

# Minería de Datos Aplicada a la Estimación del Riesgo Cardiovascular

Mtra. Norma Karen Valencia<sup>1</sup>, Mtro. Edgar Corona Organiche<sup>2</sup>, Dr. Abraham Jorge Jiménez Alfaro<sup>3</sup>, Dra. Griselda Cortés Barrera<sup>4</sup>.



## Acerca de los autores...

<sup>1</sup>Académico del Tecnológico de Estudios Superiores de Chimalhuacán.

<sup>2,3 y 4</sup>Académico del Tecnológico de Estudios Superiores de Ecatepec.

## Resumen

Las enfermedades cardiovasculares (ECV) son causantes de la mayoría de las muertes en el mundo, aproximadamente 30%, por lo que es importante estimar el riesgo de desarrollar este tipo de padecimiento con fines de prevención y reducción del índice de mortalidad. El presente trabajo expone el resultado de una Red Neuronal para la Estimación de Riesgo Cardiovascular, tomando como entrada

siete factores y como salida tres niveles: alto, intermedio y bajo. La evaluación de algoritmos permite identificar la mejor solución para estimar el riesgo cardiovascular; en primer término, mostraremos una red neuronal compuesta por seis entradas, ocho capas ocultas y una salida, que es comparada con la aplicación del algoritmo Naive Bayes, al igual que árboles de decisión y una máquina de soporte vectorial (SVM), mostrando la matriz de confusión de cada uno de los algoritmos implementados. Como resultado, se presenta una comparación de precisión entre los cuatro modelos, teniendo en referencia la estratificación de riesgo cardiovascular por medio de la tabla Framingham, y utilizando la matriz de confusión para validar los datos, la cual permite establecer el grado de error en cada uno ellos.

**Palabras clave:** Árboles de Decisión, Tabla Framingham, Máquina de Soporte Vectorial, Naive Bayes, Red Neuronal.

## Abstract

*Cardiovascular diseases (CVD) are the cause of the majority of deaths in the world, approximately 30%, so it is important to estimate the risk of developing this type of disease for the purposes of prevention and reduction of the mortality rate. The present work exposes the result of a Neural Network for Cardiovascular Risk Estimation, taking seven factors as input and three levels as output: high, intermediate and low. The evaluation of algorithms allows identifying the best solution to estimate cardiovascular risk; In the first place, we will show a neural network composed of six inputs, eight hidden layers and one output, which is compared with the application of the Naive Bayes algorithm, as well as decision trees and a support vector machine (SVM), showing the matrix confusion of each of the implemented algorithms. As a result, a precision comparison between the four models is presented, taking into account the stratification of cardiovascular risk through the Framingham table, and using the confusion matrix to validate the data, which allows establishing the degree of error in each one them.*

**Keywords:** Decision Trees, Framingham Table, Support Vector Machine, Naive Bayes, Neural Network.

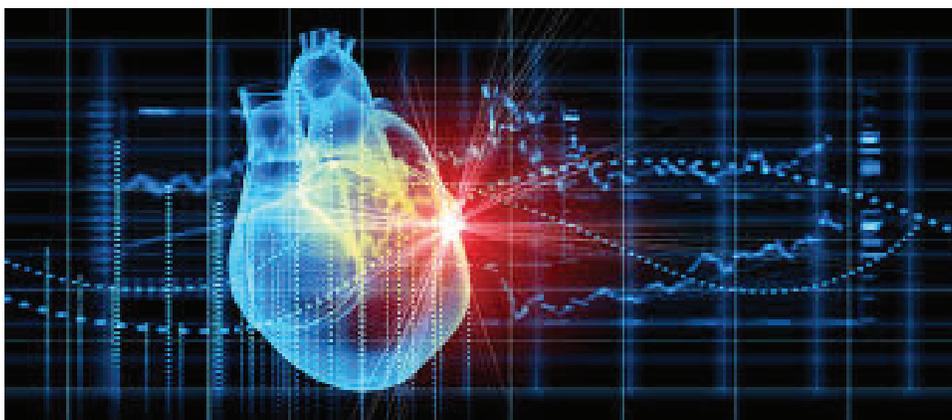
## Introducción

La Organización Mundial de la Salud (OMS) define como Enfermedad Cardiovascular (ECV) a todos aquellos padecimientos que generan desórdenes del corazón y de los vasos sanguíneos, entre los que se incluyen: la cardiopatía coronaria que se describe como una enfermedad de los vasos sanguíneos que irrigan el músculo cardíaco.

Se incluyen también las siguientes enfermedades cerebrovasculares: enfermedades de los vasos sanguíneos que irrigan el cerebro; las arteriopatías periféricas: enfermedades de los vasos sanguíneos; la cardiopatía reumática, cardiopatías congénitas generadas por malformaciones del corazón presentes desde el nacimiento, agregando a las trombosis venosas profundas y embolias pulmonares, que pueden desprenderse (émbolos) y alojarse en los vasos del corazón y los pulmones (OMS, 2017).

En México se estima que las ECV representan el 20% del total de muertes en los adultos. De acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI), en 2016 se reportaron 136,342 fallecimientos a causa de enfermedades del corazón, un aumento de 7,611 muertes respecto al 2015. Las causas de muerte incluyen enfermedad isquémica del corazón, enfermedades cerebrovasculares, hipertensivas, entre otras (OMENT, 2018).

La valoración de riesgos cardiovasculares ha generado diversos estudios, desde el año 2015 el Gobierno Federal proporcionó el “Cuestionario de Factores de



Riesgo” para ser utilizado como instrumento para determinar el Riesgo Cardiovascular, basado en su estratificación mediante la tabla Framingham.

## I. Problemática

El 25 de septiembre de 2015 se desarrolló el documento titulado “Transformar Nuestro Mundo: la Agenda 2030 para el Desarrollo Sostenible”, adoptado por países miembros de las Naciones Unidas, entre ellos México, y uno de los objetivos del documento

es desarrollar estrategias para prevenir las Enfermedades No Transmisibles, de las cuales forman parte las ECV (ONU, 2019).

La entidad de estudio posee históricos de más de 10 años, información concentrada en las Cédulas de Microdiagnóstico de su población. Sin embargo, no cuenta con una herramienta de análisis de información que le permita estimar o cuantificar la morbilidad, al igual que la mortalidad generada por las enfermedades cardiovasculares.

Aunado a lo anterior, la institución de salud tampoco cuenta con un modelo que le permita estimar el riesgo del desarrollo de enfermedades cardiovasculares en la población que atiende. Asimismo, carece del nivel de correlación que existe entre los diferentes los factores riesgo que influyen al desarrollo de este tipo de enfermedades.

Al iniciar las campañas de prevención, no se capta en su totalidad la población mencionada, puesto que se omite el análisis exhaustivo de datos y factores de riesgo por cada individuo, lo que genera un informe impreciso de la población vulnerable a dichas enfermedades.

## II. Marco Teórico

La función clásica de Framingham estima el riesgo de padecer un acontecimiento en los próximos 10 años, considerando como acontecimiento la muerte de origen coronario, el infarto agudo de miocardio no fatal, la angina estable o la angina inestable (insuficiencia coronaria).

El National Cholesterol Education Program, en su último documento *Adult Treatment Panel III (ATP-III)* ha modificado esta función para calcular exclusivamente el riesgo de los denominados acontecimientos “duros”, es decir, infarto agudo de miocardio no fatal y muerte coronaria, excluyendo de esta estimación a los diabéticos, al considerarlos directamente como de riesgo alto (Pacheco A. M. and Jaquez T. J., 2017; Gonzalo C. V., Ruiz-Villaverde G., y Gracia M. A. 2009).

Derivado de lo anterior, en la Tabla 1 se resumen los riesgos estudiados de eventos cardiovasculares, pero el nivel de riesgo “Muy Alto” no se toma en cuenta, ya que se engloba dentro del riesgo alto, todos ellos pertenecen a la estratificación basada en la tabla Framingham, que es utilizada en el Sector Salud por medio de la Guía de Práctica Clínica: Detección y Estratificación de Factores de Riesgo Cardiovascular (Secretaría de Salud, 2010).

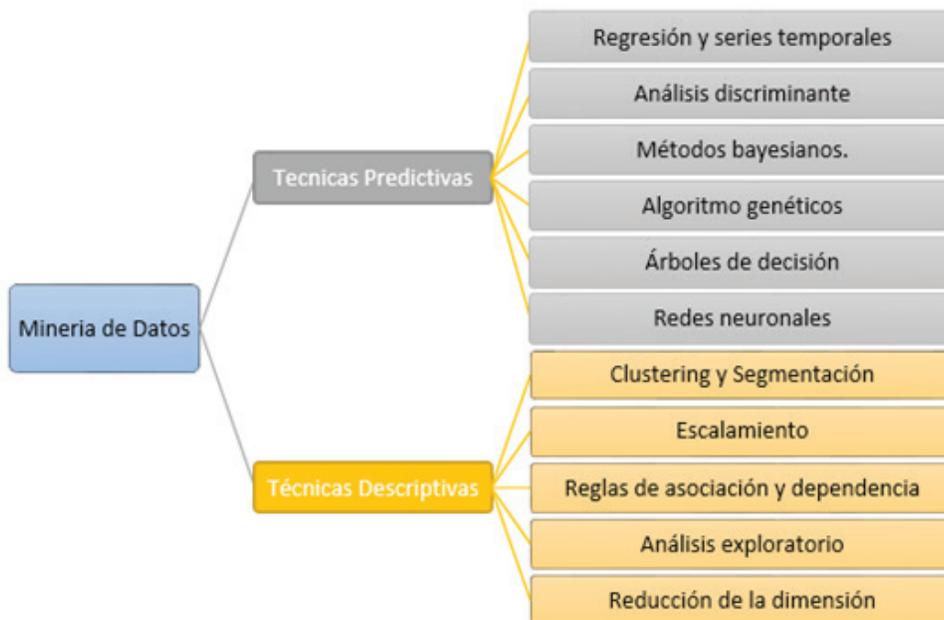
**TABLA 1**  
**RIESGO DEL EVENTO CARDIOVASCULAR EN UN PERIODO DE 10 AÑOS**

NIVEL DE RIESGO	PORCENTAJE DE ESTIMACIÓN DE RIESGO A 10 AÑOS	AÑOS DE POSIBLE EVENTO CARDIOVASCULAR
<b>Bajo</b>	<15%	Mayor a 7 años
<b>Intermedio</b>	15-20%	Mayor a 4 años
<b>Alto</b>	>20%	Menor a 5 años

El Cuestionario de Factores de Riesgo consta de las siguientes variables para detectar estos factores: Diabetes, Glucemia, Hipertensión Arterial, Cáncer (colorectal, pulmonar, bucal, gástrico), Enfermedades cardiovasculares (cardiopatías, embolias e hipertensión). Para el presente estudio, se toman las mostradas en la Tabla 1.

Las aplicaciones comúnmente desarrolladas con análisis predictivo son: predecir riesgos, predecir activación de nuevos clientes, predicción de ventas, entre otras. Este tipo de análisis se caracteriza porque requiere un conjunto de entrenamiento, el cual está formado por un histórico de datos. En el análisis predictivo se pueden desarrollar tareas de predicción discreta y predicción continua (Lanzarini L. C. *et al.*).

Dentro de la minería de datos, los clasificadores son capaces de obtener la incidencia de eventos cardiovasculares. Los clasificadores bayesianos permiten clasificar eventos discretos y limitados (variables independientes) en un número determinado de clases, definiendo una función estadística para cada una. En la Figura 2, se muestra el gráfico con su clasificación.



*Figura 1*

Clasificación de las técnicas de Minería de Datos.

Las redes bayesianas son un modelo probabilístico por medio del cual es factible construir un grafo entre las causas de un evento (Variables Independientes) y sus consecuencias (Variables Dependientes) (Castrillón O. D., Sarache W., and Castaño E., 2017).

### III. Marco Metodológico

Existen diversas metodologías que proporcionan una serie de pasos a seguir con el fin de realizar una implementación adecuada de la minería de datos. Según sondeos publicados (KDnuggets, 2014), las metodologías más utilizadas son: CRISP-DM (Cross Industry Standard Process for Data Mining), SEMMA, KDD y Catalyst. Se puede consultar esta información en la liga <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>

La CRISP-DM es una metodología de libre distribución que puede trabajar con cualquier herramienta para desarrollar cualquier proyecto, estructura el ciclo de vida de un proyecto de Minería de Datos, sus fases descritas en la Tabla 2, interactúan entre ellas de forma iterativa durante el desarrollo del proyecto y de forma neutra a la herramienta que se utilice para el desarrollo del proyecto (Carpio J. C., 2016).

**TABLA 2**  
DESCRIPCIÓN DE FASES: METODOLOGÍA CRISP-DM

Fase	Descripción
Comprensión del Negocio o Entendimiento	Reúne las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto (Betancur S., 2019).
Comprensión de los Datos	Comprende la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema (Huber S, <i>et al.</i> , 2019).
Preparación de los Datos	Consta de algunas tareas, como la selección de datos escogiendo un subconjunto de los datos recopilados en la etapa anterior. Limpieza de los datos, preparándolos para la fase de modelación, ya sea aplicando técnicas de normalización, discretización de campos numéricos, tratamiento de valores nulos, entre otros (Schäfer J. B. y H. O. F., Zeiselmaier C., 2018).
Modelado	Se sugiere realizar una comparación de las técnicas de modelado que sean más apropiadas para resolver el problema (Wirth R, 2000).
Evaluación	Se evalúa el modelo con base en el cumplimiento de los criterios de éxito del problema, revisar el proceso seguido, teniendo en cuenta los resultados obtenidos (Galán V., 2015).
Implementación	Esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación (Galán V., 2015).

## IV. Desarrollo Metodológico

### A. Comprensión del Negocio o Entendimiento

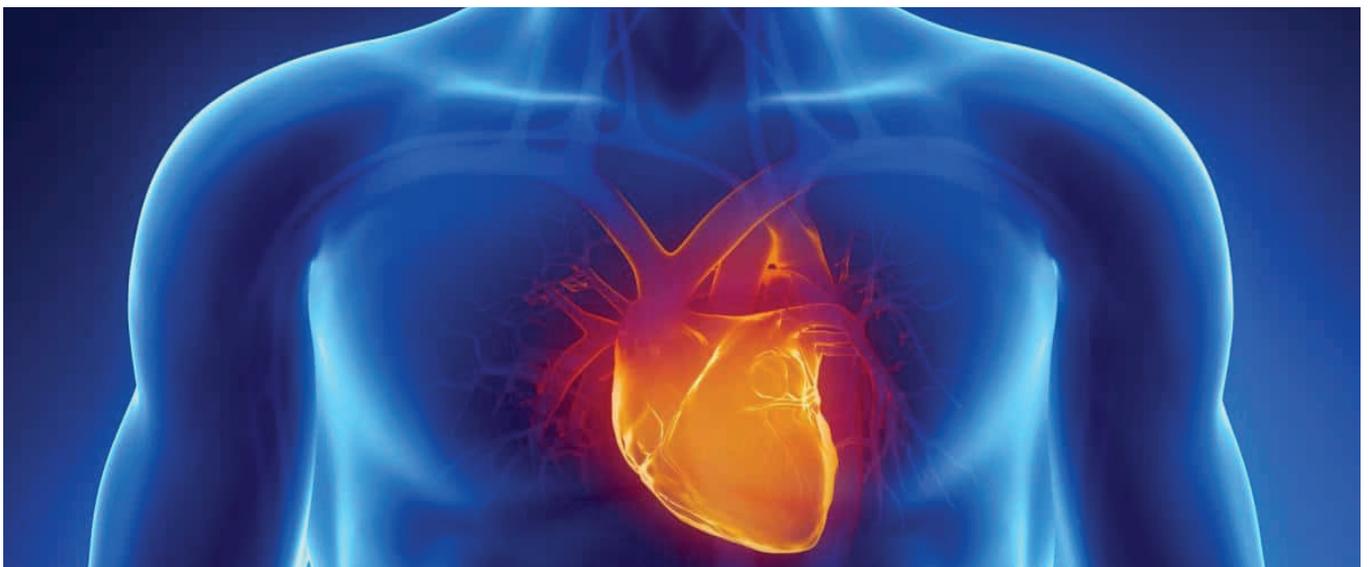
La tabla Framingham trabaja con nueve factores de riesgo, al igual que otras tablas de estratificación de riesgos, como lo es Systematic Coronary Risk Evaluation (SCORE), que mide la mortalidad por patología cardiovascular a 10 años en población europea de 40 a 65 años (Lozano Merino P. and Cid Galán M. L., 2016). El presente estudio se concentra en la primera tabla y en la práctica de guía clínica para la estratificación de riesgo cardiovascular.

**TABLA 3**  
FACTORES DE RIESGO UTILIZADOS EN LOS MODELOS

Factor de riesgo	Red Neuronal	Naive Bayes	Árboles de Decisión
Edad	X	X	X
Presión diastólica	X	X	X
Presión sistólica	X	X	X
Diabetes	X	X	X
Hipertensión	X	X	X
Herencia	X	X	X

### B. Comprensión de los Datos

Se tomaron en cuenta nueve factores de riesgo que se presentan en la Tabla 3; se descartaron cuatro características, en primer lugar un factor no modificable, como es el sexo, y tres elementos modificables, los cuales son peso, tabaquismo y actividad física.



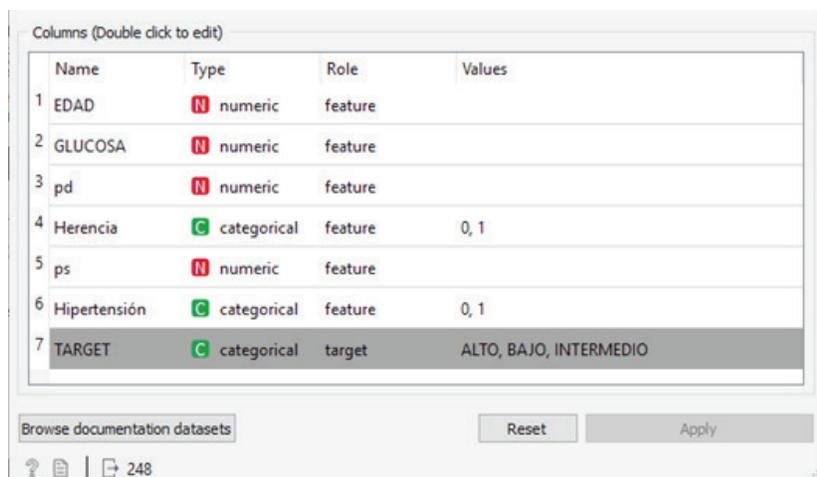
**TABLA 4**  
REVISIÓN DE VARIABLES EN DIVERSAS TABLAS DE ESTRATIFICACIÓN DE RIESGO CARDIOVASCULAR

Variables	Cédula de Microdiagnóstico	SCORE	Framingham	PCE
Edad	*	*	*	*
Sexo	*	*	*	*
Nivel de colesterol	Sólo personas valoradas	*	*	*
Presión Sistólica mmg	Sólo personas valoradas	*	*	*
Tabaquismo	*	*	*	*
Diabetes	*	*	*	*
Dieta	*	*	*	*
Actividad Física	*	*	*	*
Peso	*	*	*	*

Los factores de riesgo que no se tomaron en cuenta, poseen una estrecha relación con los que se utilizaron como características; en el caso del factor de riesgo “peso” está ligado a la presión sistólica y diastólica, así como a la hipertensión y diabetes.

### C. Preparación de los Datos

Para que el modelo funcione óptimamente, los datos deben estandarizarse cumpliendo con las características, se establecen tres categorías Alto, Intermedio y Bajo como salida, de entradas se toman seis características conteniendo un total de 248 registros, los datos de entrada al igual que de salida se muestra en la Figura 3.



**Figura 3**

Características de entrada y salida de los modelos.

## D. Modelado

La arquitectura de la red Neuronal consta de seis entradas, ocho capas ocultas con ocho neuronas y tres salidas, como se muestra en la Figura 4.

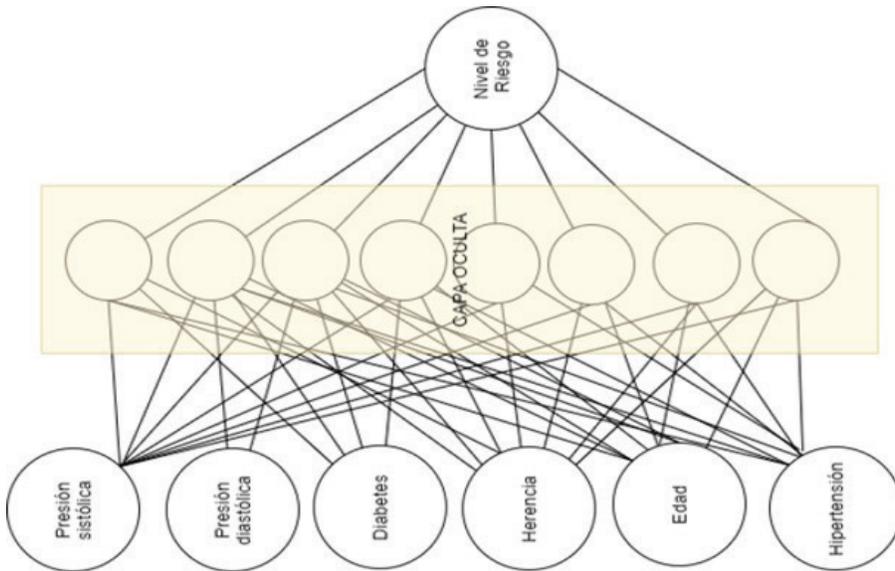


Figura 4

Arquitectura Red Neuronal

### 1) Red Neuronal

Las características esenciales de una red neuronal son los nodos (organizados en capas), la arquitectura de la red, que describe la conexión entre los nodos, y el algoritmo utilizado para encontrar los valores de los parámetros de la red (pesos) (Caicedo E., 2018).

Las capas de una red, pueden ser: a) De entrada, conformadas por las neuronas que introducen los seis factores de riesgo; en estas neuronas no se realiza procesamiento. b) Oculta, conformadas por las neuronas cuyas entradas vienen de la capa anterior. y c) De salida, formada por neurona que indica el nivel de riesgo.

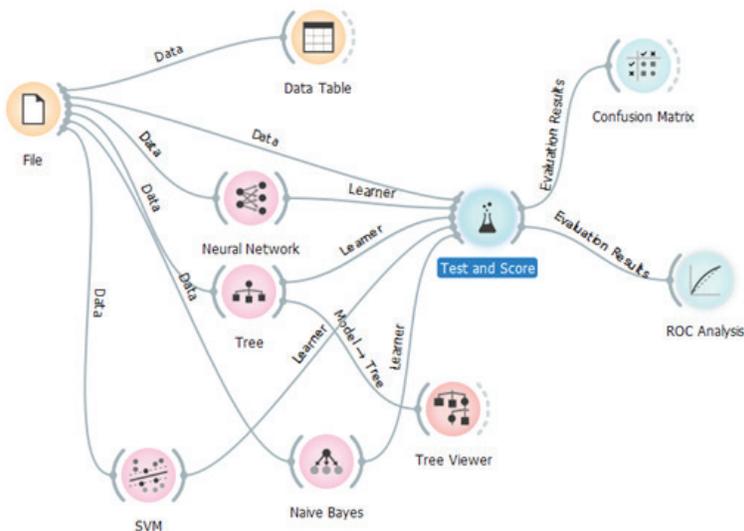


Figura 5

Diagrama Generado en Orange Versión 3.26.0

El método de optimización utilizado es Adam, que solo requiere gradientes de primer orden con pocos requisitos de memoria. El método calcula las tasas de aprendizaje adaptativo individual para diferentes parámetros a partir de las estimaciones del primer y segundo momento de los gradientes; el nombre Adam, se deriva de la estimación adaptativa del momento (Kingma D. P. and Ba J. L., 2015).

Se utilizó el software Orange Datamining como herramienta de análisis de datos. En la ilustración se muestran los widgets empleados para trabajar con los modelos.

## 2) Algoritmo Naive Bayes

El modelo que se plantea utiliza el algoritmo Naive Bayes; toma como base la edad y posteriormente los demás factores, generando la probabilidad de sufrir padecimientos cardiovasculares. En la siguiente tabla se muestra las probabilidades obtenidas.

**TABLA 5**  
PROBABILIDADES DE RIESGO CARDIOVASCULAR POR MEDIO DEL TEOREMA DE BAYES

Factor de Riesgo	Riesgo de Padecimiento Cardiovascular (RECV)	Descripción
Edad	$P(RECV E)$	Probabilidad de riesgo cardiovascular, tomando como factor de riesgo la edad.
Presión diastólica	$P(RECV E \cap Pd)$	Probabilidad de riesgo cardiovascular, tomando como factor de riesgo la edad y la presión diastólica
Presión sistólica	$P(RECV E \cap Pd \cap Ps)$	Probabilidad de riesgo cardiovascular, tomando como factor de riesgo la edad, la presión diastólica y presión sistólica.
Diabetes	$P(RECV E \cap Pd \cap Ps \cap D)$	Probabilidad de riesgo cardiovascular, tomando como factor de riesgo la edad, la presión diastólica, presión sistólica y la diabetes.
Hipertensión	$P(RECV E \cap Pd \cap Ps \cap D \cap Hip)$	Probabilidad de riesgo cardiovascular tomando como factor de riesgo la edad, la presión diastólica, presión sistólica, diabetes e hipertensión.
Herencia	$P(RECV E \cap Pd \cap Ps \cap D \cap Hip \cap Her)$	Probabilidad de riesgo cardiovascular tomando como factor de riesgo la edad, la presión diastólica, presión sistólica, diabetes, hipertensión y herencia.

## 3) Árboles de Clasificación

Éstos suponen una alternativa a las técnicas estadísticas más clásicas, como la regresión múltiple, los análisis ANOVA, regresión logística, análisis discriminante y modelos de supervivencia. Además, los árboles de clasificación parecen obtener

mejores tasas predictivas que el resto de las técnicas de minería de datos cuando se utiliza información mayoritariamente de tipo categórico (Ortiz-Lozano J. M., Rua Vieites A., and Bilbao Calabuig P., 2017). La Figura 6 muestra la arquitectura del árbol de decisión utilizado.

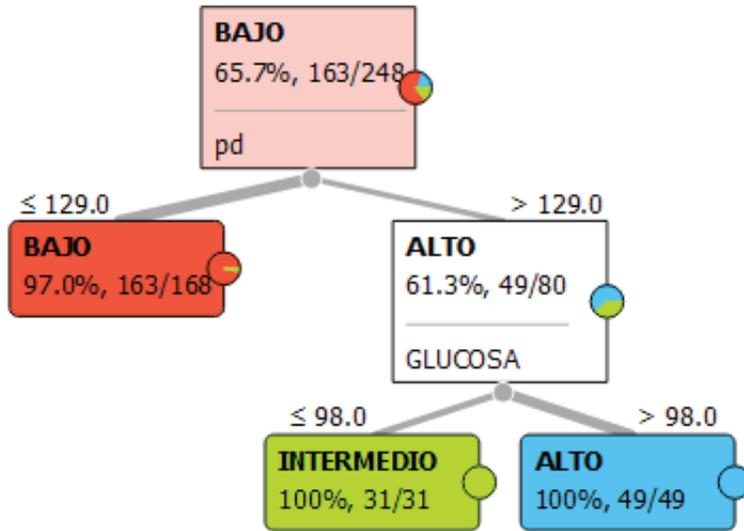


Figura 6  
Arquitectura del Árbol

#### 4) Máquinas de Soporte Vectorial

Este modelo utiliza la técnica de máquinas de soporte vectorial (SVM, por sus siglas en inglés) para realizar hiperplanos que dividen la población entre usual e inusual (en nuestro estudio riesgo bajo, intermedio y alto), a partir de la población de entrenamiento, con el cual se logra clasificar la población de validación (Gracia Granados M. E., 2019).

Para la configuración de la SVM se elige un costo de 1 y una pérdida de 0.10 además de un kernel polinomial.

### V. Resultados.

Los datos que fueron tratados para este estudio fueron de 248 pacientes, correspondientes a una sección del Centro de Salud estudiado. La Figura 7 proporciona una muestra de los datos introducidos a los modelos atendidos en el presente trabajo.

	TARGET	EDAD	GLUCOSA	pd	Herencia	ps	Hipertensión
1	ALTO	47.0	101.0	134.0	0	81.0	1
2	BAJO	56.0	98.0	126.0	0	86.0	0
3	BAJO	55.0	95.0	125.0	0	76.0	0
4	BAJO	39.0	89.0	123.0	0	83.0	0
5	INTERMEDIO	66.0	113.0	115.0	0	80.0	0
6	BAJO	42.0	98.0	102.0	0	70.0	0
7	BAJO	40.0	440.0	123.0	1	87.0	0
8	BAJO	35.0	97.0	122.0	0	83.0	0
9	BAJO	46.0	104.0	126.0	0	83.0	0
10	INTERMEDIO	36.0	91.0	132.0	0	83.0	1
11	BAJO	55.0	92.0	117.0	0	81.0	0

Figura 7  
Muestra parcial de los datos introducidos a los modelos

En la Figura 9 se proporcionan las matrices de confusión que produce cada modelo desarrollado. Cada una de ellas proporciona la precisión con la que el modelo predice cada nivel de riesgo.

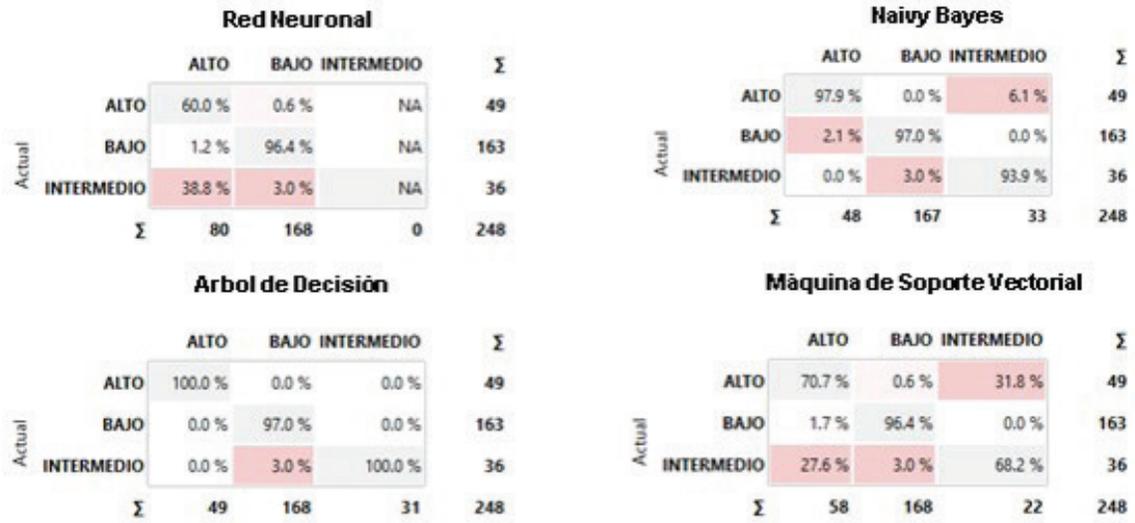


Figura 9

Matrices de confusión obtenidas para la Red Neuronal, Naivy Bayes, Árbol de Decisión y Máquina de Soporte Vectorial

En las matrices de confusión se puede observar que el perceptrón no identificó correctamente el nivel de Riesgo Intermedio, de tal forma, que solo trabajó en dos niveles de riesgo: bajo y alto. Asimismo, se puede observar que el mejor estimador lo constituye el árbol de decisión, presentando una eficacia del 100% para el nivel de riesgo alto e intermedio.

Una manera más global de conocer la calidad de la prueba en el espectro completo de puntos de corte, es mediante el uso de curvas ROC (Receiver Operating Characteristics, características operativas del receptor) que, constituyen una herramienta fundamental y unificadora en el proceso de evaluación, ver Figura 10 (Diez Herranz A. and Tobal González M., 1997).

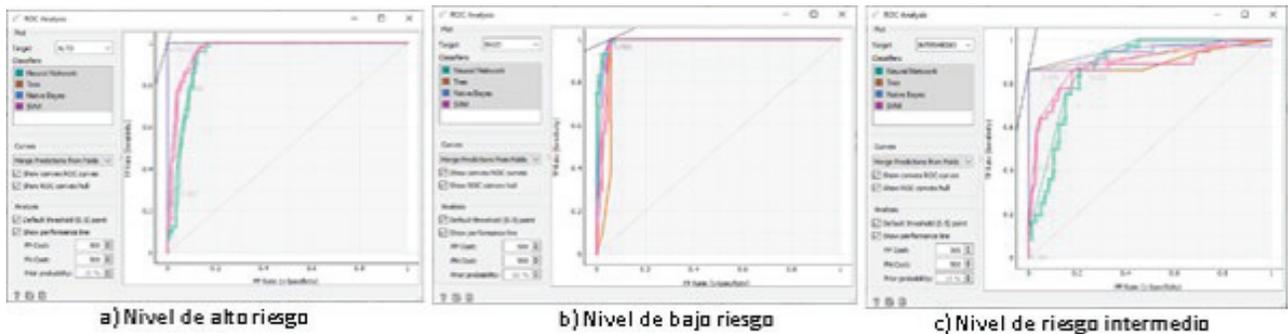


Figura 9

Curvas ROC de los modelos para el nivel de riesgo alto, bajo e intermedio

Con base en el análisis de las matrices de confusión y de las curvas ROC se construye la Tabla 6, donde se interpreta la eficacia de cada modelo para estimar el nivel de riesgo. Se puede observar y confirmar que el árbol de decisión es el que entrega mejores estimaciones, con una precisión de 0.980, y que la red neuronal que se configura con el modelo Adam y un algoritmo ReLu es la menos efectiva, con una precisión de 0.875.

**TABLA 6**  
RESULTADOS DE LOS MODELOS

Modelo	*AUC	Precisión de Clasificación	Precisión	Especificidad
Naive Bayes	0.977	0.976	0.976	0.961
Neural Network	0.944	0.843	0.745	0.907
SVM	0.943	0.875	0.871	0.930
Tree	0.955	0.980	0.980	0.961

\*AUC= Area under ROC curve

## Conclusiones

Tomando como base la información que recaban las entidades de salud, los pacientes al enterarse del nivel de riesgo de enfermedad cardiovascular, se prevé que tomen ciertas medidas de cuidado, al igual que se les invita a campañas de control y prevención de dichas enfermedades. El uso de herramientas que permitan el análisis de datos de modelos para la estimación de riesgo, apoya a que la investigación pueda mejorar, de acuerdo con los resultados del perceptrón multicapa, cuyo resultado estuvo por debajo del Árbol de Clasificación y del Algoritmo Naive Bayes. Los árboles de clasificación son una buena elección para determinar el nivel de riesgo cardiovascular; tomando los resultados de la matriz de confusión, el error de dicho modelo respecto a clasificar en riesgo intermedio es del 3% en riesgo bajo, similar a los resultados de Naive Bayes. Los resultados de los modelos son comparables y podrían examinarse otros modelos; cabe mencionar que a la red neuronal se le pueden agregar capas ocultas, implementando otras funciones de activación, mismas que pueden aplicarse a los modelos presentados.



## Referencias

- Betancur S., 2019, "Inteligencia de negocios aplicada al ecoturismo en Colombia: Un caso de estudio aplicando la metodología CRISP-DM," IEEE Comput. Soc.
- Caicedo E., 2018, "Modelos para otorgamiento y seguimiento en la gestión de riesgo de crédito", *Rev. Métodos Cuantitativos para la Econ., y la Empres.*
- Carpio J. C., 2016, "Modelo de Predicción de la Morosidad en el Otorgamiento de Crédito Financiero Aplicando Metodología Crisp-DM".
- Castrillón O. D., Sarache W., and Castaño E., 2017, "Sistema bayesiano para la predicción de la diabetes," *Inf. Tecnol.*, Vol. 28, No. 6, pp. 161-168.
- Díez Herranz A. and Tobal González M., 1997, "Las curvas ROC en la evaluación de las pruebas diagnósticas," *Med. Clin. (Barc)*, Vol. 108, No. 1, pp. 34-35.
- Galán V., 2015, "Crisp-Dm a un Proyecto de Minería," p. 120, 2015.
- Gracia Granados M. E., 2019, "Máquinas de soporte vectorial y árboles de clasificación para la detección de operaciones sospechosas de lavado de activos," *Lámpasakos*, Vol. 21, pp. 26-38.
- Gonzalo Ruiz-Villaverde C. V.-G., Gracia María Abalos-Medina, 2009, "Estimación de riesgo cardiovascular en pacientes con síndrome metabólico (SM)," *Medwave*.
- Huber S., Wiemer H., Schneider D., and Ihlenfeldt S., 2019, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," *Procedia CIRP*, Vol. 79, pp. 403-408.
- KDnuggets, 2014, "CRISP-DM, sigue siendo la mejor metodología para proyectos de análisis, minería de datos o ciencia de datos." [On line]. Available: <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>.
- Kingma D. P. and Ba J. L., 2015, "Adam: A method for stochastic optimization," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1-15.
- Lanzarini L. C. *et al.*, 2018, "Minería de datos y big data: aplicaciones en riesgo crediticio, salud y análisis de mercado," pp. 350-354.
- Lozano Merino P. and Cid Galán M. L., 2016, "Papel de la enfermera en prevención y rehabilitación cardiaca. Revisión narrativa," Sección Dep. Enfermería, pp. 1-45.
- OMENT, 2018. "Un panorama de las enfermedades cardiovasculares," Observatorio Mexicano de Enfermedades No Transmisibles.
- OMS, 2017 "Enfermedades Cardiovasculares," *Enfermedades Cardiovasculares Datos y Cifras*.
- ONU, 2019. "Informe de los objetivos del desarrollo sostenible," p. 64,
- Ortiz-Lozano J. M., Rúa Vieites A., and Bilbao Calabuig P., 2017, "Aplicación de árboles de clasificación a la detección precoz de abandono en los estudios universitarios de administración y dirección de empresas," *Recta*, Vol. 18, No. 2, pp. 177-201.
- Pacheco A. M. and Jaquez T. J., 2017, "Prevalencia de síndrome metabólico en la consulta externa," *Rev Sanid Milit Mex*, Vol. 71, pp. 264-275.
- Schäfer J. B. y H. O. F., Zeiselmaier C., 2018, "Sintetizar CRISP-DM y gestión de calidad: un enfoque de minería de datos para procesos de producción," Conf. Int. IEEE 2018 sobre Gestión Tecnol. Operaciones y Decis.
- Secretaría de Salud, 2010, *Detección y Estratificación de factores de riesgo cardiovascular*.
- Wirth R., 2000, "CRISP-DM: Towards a Standard Process Model for Data Mining," Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min., No. 24959, pp. 29-39.