se elige utilizar la herramienta *RAPIDMINER*.

Análisis y Selección de Variables

De las variables iniciales previstas para poder analizar, contamos que el hospital solamente nos podía brindar algunas de ellas. Se presenta a continuación el cuadro (Tabla 2) en el que se destacan en color verde aquellas variables con las que se cuenta para su análisis, y en rojo aquellas de las que se tuvo que prescindir:

Tabla 2. Variables utilizadas para análisis luego del primer estudio

Nombre	Tipo	Descripción	Descriptor
Edad	Continuo	Edad en años	(se calcula con fecha de nacimiento y fecha del evento cardíaco)
Sexo	Discreto	F - Femenino M - Masculino I - Indeterminado	
Etnia	Discreto	Caucásico blanco; Asia meridional; Negro / Afro-Caribeño; Chino / asiático oriental; Otro / Mezclado; Desconocido	
Actividad Física	Discreto	1 - SI 0 - NO	CIE Z72.3 (falta de ejercicio)
Estrés	Discreto	1 - SI 0 - NO	CIE Z73.3
Fecha Nacimiento	Continuo		(para calcular la edad)
Peso	Continuo		(para calcular IMC)
Talla	Continuo		(para calcular IMC)
IMC (Índice Masa Corporal)	Continuo		(se calcula con peso y talla)

Tabla 2, continuación

Presión Arterial Diastólica	Continuo	mmHg	
Presión Arterial Sistólica	Continuo	mmHg	
Frecuencia Cardíaca (máxima)	Continuo	82 a 188	
Antecedentes Familiares de Enfermedad Cardíaca	Discreto	1 – SI 0 – NO	
Enfermedad Renal Crónica	Discreto	1- SI 0- NO	CIE N18
Obesidad	Discreto	1 – SI 0 – NO	CIE E66 En caso de no tener estos datos, se puede obtener considerando un peso mayor al 20% de la media. Hay que tener tabla de la media por edad
Fumador (tabaquismo)	Discreto	1 – SI 0 – NO	CIE Z72.0
APNEA obstructiva del sueño	Discreto	1 – SI 0 – N	CIE G47.3
Hipercolesterolemia	Discreto	1 – SI 0 – NO	CIE E78.0
Colesterol total	Continuo		El colesterol total es la suma de 3 tipos diferentes de colesterol (colesterol total = colesterol de LDL + colesterol de HDL + colesterol de VLDL)
Colesterol LDL	Continuo	mmol/L	Para obtener colesterol total
Colesterol HDL	Continuo	mmol/L	
Colesterol de VLDL (colesterol de lipoproteínas de muy baja densidad)	Continuo	mmol/L	Considerado para obtener colesterol total si no se tiene
Triglicéridos	Continuo	mmol/L	
PCR (Prueba de Proteína C Reactiva)	Continuo	mg/L	
Creatinina	Continuo	mg/dl	
Proteinuria	Continuo	mg/dl	
Hemoglobina glicosilada (hba1c) (Azúcar en Sangre)	Continuo	g/dl	
Uso de antihipertensivos	Discreto	1 – SI 0 – NO	Esto puede ayudar a identificar al hipertenso, si no se tiene medidas de hipertensión del paciente
VES (velocidad de eritrosedimentación)	Continuo	mm/h (milímetros por hora)	

Tabla 2, continuación

Cateterismo	Discreto	0 - Normal 1 - Anormal	En caso de SI, considerar todos los datos antes del cateterismo. Si tiene más de uno, se considera el primero antes del anormal
Fecha del Cateterismo	Continuo		
Diabetes Mellitus	Discreto	1 - SI 0 - NO	Se puede identificar por CIE10 o SNO-MED E10 a E14
Hipertensión Arterial	Discreto	1 - SI 0 - NO	Se puede identificar por CIE10 o SNOMED o con medidas sucesivas de presión arterial alta CIE I10.x
Dislipemia (Patología)	Discreto	1 - SI 0 - NO	Se puede identificar por CIE10 o SNO-MED ó con medidas altas de colesterol, triglicéridos o ambas - E78
Hipertrofia Ventricular Izquierda	Discreto	1 - SI 0 - NO	
EPOC	Discreto	1 - SI 0 - NO	CIE 10 J44
Volumen Espiratorio Forzado (VEF1)	Continuo	%	
Gamma glutamyltransferase (gamma GT)	Continuo	IU/L	
Intolerancia a la glucosa	Discreto	1 - SI 0 - NO	Glucosa entre 140 y 200 mg/dL
Enfermedad Cardíaca	Discreta	CIE10	Se consideran como códigos de CIE10s válidos: "I20%", "I21%", "I22%", "I23%", "I24%", "I25%". Se incluyen códigos que hacen referencia a enfermedades cerebrovasculares: "I60%", "I61%", "I62%", "I63%", "I64%", "I65%", "I66%", "I67%", "I68%", "I69%"

	Variable con la que no se cuenta
	Variables con las que se cuenta para realizar el análisis

Para realizar el análisis se cuenta con un set de datos de 3162 registros. Se destaca que los datos considerados para formar el set de datos de entrenamiento en lo que se refiere a valores clínicos y diagnósticos, son aquellos cuya fecha de valoración fueron

inmediatamente anteriores al diagnóstico de enfermedad cardiovascular para aquellos pacientes diagnosticados y los valores clínicos más actuales para aquellos pacientes no diagnosticados.

Análisis de algoritmos de *Machine Learning* (ML)

De acuerdo al estudio realizado por Hasan, S. y otros [15], en el que se realiza la evaluación de diferentes algoritmos de ML, se destaca la muy buena performance lograda por el algoritmo de Regresión Logística comparado con otros algoritmos como K-NearestNeighbors, DecisionTree, GaussianNaïveBayes y RandomForest en lo que respecta a la predicción de enfermedades cardíacas.

Por otro lado, Dangare, C. y Apte, S. [16] realizan un estudio vinculado al mismo tema, pero en este caso evaluando los algoritmos Redes Neuronales, DecisionTrees y NaiveBayes. De dicho estudio se desprende que las Redes Neuronales proporcionan resultados precisos en comparación con los algoritmos DecisionTrees y NaiveBayes.

Suhendra, R. y otros [17], realizaron un estudio para determinar la efectividad del clasificador de refuerzo gradiente (Gradient Boosting) en la predicción de enfermedades cardiovasculares. El estudio fue realizado utilizando dos conjuntos de datos distintos para tareas de clasificación binaria, obteniéndose una precisión del 97.63% al aplicar el algoritmo, lo cual destaca la capacidad del mismo para detectar con precisión las enfermedades cardiovasculares.

En el estudio realizado por Yang, J. [18] en 2023, vinculado al análisis de enfermedades del corazón con XGBoost utilizan una recopilación de registros médicos proveniente de una destacada base de datos de investigación médica, que ofrece indicadores vitales que sugieren la posibilidad de enfermedad cardíaca. A través de este estudio Yang concluye que los resultados experimentales obtenidos aplicando el modelo XGBoost, al ser ajustado con los hiperparámetros adecuados, mostró un rendimiento predictivo impresionante en la predicción de enfermedades cardíacas, destacando su eficacia como herramienta de predicción.

Teniendo presente las valoraciones realizadas en distintas investigaciones acerca de la aplicación de diferentes algoritmos de ML para la predicción de enfermedades cardiovasculares, se opta por realizar el estudio utilizando las técnicas de Regresión Logística, Redes Neuronales, Gradient Boosting y XGBoost, y aplicarlas en nuestra investigación con el fin de analizar y predecir la presencia de enfermedades cardiovasculares en pacientes del Hospital “Delicia Concepción Masvernat”. En general, las técnicas seleccionadas ofrecen la capacidad de analizar grandes conjuntos de datos clínicos y biomédicos para identificar patrones y factores de riesgo que pueden pasar desapercibidos en los métodos tradicionales.

XGBoost es un algoritmo de aprendizaje automático desarrollado por Tianqi Chen y Carlos Guestrin, que se basa en un árbol de decisiones, utilizando un marco de potenciación de gradientes cuyo objetivo es resolver problemas de ciencia de datos de manera eficiente y precisa. El mismo puede ejecutarse en entornos distribuidos como Hadoop, pudiendo resolver problemas con grandes conjuntos de datos [19]. Se destaca que dicho algoritmo recibió reconocimientos por su influencia en diversas aplicaciones de la industria de vanguardia, lo que ha generado que muchos equipos de científicos de datos realicen diferentes aportes a distintos proyectos de código abierto relacionados a XGBoost.

XGBoost es ampliamente conocido por su eficiencia dado que se encuentra diseñado para ser altamente eficiente y escalable, con capacidades de paralelización y optimización que lo hacen adecuado para grandes conjuntos de datos [19]. La computación paralela y distribuida hace que el aprendizaje sea más rápido, lo que permite una exploración de modelos más rápida [20]. Además, es muy utilizado por su análisis de importancia de características y manejo de valores faltantes, siendo ampliamente utilizado en tareas como regresión, clasificación y ranking.

El algoritmo funciona añadiendo de forma secuencial aprendices débiles al conjunto de datos, donde cada nuevo aprendizaje se enfoca en corregir los errores cometidos por los existentes. Utiliza una técnica de optimización de descenso de gradiente para minimizar una función de pérdida predefinida durante el entrenamiento.

Las características claves del algoritmo XGBoost incluyen su capacidad para manejar relaciones complejas en los datos, técnicas de regularización para prevenir el sobreajuste e incorporación de procesamiento paralelo para una computación eficiente. XGBoost se utiliza ampliamente en varios dominios debido a su alto rendimiento predictivo y versatilidad en diferentes conjuntos de datos.

De acuerdo a Yang, C. [18] el modelo XGBoost supera a otros modelos en precisión y rendimiento, lo que resalta su potencial en la predicción de enfermedades cardíacas y su relevancia para mejorar las intervenciones médicas tempranas y el manejo de pacientes. Por ello, en su investigación Yang concluye que XGBoost, ha demostrado consistentemente un rendimiento excepcional en varias tareas en el ámbito del aprendizaje automático.

Adquisición de datos y procesamiento

Se realiza el procesamiento con Pandas, quedando la disposición de los datos de forma contigua, por columnas, permite que el procesador agrupe o fusione los accesos a la memoria para el procesamiento a nivel de fila; esto logra un uso eficiente del almacenamiento en caché mientras se reduce el número de accesos a la memoria principal (Ver Figura 1).

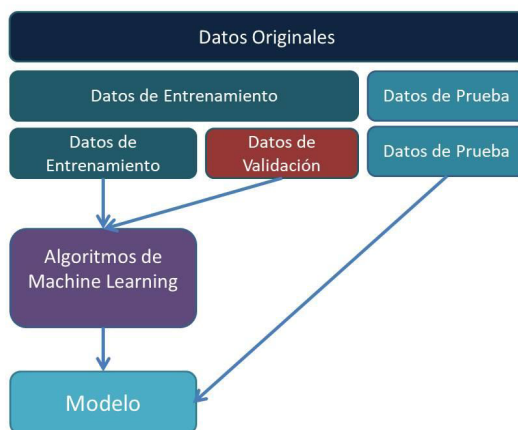
Figura 1. Variables utilizadas para el estudio

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3162 entries, 0 to 3161
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Edad                                  3162 non-null  int64
1   Sexo                                  3162 non-null  object
2   estres                                3162 non-null  int64
3   renal_cronica                         3162 non-null  int64
4   obesidad                               3162 non-null  int64
5   fumador                                3162 non-null  int64
6   estilo_vida                           3162 non-null  int64
7   apnea                                  3162 non-null  int64
8   hipercolesterolemia_x_diag            3162 non-null  int64
9   diabetes                               3162 non-null  int64
10  hipertension                           3162 non-null  int64
11  dislipemia                             3162 non-null  int64
12  epoc                                    3162 non-null  int64
13  pcr_valor                              3162 non-null  float64
14  creatinina_valor                       3162 non-null  float64
15  proteinuria_valor                      3162 non-null  object
16  hemoglobina_valor                      3162 non-null  float64
17  colesterol_valor                       3162 non-null  object
18  Evento_Cardiologico                    3162 non-null  int64
dtypes: float64(3), int64(13), object(3)
memory usage: 469.5+ KB
```

Configuración del entrenamiento

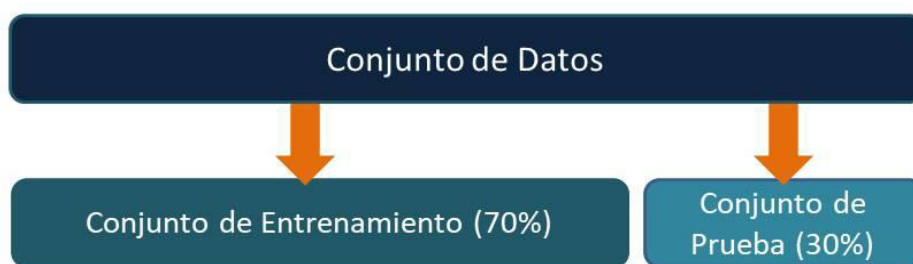
Para entrenar el algoritmo de Aprendizaje Automático, se dividió el conjunto de datos en dos subconjuntos: conjunto de entrenamiento y conjunto de prueba, donde el conjunto de entrenamiento es utilizado para realizar el entrenamiento del algoritmo mientras que el de prueba es utilizado para realizar las validaciones del modelo. A su vez, el conjunto de datos de entrenamiento es dividido en un conjunto de entrenamiento y otro para validación como muestra la Figura 2.

Figura 2. Esquema de división de conjunto de datos para el entrenamiento



Del conjunto total de datos originales asignados al entrenamiento (3162), el 70% (2213) fueron utilizados para entrenamiento de los modelos, y el 30% (949) fueron utilizados como datos de validación, como se aprecia en la Figura 3.

Figura 3. Esquema de división del conjunto de datos



Se destaca que, del total de los datos de entrenamiento (2213), fueron identificados 100 pacientes con evento cardiológico registrado y 2113 pacientes que no tenían registro de haber tenido un diagnóstico de enfermedad cardiológica previa.

Discusión de la evaluación

En los resultados obtenidos de la utilización de los distintos modelos (ver Tabla 3) se puede observar que las tasas de *accuracy* (proporción de predicciones correctas) son altas, generalmente alrededor del 94%. Sin embargo, este no es un resultado totalmente representativo, dado que la tasa de la detección de eventos cardiológicos varía desde un piso de 15% para modelos con bajo desempeño como Logistic Regression

(LR), logrando un desempeño de hasta el 43% para Random Forests (RF), o más precisamente hasta 54% para Extreme Gradient Boosting (XGBoost).

Tabla 3. Evaluación empírica del modelo

Nro.	Algoritmo	Umbral	Accuracy	Precision	Recall	F2-Score
λ_1	Regresión Logística	0.5	93.68	0.00	0.00	-
λ_2	Gradient Boosting	0.5	94.52	66.67	14.29	-
λ_3	Random Forest	0.5	94.84	68.42	23.21	
λ_4	Red Neuronal	0.5	94.10	0.00	0.00	-
λ_5	XGBoost	0.5	94.94	100.00	14.29	-
λ_6	Random Forest	0.283019	94.94	60.00	42.86	0.510
λ_7	XGBoost	0.178965	94.52	53.57	53.57	0.536

Del análisis de algoritmos realizado se opta por el algoritmo XGBoost, en primer lugar, por el resultado prometedor de los ensayos de laboratorio que realizamos con distintos algoritmos, superando inclusive las métricas de algoritmos más recientes, o con mayor interés por los investigadores; su fundamentación académica y en el gran impacto que ha tenido en la comunidad científica desde su presentación. Chen y Guestrin [2] introdujeron este sistema como una solución escalable de boosting, que incluye innovaciones técnicas tales como un algoritmo consciente de la dispersión de datos (*sparsity-aware*), el esquema de cuantiles ponderados (*weighted quantile sketch*) y optimizaciones a nivel de caché y particionado de datos. Con estas mejoras, XGBoost ha sido capaz de manejar eficientemente conjuntos de datos y escalado, convirtiéndose en un estándar de referencia y aplicaciones reales de ciencia de datos.

En segundo término, la robustez del algoritmo se ve reforzada por sus mecanismos de regularización y control del sobreajuste, los cuales son fundamentales en aplicaciones de carácter riguroso, como en el dominio médico. XGBoost incorpora penalizaciones L1 y L2 en su función objetivo, así como estrategias de *shrinkage* (reducción de tasa de aprendizaje) y *early stopping* basadas en validación, que permiten optimizar el balance entre sesgo y varianza. Estas capacidades mejoran la capacidad de generalización de los modelos, reduciendo la propensión al sobreajuste, lo que se alinea con los principios metodológicos deseables [23].

Finalmente, la capacidad interpretativa del modelo ha recibido creciente atención en los últimos años, particularmente mediante el uso de métodos basados en teoría de juegos. Los valores SHAP (*SHapley Additive exPlanations*), desarrollados por Lundberg y Lee [24], se han consolidado como un enfoque consistente para atribuir la contribución de cada variable en la predicción del modelo. Estudios recientes, como el de Tarabanis, Kalampokis, Khalil, Alviar, Chinitz y Jankelson [25], destacan cómo la combinación de XGBoost y SHAP ofrece transparencia y puede explicar los resultados, lo cual es crucial en contextos académicos y profesionales donde se requiere no solo obtener resultados precisos, sino también interpretables y justificables.

Durante la construcción del modelo predictivo se aplicó una metodología de ajuste del umbral de clasificación en el algoritmo XGBoost con el objetivo de optimizar la de-

tección de eventos cardiológicos. Este enfoque fue evaluado en el marco del presente proyecto, a través de la construcción de un *framework* que ha sido implementado, demostrando su potencial en condiciones de registro de datos clínicos fidedignos.

Se priorizó la reducción de falsos negativos, dada la alta criticidad de este tipo de error en medicina, mediante el ajuste del umbral que maximiza el *F2-Score*. Esta estrategia permitió incrementar la sensibilidad en la detección de casos clínicamente relevantes, asumiendo un ligero aumento en la tasa de falsos positivos.

Arquitectura utilizada

El sistema predictivo desarrollado fue desplegado como un servicio web utilizando una arquitectura *RESTful* accesible desde múltiples interfaces. La configuración del servidor contempló tanto el acceso local como remoto, facilitando pruebas de integración y posibles futuras implementaciones clínicas.

Una vez que los datos clínicos son normalizados, se envían al modelo predictivo mediante un servicio *web RESTful*, desarrollado en *Python* y desplegado en un servidor público, siendo las principales características de esta arquitectura:

- **API RESTful**

El servicio está disponible a través del protocolo HTTP, y admite solicitudes mediante los métodos **GET** y **POST**. Su principal punto de acceso es la ruta `/predict`, disponible en el puerto 8500, tanto en entornos locales como en redes públicas.

- **Respuesta estructurada:**

Cada solicitud enviada al servicio es procesada por el servidor, que realiza la inferencia del modelo y devuelve una respuesta en formato JSON con la siguiente estructura:

- Una **predicción binaria** (0 o 1) correspondiente al riesgo cardiovascular.
- Un **mensaje explicativo breve**.
- Un **mensaje extendido** con recomendaciones clínicas asociadas.

- **Validación y robustez**

El sistema incorpora mecanismos de validación del formato de entrada, transformación de tipos de datos, y manejo de errores. Esto garantiza la estabilidad del servicio, incluso ante entradas incompletas o malformadas.

- **Accesibilidad e integración**

Gracias a su arquitectura abierta, el servicio puede ser consumido desde múltiples plataformas, incluyendo formularios web (*HTML*), aplicaciones móviles, o como componente embebido en sistemas de historia clínica electrónica. Esta flexibilidad lo convierte en una herramienta apta tanto para uso individual como institucional.

Figura 4. Arquitectura del sistema predictivo



Descripción Funcional de la Aplicación desarrollada

La aplicación guía al usuario a través de una secuencia de pantallas donde se recopilan datos médicos relevantes para evaluar el riesgo de enfermedades cardíológicas. Al finalizar, se presenta un resultado basado en la predicción generada por el modelo entrenado (en las Figuras 5 y 6 se pueden observar una serie de pantallas que presenta la app desarrollada).

Figura 5. Muestra de serie de pantallas de la app desarrollada (Parte 1)

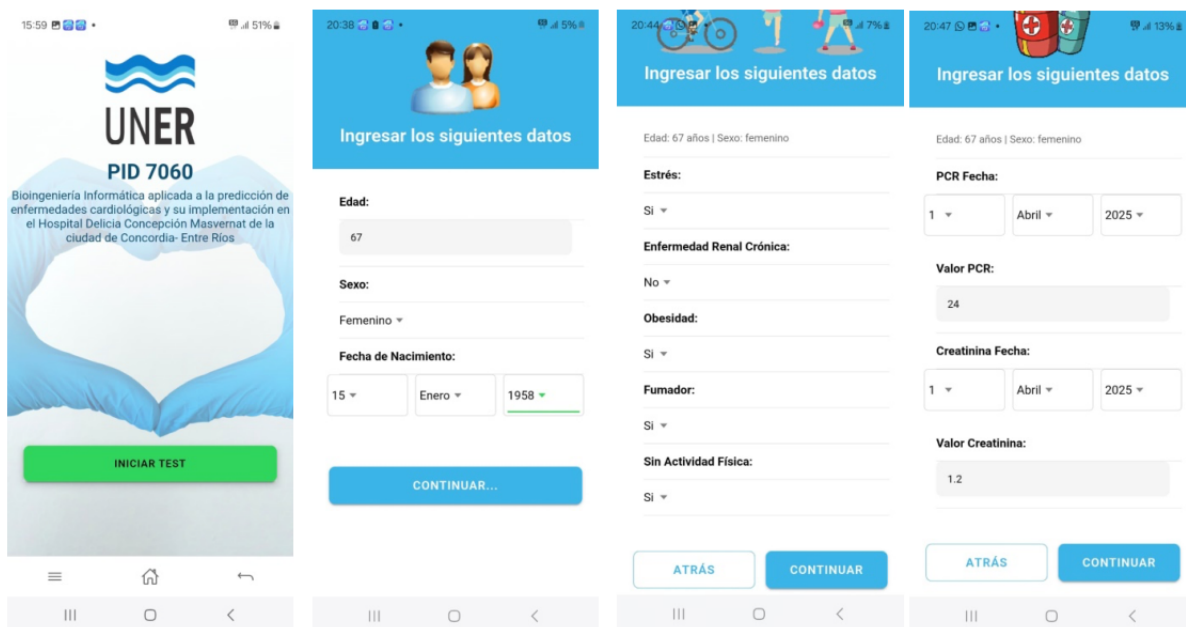
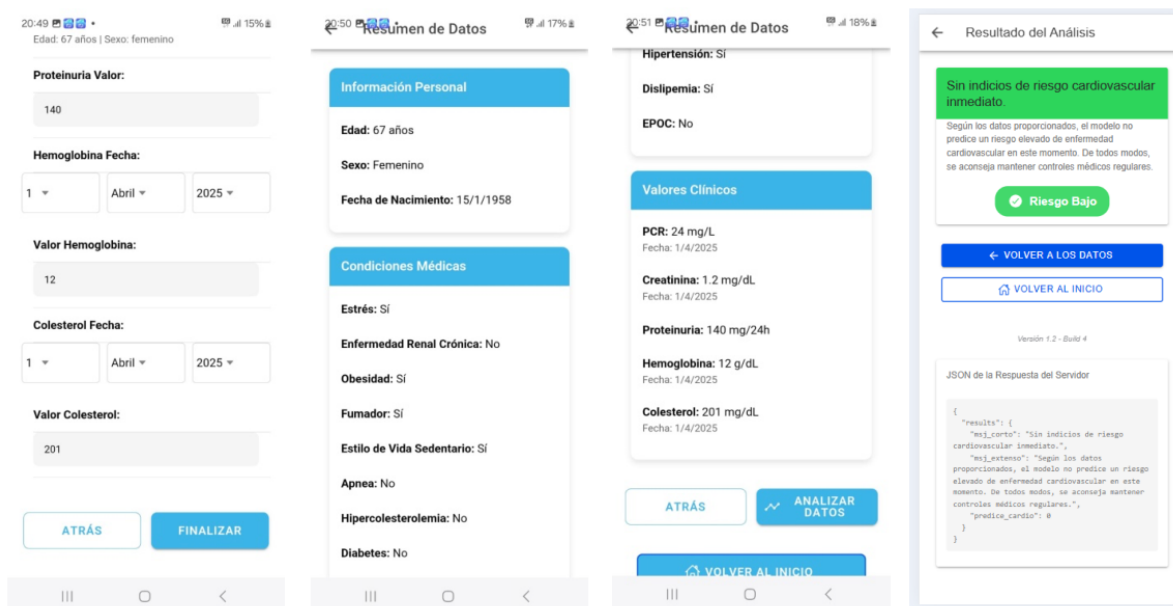


Figura 6. Muestra de serie de pantallas de la app desarrollada (Parte 2)



Utilidad clínica de la app

Este enfoque permite que los profesionales médicos -así como otros usuarios habilitados- puedan obtener una evaluación inmediata del riesgo cardiovascular a partir de datos clínicos básicos. De esta manera, el sistema contribuye activamente a fortalecer estrategias de medicina preventiva y apoyo a la toma de decisiones en escenarios de atención primaria o seguimiento ambulatorio.

Análisis de resultados

Se le fue solicitado al médico experto en cardiología que nos ha brindado cuatro casos para poder realizar las pruebas del modelo generado. Durante la evaluación de estos casos se observó que el modelo no clasificó adecuadamente los perfiles como de riesgo elevado, a pesar de presentar condiciones clínicas relevantes. Esta situación motivó un análisis más profundo del comportamiento del sistema y de los datos utilizados.

El modelo mostró alta incidencia en las variables que indican “patologías previas”. Debido a que la fuente de datos del hospital contiene muy pocas patologías catalogadas debido al modelo de atención que posee. El algoritmo aprendió a asociar de forma desproporcionada la presencia o ausencia de estas patologías con el riesgo, lo que provoca que cuando se incluyen las patologías, el modelo tiende a predecir en base a ellas, incluso si su peso clínico real es menor. Considerando los cuatro casos planteados, cuando se retiran las “patologías previas” dejando solamente las valoraciones clínicas, la predicción resultante es más coherente y representativa en términos generales, evitando que el modelo se sesgue excesivamente hacia patologías poco frecuentes, o escasamente registradas en la fuente de datos del entrenamiento. Se evidencia un sesgo hacia las pocas patologías que sí están registradas, por lo que el modelo no generaliza bien a casos donde las patologías difieren o no están en el conjunto de entrenamiento.

Es importante destacar que el desempeño del modelo está fuertemente condicionado por la calidad, cantidad y representatividad de la información disponible [21]. En este caso, la presencia de un número limitado de patologías registradas en los datos generó un sesgo hacia inexistencia de patologías, afectando la capacidad de generalización del modelo a otros contextos clínicos. Se observó que, al eliminar estas variables y considerando los cuatro casos planteados, la predicción se volvió más coherente y generalista, reduciendo la dependencia de patrones poco representativos.

La causa principal identificada es la distribución desigual y escasez de registros de patologías con la que contamos, lo que limita la capacidad del modelo para aprender patrones clínicos más amplios.

Para mitigar el problema planteado, se analizó considerar a futuro alternativas como:

1. Ampliar y balancear el registro de patologías en el conjunto de entrenamiento, asegurando que la información clínica proporcionada por el hospital sea más completa, consistente y representativa de la población atendida.
2. Aplicar técnicas de regularización y selección de características para reducir la dependencia de variables poco representativas.
3. Evaluar el modelo con datos de otras fuentes para medir su generalización.

Conclusiones

Las tecnologías de inteligencia artificial y *machine learning* ofrecen una alternativa poderosa para abordar la detección temprana de enfermedades. A partir de tecnologías abiertas y accesibles (Python, Scikit-learn, Flask, entre otras), se logró construir un sistema funcional con capacidad de integración en distintos niveles institucionales.

Hoy en día, incluso instituciones con recursos limitados pueden acceder a bibliotecas de algoritmos robustos, servidores de aplicaciones *open source*, e incluso servicios web ya publicados para funciones específicas de IA. Este ecosistema digital democratiza el acceso a la innovación, haciendo posible su aplicación tanto en hospitales de alta complejidad como en centros de atención primaria.

La arquitectura adoptada, basada en principios de escalabilidad, modularidad y reutilización, se presenta como una opción viable para enfrentar uno de los mayores desafíos de la salud pública: la detección temprana de enfermedades cardiovasculares.

Esta solución no solo apunta a ser técnica y científicamente robusta, sino que también promueve un modelo de atención centrado en el paciente, anticipándose a eventos clínicos mediante el uso responsable y ético de la inteligencia artificial.

Indicadores de producción

Artículos publicados en revistas de difusión científica

Bioingeniería Informática Aplicada a la Predicción de Enfermedades Cardiológicas y su Implementación en el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, Provincia De Entre Ríos. XXII Workshop de Investigadores en Ciencias de la Computación WICC'2020 (Virtual). Junio 2020, Universidad Nacional de la Patagonia Austral.

Selección de algoritmos de preprocesamiento de datos del Hospital Delicia Concepción Masvernat (Concordia, provincia de Entre Ríos) que permita el desarrollo de un componente de software para predicción de enfermedades cardiológicas. (2022). Memorias De Las JAIIO, 8(5), 84-101.

Evaluation of data processing tools for a project on computer biomedical engineering applied to the prediction of cardiological diseases and its implementation at a hospital in Argentina, Book of industry papers, poster papers and abstracts of the CENTERIS 2024, ISBN 978-989-54617-6-9 E-book edition 2024 bySciKA.

Artículos aceptados para presentación en congresos

Aplicación de bioingeniería informática a la predicción de enfermedades cardiológicas 20° Congreso de Investigación en Salud Pública y VIII Congreso Latinoamericano y del Caribe de Salud Global, Cuernavaca, Morelos, México. (2025)

Predicción de enfermedades cardiológicas mediante inteligencia artificial: Integración de herramientas, modelos y desarrollo de una app de evaluación. XV Congreso Universitario Internacional sobre Contenidos, Investigación, Innovación y Docencia (CUICIID 2025), Madrid, España

Bibliografía

- [1] Carabio, A., Silva Layes, M.E.&Frola, F., (2016) Bioingeniería aplicada en el diagnóstico de enfermedades. 45 JAIIO, CAIS 2016, Séptimo Congreso Argentino de Informática y Salud, 240–248.
- [2] Gillespie, G. (2000) There's gold in them thar' databases. Health Data Management 8(11), 40–52.
- [3] Techentin, B. (2012) Big data and graph analytics in a health care setting. http://www.graphanalysis.org/SC12/03_Techentin.pdf. Online; accedido en 2018.
- [4] Ayad Ghany Ismaeel, D. Y. M. (2016) Effective data mining technique for classification cancers via mutations in gene using neural network. International Journal of Advanced Computer Science and Applications 7(7), 69–76.
- [5] Shantakumar Patil, Y. K. (2009) Intelligent and effective heart attack prediction system using data mining and artificial neural network. European Journal of Scientific Research Vol.31 No.4, 642–656.
- [6] Marr, B. (2016) Big Data In Healthcare: Paris Hospitals Predict Admission Rates Using Machine Learning. <https://www.forbes.com/sites/bernardmarr/2016/12/13/big-data-in-healthcare-paris-hospitals-predict-admission-rates-using-machine-learning/#1532778779a2>
- [7] Silva Layes, M.E.&Falappa, M., (2011) Sistemas de soporte a las decisiones clínicas. XIII Workshop de Investigadores en Ciencias de la Computación, 113-115
- [8] Kannel, W. B., McGee, D., & Gordon, T. (1976). A general cardiovascular risk profile: the Framingham Study. The American journal of cardiology, 38(1), 46-51.
- [9] Chaurasia, V., Pal, S. (2013). Data Mining Approach to Detect Heart Diseases. International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol, 2, 56-66. Recuperado de <https://www.researchgate.net/publication/258374505>
- [10] Dimopoulos, A. C., Nikolaidou, M., Caballero, F. F., Engchuan, W., Sanchez-Niubo, A., Arndt, H., & Panagiotakos, D. B. (2018). Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. BMC medical research methodology, 18, 1-11.
- [11] Kannel, W. B., McGee, D., & Gordon, T. (1976) A General Cardiovascular Risk Profile: The Framingham Study. A general cardiovascular risk profile: the Framingham Study. The American journal of cardiology, 38(1), 46-51.

- [12] Clínica Mayo (2023). Hipertrofia ventricular izquierda Recuperada de <https://www.mayoclinic.org/es-es/diseases-conditions/left-ventricular-hypertrophy/symptomscauses/syc-20374314>
- [13] Weng, S. F., Reys, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one*, 12(4), e0174944.
- [14] Batista, G. E., & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied artificial intelligence*, 17(5-6), 519-533.
- [15] Hasan, S. M. M., Mamun, M. A., Uddin, M. P., & Hossain, M. A. (2018). *Comparative analysis of classification approaches for heart disease prediction*. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), pp.1-4. IEEE.
- [16] Dangare, C. S., & Apte, S. S. (2012). *Improved study of heart disease predictions system using data mining classification techniques*. *International Journal of Computer Applications*, 47(10), 44-48.
- [17] Suhendra, R., Husdayanti, N., Suryadi, S., Juliwardi, I., Sanusi, S., Ridho, A., & Ikhsan, I. (2023). *Cardiovascular Disease Prediction Using Gradient Boosting Classifier*. *Infolitika Journal of Data Science*, 1(2), 56-62.
- [18] Yang, J. (2024). *The prediction and analysis of heart disease using XGBoost algorithm*. *Proceedings of the 2023 International Conference on Machine Learning and Automation*, 41, 61-68. <https://doi.org/10.54254/2755-2721/41/20230711>
- [19] XGBoost Documentation. (s.f.). (2023) <https://xgboost.readthedocs.io/en/stable/Online>; accessed on 2023
- [20] Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- [21] Cai, Y., Gong, D., Tang, L., Cai, Y., Hui-Jun, J., Tian-Ci, G., Mengchun, H., Wei, Z., Zhen-Wei, Z., Xingang, Z. & Guang-Wei, Z. (2024). *Pitfalls in Developing Machine Learning Models for Predicting Cardiovascular Diseases: Challenge and Solutions*. *Journal Of Medical Internet Research*. vol. 26. e47645. p. 1
- [22] Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. <https://doi.org/10.1145/2939672.2939785>
- [23] Nielsen, D. (2016). *Tree boosting with XGBoost - Why does XGBoost win "every" machine learning competition? (Master's thesis, Norwegian University of Science and Technology)*. <http://hdl.handle.net/11250/2433761>
- [24] Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. *Advances in Neural Information Processing Systems*, 30. <https://arxiv.org/abs/1705.07874>
- [25] Tarabanis C, Kalampokis E, Khalil M, Alviar CL, Chinitz LA, Jankelson L (2023). *Explainable SHAP-XGBoost models for in-hospital mortality after myocardial infarction*. *Cardiovasc Digit Health J*. 2023 Jun 14;4(4):126-132. doi: 10.1016/j.cvdhj.2023.06.001. PMID: 37600443; PMCID: PMC10435947.

PID 7060

Denominación del Proyecto

Bioingeniería informática aplicada a la predicción de enfermedades cardiológicas y su implementación en el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, provincia de Entre Ríos

Unidad de ejecución

Universidad Nacional de Entre Ríos

Dependencia

Facultad de Ciencias de la Administración

Cátedra, Área o disciplina científica

Lenguajes de Programación, Sistemas Operativos, Taller de Investigación, Programación Orientada a Objetos, Bases de Datos

Instituciones intervinientes públicas o privadas

a) Acuerdo de Colaboración Académica entre el Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur y la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos

b) Acta Compromiso entre la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos y el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, provincia de Entre Ríos.

c) Convenio Específico entre la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos y el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, provincia de Entre Ríos para el desarrollo del proyecto de investigación denominado "Bioingeniería informática aplicada a la predicción de enfermedades cardiológicas y su implementación en el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, provincia de Entre Ríos".

Contacto: elizabeth.silva@uner.edu.ar

Integrantes del proyecto

Director/a

Silva Layes, María Elizabeth

Integrantes internos/docentes UNER

Benedetto, Marcelo Gabriel (FCAD-UNER); Benitez, Duval Horacio (FCAD-UNER); Carabio, Ana Lía Ramona (FCAD-UNER); Aguirre, Juan José (FCAD-UNER); Costen, Elio Darío (FCAD-UNER). Integrante estudiante de posgrado: Frola Da Silva, Jesús Fabian(FCAD-UNER)

Becario de formación vinculado al PID: Diez, Joaquin

Fecha de iniciación y finalización efectivas

01/10/2020 y 22/02/2025

Aprobación del Informe Final mediante resolución C.S. N°427/25 (12-12-2025)