

Artículo de Investigación

Modelo predictivo de enfermedades cardiovasculares basado en Redes Bayesianas

Predictive Model of cardiovascular diseases based on Bayesian Networks

Dayron Rumbaut Rangel¹: Universidad Bolivariana del Ecuador (UBE), Ecuador.

drumbautr@ube.edu.ec

Milton Rafael Maridueña Arroyave: Universidad Bolivariana del Ecuador (UBE), Ecuador.

mrmariduenaa@ube.edu.ec

Fecha de Recepción: 01/09/2024

Fecha de Aceptación: 20/10/2024

Fecha de Publicación: 27/12/2024

Cómo citar el artículo:

Rumbaut Rangel, D. y Maridueña Arroyave, M. R. (2024). Modelo predictivo de enfermedades cardiovasculares basado en Redes Bayesianas [Predictive Model of cardiovascular diseases based on Bayesian Networks]. *European Public & Social Innovation Review*, 9, 01-22. <https://doi.org/10.31637/epsir-2024-1804>

Resumen

Introducción: Este estudio presenta un análisis y comparación de varios modelos de redes bayesianas para la predicción de enfermedades cardiovasculares, utilizando datos clínicos para identificar los modelos más efectivos. **Metodología:** Los algoritmos evaluados incluyen Naive Bayes, TAN_cl, TAN_hcsp, FSSJ, BSEJ y KDB, los cuales fueron entrenados y validados para medir su desempeño. Para evaluar el rendimiento se utilizó un conjunto de datos clínicos de pacientes, combinando cinco bases de datos públicas. Las métricas usadas fueron como precisión, sensibilidad, especificidad, F1 Score y además una validación cruzada para asegurar la consistencia de los resultados. **Resultados:** El modelo BSEJ presentó el mejor rendimiento en todas las métricas evaluadas, destacándose por su capacidad de eliminar dependencias irrelevantes, manteniendo así un equilibrio óptimo entre complejidad y precisión. **Discusión:** Las redes bayesianas ofrecen una potente herramienta para la predicción de ECV, ya que no solo proporcionan predicciones precisas, sino que también facilitan la interpretación de las relaciones entre los factores de riesgo, lo cual es crucial en el

¹ Autor Correspondiente: Dayron Rumbaut Rangel. Universidad Bolivariana del Ecuador (Ecuador).

ámbito clínico. **Conclusiones:** Las redes bayesianas, y en particular los modelos BSEJ y KDB, se destacan por su efectividad en la predicción de enfermedades cardiovasculares, proporcionando soporte a la toma de decisiones clínicas de manera informada.

Palabras clave: redes bayesianas; enfermedades cardiovasculares; prevención; modelos predictivos; BSEJ; KDB; naive bayes; aprendizaje automático.

Abstract

Introduction: This study presents an analysis and comparison of several Bayesian network models for cardiovascular disease prediction, using clinical data to identify the most effective models. **Methodology:** The algorithms evaluated include Naive Bayes, TAN_cl, TAN_hcsp, FSSJ, BSEJ and KDB, which were trained and validated to measure their performance. A clinical dataset of patients, combining five public databases, was used to evaluate performance. The metrics used were accuracy, sensitivity, specificity, F1 Score and also a cross validation to ensure consistency of results. **Results:** The BSEJ model presented the best performance in all the metrics evaluated, standing out for its ability to eliminate irrelevant dependencies, thus maintaining an optimal balance between complexity and accuracy. **Discussions:** Bayesian networks offer a powerful tool for CVD prediction, as they not only provide accurate predictions, but also facilitate the interpretation of relationships between risk factors, which is crucial in the clinical setting. **Conclusions:** Bayesian networks, and in particular the BSEJ and KDB models, stand out for their effectiveness in predicting CVD, providing informed clinical decision support.

Keywords: Bayesian networks; cardiovascular disease; prevention; predictive models; BSEJ; KDB; naive Bayes; machine learning.

1. Introducción

Las enfermedades cardiovasculares (ECV) en general se refiere a las condiciones que implican el estrechamiento o bloqueo de los vasos sanguíneos, causada por daño al corazón o a los vasos sanguíneos por aterosclerosis. Cada año mueren más personas por enfermedades cardiovasculares (ECV) que por cualquier otra causa. (*Enfermedades Cardiovasculares - OPS/OMS | Organización Panamericana de La Salud, n.d.*) Se estima que 17,9 millones de personas murieron por ECV en 2019, lo que representa el 32% de todas las muertes a nivel mundial. De estas muertes, el 85% se debieron a ataques cardíacos y accidentes cerebrovasculares. (*Cardiovascular Diseases (CVDs), n.d.*) En la Región de las Américas en el año 2019, se registró que 2.0 millones de personas murieron a causa de las enfermedades cardiovasculares (*La Carga de Enfermedades Cardiovasculares - OPS/OMS | Organización Panamericana de La Salud, n.d.*).

Las enfermedades cardiovasculares suelen ser causadas por una combinación de factores como la aterosclerosis, donde se acumula placa en las arterias, la hipertensión, y niveles elevados de colesterol, lo que contribuye al estrechamiento y endurecimiento de los vasos sanguíneos. Factores como la diabetes mal controlada, el tabaquismo, la obesidad y el sedentarismo también juegan un papel clave, al igual que una dieta poco saludable rica en grasas y sal. Además, el estrés crónico, el consumo excesivo de alcohol y los antecedentes familiares incrementan el riesgo cardiovascular. Con la edad, este riesgo aumenta, especialmente en hombres y mujeres postmenopáusicas.

En Ecuador, las enfermedades cardiovasculares son la primera causa de consultas en los establecimientos del Ministerio de Salud Pública (MSP). Entre 2018 y 2022 se registró un promedio anual de 247 000 primeras consultas y casi 1,5 millones de consultas subsecuentes.

También son la primera causa de muerte en el país, acumulando el 25% del total de decesos anuales (*Ecuador Acumula Pacientes Con Enfermedades Cardiovasculares | CEAP :: Centro de Estudios Asia-Pacífico, n.d.*).

Los modelos predictivos son herramientas poderosas en la prevención y manejo de enfermedades cardiovasculares, ya que permiten identificar a las personas con mayor riesgo antes de que desarrollen síntomas graves.[5] Estos modelos utilizan datos históricos y actuales de los pacientes, como la edad, presión arterial, niveles de colesterol, hábitos de vida (fumar, actividad física, dieta), antecedentes familiares y condiciones médicas como la diabetes. Con esta información, los modelos aplican algoritmos estadísticos y técnicas de aprendizaje automático para analizar patrones y predecir la probabilidad de que una persona desarrolle una enfermedad cardiovascular.

El uso de estos modelos permite una detección temprana de riesgos, lo que facilita la intervención a tiempo, como cambios en el estilo de vida o tratamientos médicos, para prevenir eventos graves como infartos o accidentes cerebrovasculares. Además, ayudan a los profesionales de la salud a tomar decisiones más informadas y personalizadas, optimizando la atención al paciente al enfocarse en quienes tienen mayor probabilidad de complicaciones. De este modo, los modelos predictivos contribuyen a mejorar la salud cardiovascular y a reducir la carga de estas enfermedades en la población.

El objetivo de esta investigación es analizar múltiples modelos de redes bayesianas para la predicción de enfermedades cardiovasculares, con el propósito de identificar el modelo que ofrezca la mayor precisión y eficacia en la estimación del riesgo, facilitando así la detección temprana y la toma de decisiones clínicas informadas para la prevención y el manejo de estas enfermedades.

1.1. Marco Teórico

1.1.1. El papel del aprendizaje automático en la predicción de enfermedades cardíacas

El aprendizaje automático ha revolucionado la manera en que se abordan las enfermedades cardiovasculares (ECV), especialmente en la predicción y prevención de estos padecimientos (Nagavelli *et al.*, 2023). El principal objetivo de los modelos predictivos basados en aprendizaje automático es identificar de manera temprana a individuos con alto riesgo de desarrollar enfermedades cardíacas, permitiendo que se implementen intervenciones preventivas a tiempo. La capacidad de estos modelos para analizar grandes cantidades de datos y encontrar patrones complejos ha demostrado ser valiosa en la medicina personalizada y en la toma de decisiones clínicas. Los algoritmos de aprendizaje automático pueden utilizar datos demográficos, clínicos y de estilo de vida, como la presión arterial, el colesterol, la edad y hábitos de vida, para predecir la probabilidad de sufrir un evento cardiovascular.

En el ámbito de las enfermedades cardiovasculares, los modelos de predicción más utilizados incluyen regresiones logísticas, máquinas de soporte vectorial, redes neuronales y redes bayesianas. Las redes bayesianas en particular son una herramienta poderosa, ya que no solo permiten hacer predicciones, sino que también pueden modelar las relaciones causales entre las variables de entrada, proporcionando interpretaciones más transparentes y explicativas (Moscatelli *et al.*, 2018). Esta capacidad resulta esencial en el contexto de las enfermedades cardíacas, donde la combinación de múltiples factores de riesgo necesita ser comprendida de manera integral.

1.1.2. Algoritmo NAÏVE BAYES

El algoritmo Naïve Bayes es uno de los modelos probabilísticos más simples y efectivos para la clasificación binaria, lo que lo hace adecuado para la predicción de enfermedades cardíacas (Kumar y Sahoo, 2015). Este algoritmo se basa en el teorema de Bayes y supone que las características utilizadas para la predicción son condicionalmente independientes entre sí. A pesar de esta suposición, que en muchas ocasiones no se cumple en problemas del mundo real, el modelo ha demostrado ser robusto en una amplia variedad de tareas de clasificación, incluidas las relacionadas con la medicina.

En el contexto de las ECV, el Naïve Bayes puede integrar factores como la edad, el nivel de colesterol y la presión arterial para estimar la probabilidad de que un individuo desarrolle una enfermedad cardíaca. Su simplicidad y capacidad para manejar grandes cantidades de datos lo convierten en una herramienta útil para tareas de predicción rápida, aunque su desempeño puede limitarse en escenarios donde las variables no son completamente independientes.

1.1.3. Algoritmo TAN_CL (TREE AUGMENTED NAÏVE BAYES)

El Tree Augmented Naïve Bayes (TAN_cl) es una extensión del modelo Naïve Bayes que relaja la suposición de independencia condicional entre las características (Madden, 2009). En lugar de tratar cada variable de forma independiente, el TAN_cl permite que algunas variables dependan entre sí, construyendo una estructura de árbol para modelar estas relaciones. Esto mejora significativamente el rendimiento predictivo en comparación con Naïve Bayes puro, ya que puede capturar interacciones entre las variables que son relevantes para el riesgo cardiovascular.

El TAN_cl ha sido utilizado exitosamente en la predicción de enfermedades cardiovasculares, donde las interacciones entre factores como la hipertensión, el colesterol y el hábito de fumar pueden tener efectos conjuntos en la probabilidad de desarrollar ECV (Wang *et al.*, 2014). Este modelo ha demostrado un rendimiento superior en la predicción de riesgo, sobre todo cuando se trabaja con datos heterogéneos que involucran múltiples variables interdependientes.

1.1.4. Algoritmo FSSJ (FORWARD SEQUENTIAL SELECTION AND JOINING)

El algoritmo FSSJ (Forward Sequential Selection and Joining) es un enfoque constructivo utilizado para generar redes bayesianas más complejas de manera incremental. Este método comienza con un modelo vacío y agrega variables de manera secuencial, eligiendo en cada paso la variable que maximiza el aumento en la calidad del modelo de acuerdo con una métrica predeterminada, como la verosimilitud.

Para la predicción de enfermedades cardiovasculares, el FSSJ permite la construcción de modelos más refinados y adaptativos que Naïve Bayes, ya que puede identificar relaciones complejas entre múltiples factores de riesgo, mejorando así la capacidad predictiva. Esta capacidad es especialmente valiosa cuando se analizan grandes conjuntos de datos que incluyen una amplia gama de factores demográficos y clínicos, como edad, género, presión arterial y niveles de colesterol.

1.1.6. Algoritmo TAN_HCSP (*Tree augmented naïve bayes with hill-climbing search procedure*)

El TAN_hcsp es otra variación del TAN, pero esta vez utilizando un enfoque de búsqueda por colinas (*hill-climbing*) para optimizar la estructura del árbol. A diferencia de TAN_cl, que utiliza un procedimiento basado en puntajes para construir el árbol, el TAN_hcsp emplea una técnica de búsqueda local que evalúa los posibles ajustes a la estructura y selecciona las modificaciones que mejoran el modelo (Tovey, 1985).

Este algoritmo es particularmente útil para la predicción de enfermedades cardiovasculares porque puede adaptarse a estructuras más complejas, incorporando dependencias no triviales entre los factores de riesgo. Con este enfoque, se pueden modelar con mayor precisión las interacciones entre, por ejemplo, la hipertensión y la diabetes, proporcionando una predicción más precisa del riesgo de ECV.

1.1.7. Algoritmo KDB (*K-dependence bayesian classifier*)

El K-Dependence Bayesian Classifier (KDB) es un modelo que permite que cada variable en el conjunto de datos tenga hasta K dependencias con otras variables, lo que lo convierte en una extensión más flexible del Naïve Bayes tradicional. Este enfoque equilibra la simplicidad de Naïve Bayes con la capacidad de modelar relaciones más complejas entre variables, lo que lo hace especialmente adecuado para dominios donde hay interdependencias importantes (Jahirul, 2010).

En la predicción de enfermedades cardiovasculares, el KDB puede captar relaciones entre factores como la edad, el índice de masa corporal, el tabaquismo y la hipertensión, que son determinantes clave en la aparición de eventos cardíacos. Al permitir más dependencias, este algoritmo es capaz de mejorar la precisión sin perder la simplicidad relativa de las redes bayesianas.

1.1.8. Algoritmo BSEJ (*Backward sequential elimination and joining*)

El BSEJ (Backward Sequential Elimination and Joining) es un enfoque inverso al FSSJ, donde el modelo inicial es una red bayesiana completa y se eliminan las dependencias menos importantes de manera secuencial. Este método es útil para simplificar modelos complejos, manteniendo solo las relaciones más significativas entre las variables.

En el ámbito de las enfermedades cardiovasculares, el BSEJ ha sido utilizado para reducir la complejidad de los modelos predictivos sin sacrificar la precisión. Esto es particularmente útil en escenarios con muchos factores de riesgo, donde es importante identificar las relaciones más críticas para la toma de decisiones clínicas.

2. Metodología

2.1. Datos Clínicos

Los datos mediante los cuales se llevó a cabo la investigación se encuentran publicados en la plataforma Kaggle (*Heart Failure Prediction Dataset*, n.d.). Este dataset es una combinación de cinco conjuntos de datos previamente disponibles pero no unificados, lo que lo convierte en uno de los más extensos para el estudio de enfermedades cardíacas:

- Cleveland: 303 observaciones.
- Hungarian: 294 observaciones.
- Switzerland: 123 observaciones.
- Long Beach, VA: 200 observaciones.
- Statlog (Heart): 270 observaciones.

Total Inicial: 1190 observaciones.

Duplicados Eliminados: 272 observaciones.

Dataset Final: 918 observaciones únicas.

Este conjunto de datos contiene 11 características clave que son fundamentales para predecir la presencia de una posible enfermedad cardíaca. La utilización de modelos de aprendizaje automático, y específicamente de redes bayesianas, puede ser de gran ayuda para:

- Detección Temprana: identificar a individuos con alto riesgo de desarrollar ECV.
- Toma de Decisiones Clínicas: apoyar a los profesionales de la salud en la elección de tratamientos preventivos o terapéuticos.
- Investigación: analizar la influencia de diferentes factores de riesgo en la aparición de ECV.

Descripción de los Atributos:

1. Age: edad del paciente en años.
2. Sex: género del paciente.
 - M: Masculino.
 - F: Femenino.
3. ChestPainType (Tipo de Dolor en el Pecho):
 - TA: Angina típica.
 - ATA: Angina atípica.
 - NAP: Dolor no anginoso.
 - ASY: Asintomático.
4. RestingBP: presión Arterial en Reposo: Medida en mm Hg.
5. Cholesterol: colesterol sérico en mg/dl.
6. FastingBS (Azúcar en Sangre en Ayunas):

- 1: Si el nivel de azúcar en ayunas > 120 mg/dl.
 - 0: En caso contrario.
7. RestingECG (ECG en Reposo):
- Normal: Sin anomalías.
 - ST: Anomalía en la onda ST-T.
 - LHV: Hipertrofia ventricular izquierda probable o definitiva.
8. MaxHR: frecuencia cardíaca máxima alcanzada (entre 60 y 202).
9. ExerciseAngina (Angina Inducida por el Ejercicio):
- Y: Sí.
 - N: No.
10. Oldpeak: Depresión del ST inducida por el ejercicio en relación con el reposo.
11. ST_Slope (Pendiente del Segmento ST):
- Up: Pendiente ascendente.
 - Flat: Plano.
 - Down: Pendiente descendente.
12. HeartDisease (Enfermedad Cardíaca: Variable objetivo):
- 1: Presencia de enfermedad cardíaca.
 - 0: Ausencia de enfermedad cardíaca.

La característica a clasificar es Enfermedad Cardíaca (HeartDisease) la cual tiene dos valores 1: presencia de la enfermedad cardíaca y valores 0: ausencia de la enfermedad.

La distribución de valores positivos-negativos se obtuvo sobre los registros del conjunto de datos médicos, como se muestra en la Tabla 1. Podemos determinar que el conjunto de datos esta balanceado con una diferencia de 10% aproximadamente.

Tabla 1.

Distribución de valores positivos-negativos

Casos	Número de Registros
1: Positivo	508
2: Positivo	410
Total	918

Fuente: Elaboración propia (2024).

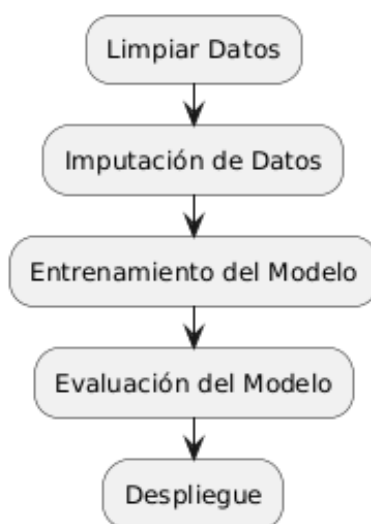
Una métrica muy popular que se utiliza para medir el rendimiento de los modelos de clasificación es la precisión.(Chicco & Jurman, 2020) Mide la proporción de muestras correctamente clasificadas respecto del total de muestras. Sin embargo, la precisión pierde su fiabilidad cuando el conjunto de datos está desequilibrado (es decir, hay significativamente más muestras en una clase que en las otras), ya que esto conduce a una evaluación demasiado optimista de la capacidad del clasificador en la clase mayoritaria. En nuestras datos *et al* estar balanceado no tenemos esta amenaza al usar esta métrica.

2.2. Diseño Metodológico

La Figura 1 muestra los pasos a considerar en el diseño metodológico, que comienza con el procesamiento de los datos y finaliza con el despliegue de la Red Bayesiana que predice enfermedades cardiovasculares.

Figura 1.

Diseño Metodológico



Fuente: Elaboración propia (2024).

A continuación, se describe en detalle cada fase de este diseño.

2.2.1. Limpieza de los datos

En esta etapa se verifica la presencia de columnas duplicadas en el conjunto de datos. Las columnas duplicadas pueden ser eliminadas. Adicionalmente, se eliminan aquellas columnas que contienen un único valor. Estas acciones se han llevado a cabo siguiendo las recomendaciones mencionadas en (Bravo *et al.*, 2019; 2020).

2.2.2. Imputación de los datos

En esta etapa se gestionan los valores ausentes en el conjunto de datos. Estos valores pueden ser sustituidos utilizando alguna técnica de imputación. La mayoría de estas técnicas se aplican solo a un tipo de variable, ya sea categórica o continua (Zhong *et al.*, 2016).

Generalmente, los diferentes tipos de variables se procesan de manera separada cuando se trabaja con datos mixtos, lo que implica que estas técnicas no consideran las posibles relaciones entre los diversos tipos de variables.

En nuestro conjunto de datos luego de explorarlo se determina que no contiene valores perdidos, por lo tanto, esta óptimo para la siguiente etapa de la metodología.

2.2.3. Entrenamiento del Modelo

Los algoritmos seleccionados para construir las redes bayesianas fueron 'Naive Bayes', 'TANcl', 'TANHcsp', 'FSSJ', 'BSEJ' y 'KDB', debido a su buen rendimiento demostrado en estudios previos. En cuanto a la división de los datos para entrenamiento y prueba, se utilizó un porcentaje de 75% para entrenamiento y 25% para prueba, siguiendo recomendaciones encontradas en la literatura científica.

2.2.4. Evaluación del Modelo

Después de entrenar los modelos de clasificación para predecir enfermedades cardiovasculares utilizando los algoritmos previamente seleccionados, se procede a comparar el desempeño de cada uno. Para esta comparación, se utilizarán métricas clave como la precisión, la especificidad y el F1 Score, las cuales son comúnmente adoptadas para evaluar el rendimiento en tareas de clasificación binaria.

La sensibilidad representa la proporción de casos positivos (personas con enfermedad cardiovascular) correctamente clasificados como positivos, mientras que la especificidad mide la proporción de casos negativos (personas sin enfermedad cardiovascular) que son clasificados correctamente como negativos.

Para obtener una estimación precisa del rendimiento, se empleará una validación cruzada de 10 veces utilizando el conjunto de entrenamiento. Además, para comparar el rendimiento de los modelos en datos que no han sido vistos previamente, se utilizará el conjunto de prueba, y se seleccionará el modelo que presente la mayor precisión en su clasificación.

2.2.5. Despliegue del modelo

La Red Bayesiana obtenida en el paso anterior fue exportada a un archivo en formato DSC para ser abierta mediante la herramienta "Bayesfusion" versión 4.1.4109.0 (también conocida como "GeNIe Academic"). El objetivo de utilizar esta herramienta es facilitar la visualización y el manejo de la red por parte de los médicos encargados de la detección y tratamiento de enfermedades cardiovasculares. Bayesfusion fue seleccionada porque es una herramienta eficaz para realizar inferencias en redes bayesianas y otros modelos probabilísticos gráficos. Además, los modelos creados con esta herramienta pueden ser fácilmente compartidos y utilizados en dispositivos móviles o a través de navegadores web. Finalmente, el proceso para establecer las evidencias en la red bayesiana para predecir una posible enfermedad cardiovascular es sencillo: basta con hacer clic en el valor de la categoría correspondiente a cada nodo.

El presente estudio fue implementado utilizando el lenguaje de programación R, elegido por su potencia y flexibilidad para el análisis estadístico y la manipulación de datos. Para llevar a cabo el preprocesamiento y la exploración de los datos, se emplearon varias librerías de R que son fundamentales en el análisis de datos. Entre ellas, *dplyr*, que se utilizó para la manipulación eficiente de los datos, permitiendo realizar operaciones complejas de filtrado,

selección y transformación de variables. Además, se utilizó la librería *arules* para la discretización de variables continuas en intervalos categóricos, dividiendo valores como la edad, el colesterol y la presión arterial en niveles como "bajo", "medio" y "alto". Esta discretización facilita el uso de modelos bayesianos que requieren variables discretas, y permite una interpretación más intuitiva de los resultados.

Para el modelado bayesiano, se emplearon las librerías *bnlearn* y *gRain*, ampliamente reconocidas en el campo de las redes bayesianas. *bnlearn* se utilizó para el aprendizaje y la estructura de la red bayesiana a partir de los datos, proporcionando herramientas para la estimación de la red y la evaluación de sus relaciones condicionales. Por otro lado, *gRain* permitió realizar inferencias probabilísticas sobre la red construida, facilitando la evaluación de probabilidades posteriores y la predicción de estados. Estas librerías combinadas proporcionan una plataforma robusta para la construcción y análisis de modelos bayesianos, permitiendo explorar las relaciones entre variables y realizar inferencias a partir de los datos disponibles.

3. Resultados

Como se menciona en la Sección 2.2.3 los algoritmos seleccionados para construir las redes bayesianas fueron 'Naive Bayes', 'TANcl', 'TANhccsp', 'FSSJ', 'BSEJ' y 'KDB'. Además, hemos considerado la variable "HeartDisease" como la variable a predecir. La división entrenamiento-prueba se usó la relación 75:25.

A continuación, se muestran todas las redes construidas, las mismas una vez generadas, se graficaron haciendo uso de la herramienta en línea *graphviz-visual-editor* disponible en <https://magjac.com/graphviz-visual-editor/> para una mejor visualización.

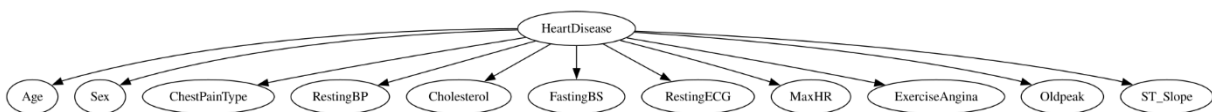
3.1. Modelado Red NAÏVE BAYES

En este modelo, Figura 2, se asume que todas las variables predictoras (edad, sexo, tipo de dolor en el pecho, colesterol, entre otras) son condicionalmente independientes entre sí dado el valor de la variable "HeartDisease".

En esta red, los factores de riesgo como la edad (Age), sexo (Sex), tipo de dolor en el pecho (ChestPainType), y otros atributos clínicos como colesterol (Cholesterol) y frecuencia cardíaca máxima (MaxHR), están conectados directamente a "HeartDisease", lo que significa que cada uno de estos factores afecta de manera directa la probabilidad de que una persona tenga o no una enfermedad cardíaca. Sin embargo, entre ellos no hay dependencias directas, lo que refleja la suposición de independencia condicional característica del Naïve Bayes.

Figura 2.

Modelado Red Naive Bayes



Fuente: Elaboración propia (2024).

3.2. Modelado Red TANcl

A diferencia del modelo Naïve Bayes estándar, TAN_cl permite dependencias adicionales entre las variables predictoras, formando una estructura más compleja que refleja las relaciones causales o dependencias condicionales entre los diferentes factores de riesgo.

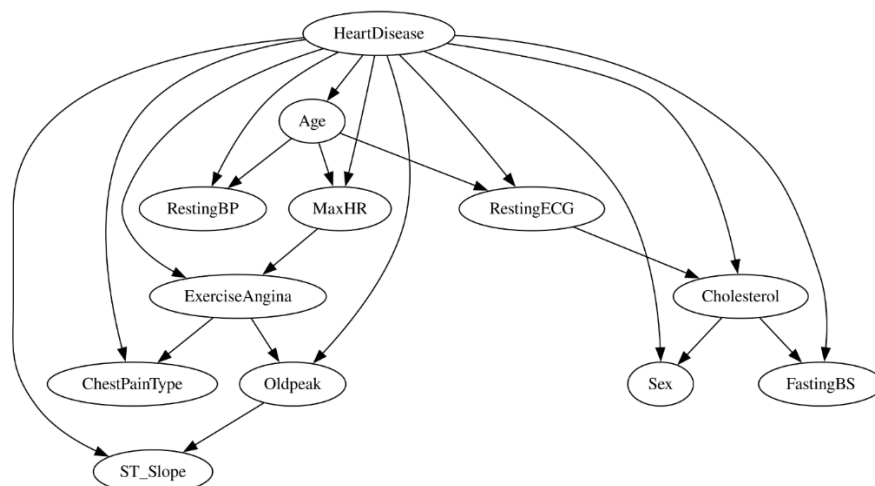
En esta red, la variable objetivo es HeartDisease (Enfermedad Cardiovascular), y se observan varias relaciones directas entre los factores de riesgo y la variable objetivo, como en los modelos anteriores. Sin embargo, a diferencia del enfoque Naïve Bayes o incluso TAN_hcsp, aquí se modelan relaciones adicionales entre los predictores mismos. Por ejemplo:

- Age influye en múltiples variables como RestingBP (Presión Arterial en Reposo), MaxHR (Frecuencia Cardíaca Máxima) y RestingECG (Electrocardiograma en Reposo), lo que refleja que la edad afecta directamente varios factores de riesgo clave.
- ChestPainType (Tipo de Dolor en el Pecho) tiene relaciones con otras variables como ExerciseAngina (Angina Inducida por Ejercicio) y ST_Slope (Pendiente del Segmento ST), lo que indica una interacción más rica entre estos síntomas y los resultados del paciente.
- Además, hay conexiones entre factores como Oldpeak, ST_Slope, y ExerciseAngina, mostrando que estas variables no son independientes entre sí.

Este tipo de estructura de árbol mostrado en la Figura 3 permite capturar interacciones entre los factores de riesgo de una manera más precisa, lo que resulta en un modelo predictivo más robusto y capaz de hacer inferencias más detalladas sobre la probabilidad de que una persona desarrolle una enfermedad cardiovascular.

Figura 3.

Modelado Red TANcl



Fuente: Elaboración propia (2024).

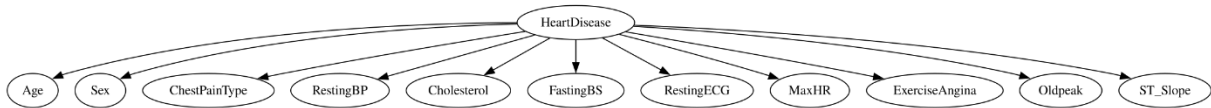
3.3. Modelado Red TANhcsp

Este modelo extiende el Naïve Bayes al permitir dependencias adicionales entre las variables predictoras, formando una estructura de árbol.

En la Figura 4, la variable objetivo sigue siendo HeartDisease (Enfermedad Cardiovascular), y al igual que en Naïve Bayes, los factores de riesgo (como la edad, el tipo de dolor en el pecho, el colesterol, y otros) están directamente conectados con HeartDisease. Sin embargo, en TANhcsp también se modelan relaciones adicionales entre las variables predictoras, que se representan mediante las conexiones adicionales entre algunos de los nodos (por ejemplo, entre ST_Slope y Oldpeak).

Figura 4.

Modelado Red TANhcsp



Fuente: Elaboración propia (2024).

El uso del enfoque de búsqueda por colinas (*hill-climbing*) optimiza la estructura del árbol al identificar las dependencias más relevantes entre las variables predictoras. Esto mejora la capacidad del modelo para capturar interacciones importantes entre los factores de riesgo, como la relación entre ciertos parámetros clínicos (como la pendiente del segmento ST y la depresión del ST), que pueden influir conjuntamente en la probabilidad de desarrollar una enfermedad cardiovascular.

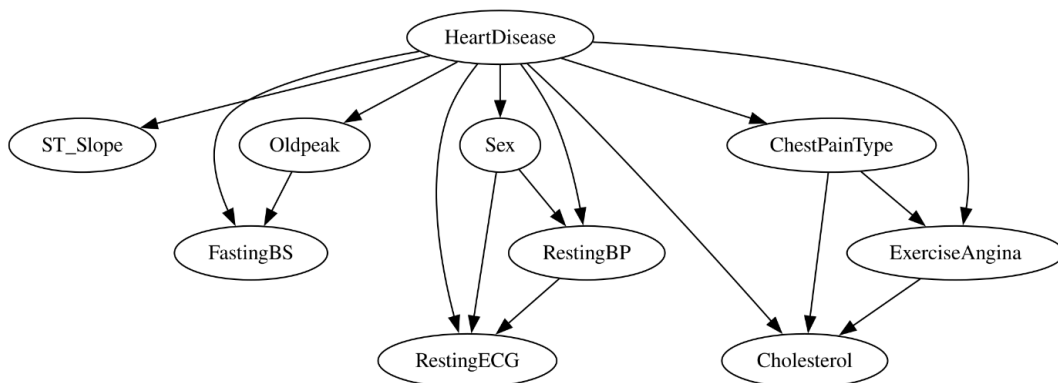
Este enfoque es más flexible y puede proporcionar predicciones más precisas que el modelo Naïve Bayes, ya que incorpora dependencias adicionales entre las variables, ofreciendo un mejor entendimiento de las interacciones entre los factores de riesgo.

3.4. Modelado Red FSSJ

La Figura 5 muestra una red bayesiana modelada utilizando el algoritmo FSSJ (*Forward Sequential Selection and Joining*). Este algoritmo construye la estructura de la red de manera secuencial, agregando variables de manera progresiva en función de su importancia para mejorar el modelo, lo que permite capturar dependencias significativas entre las variables.

Figura 5.

Modelado Red FSSJ



Fuente: Elaboración propia (2024).

En esta red, la variable objetivo sigue siendo HeartDisease (Enfermedad Cardiovascular), y está conectada con varios factores de riesgo como el Sex (Sexo), RestingBP (Presión Arterial en Reposo), FastingBS (Nivel de Azúcar en Sangre en Ayunas), ST_Slope (Pendiente del ST), entre otros.

Algunas observaciones claves de esta estructura:

- ST_Slope influye indirectamente en HeartDisease a través de Oldpeak y FastingBS, lo que muestra cómo varios factores pueden estar relacionados entre sí antes de afectar directamente a la variable objetivo.
- Existen dependencias adicionales entre variables como RestingBP y RestingECG, así como entre ChestPainType (Tipo de Dolor en el Pecho), ExerciseAngina (Angina inducida por ejercicio), y Cholesterol (Colesterol). Esto refleja que estos factores de riesgo no son independientes, sino que interactúan entre sí, afectando en conjunto la probabilidad de desarrollar una enfermedad cardiovascular.

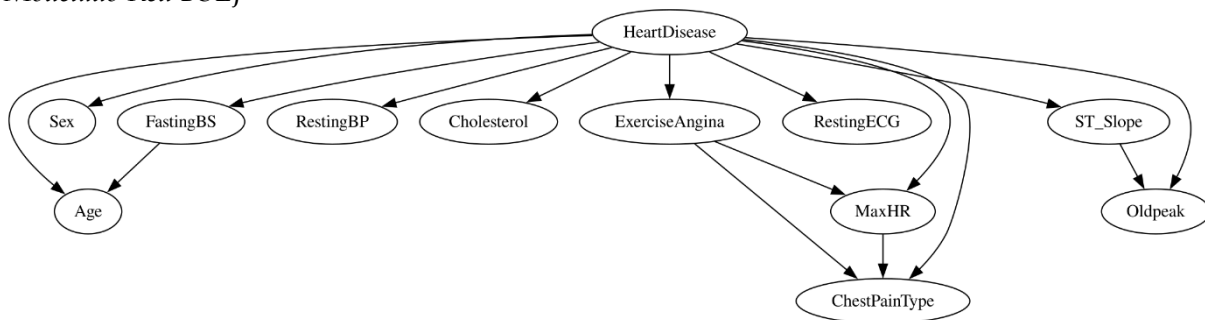
Este enfoque permite construir una red más adaptativa y eficiente, ya que las relaciones se van añadiendo progresivamente en función de su impacto en el modelo, mejorando tanto la precisión como la interpretación de las relaciones causales en el contexto de las enfermedades cardiovasculares.

3.5. Modelado Red BSEJ

La Figura 6 muestra una red bayesiana modelada utilizando el algoritmo BSEJ (Backward Sequential Elimination and Joining). Este enfoque comienza con una red completamente conectada y luego elimina progresivamente las dependencias menos significativas, dejando solo las relaciones más importantes entre las variables.

Figura 6.

Modelado Red BSEJ



Fuente: Elaboración propia (2024).

En esta red, la variable objetivo sigue siendo HeartDisease (Enfermedad Cardiovascular), y está conectada directamente con varios factores de riesgo clave, como ExerciseAngina (Angina inducida por ejercicio), RestingECG (Electrocardiograma en reposo), Cholesterol (Colesterol), entre otros. Algunas características importantes del modelo:

- Age influye en Sex, y ambos están relacionados con FastingBS (Azúcar en sangre en ayunas), RestingBP (Presión arterial en reposo), y Cholesterol, lo que refleja que estos factores están vinculados al riesgo de enfermedad cardiovascular en función del envejecimiento y el sexo del paciente.
- MaxHR (Frecuencia cardíaca máxima) y ChestPainType (Tipo de dolor en el pecho) tienen conexiones con ExerciseAngina, lo que indica cómo los síntomas asociados al ejercicio físico pueden influir en la predicción de enfermedad cardíaca.

También se destacan dependencias entre Oldpeak, ST_Slope y RestingECG, que son factores importantes en la interpretación de anomalías en los exámenes de esfuerzo y electrocardiogramas, lo que puede ayudar a identificar riesgos significativos de eventos cardíacos.

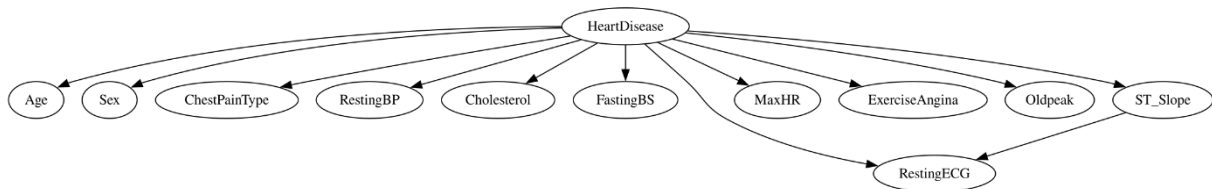
El algoritmo BSEJ optimiza la red eliminando conexiones innecesarias, simplificando el modelo sin sacrificar precisión. Esto permite que la red final sea más interpretativa y eficiente, capturando solo las relaciones más importantes entre las variables para la predicción de enfermedades cardiovasculares.

3.6. Modelado Red KDB

La Figura 7, muestra una red bayesiana modelada utilizando el algoritmo KDB (K-Dependence Bayesian Classifier). En este enfoque, cada variable predictora puede tener hasta un número máximo de dependencias con otras variables, lo que permite capturar relaciones más complejas sin sacrificar la simplicidad del modelo.

Figura 7.

Modelado Red KDB



Fuente: Elaboración propia (2024).

En esta red, la variable objetivo sigue siendo HeartDisease (Enfermedad Cardiovascular), y está conectada directamente con varias variables clave como Age (Edad), Sex (Sexo), ChestPainType (Tipo de Dolor en el Pecho), RestingBP (Presión Arterial en Reposo), Cholesterol (Colesterol), entre otras.

Observaciones importantes:

- Algunas variables, como RestingECG (Electrocardiograma en reposo), tienen conexiones adicionales con MaxHR (Frecuencia Cardíaca Máxima) y ExerciseAngina (Angina inducida por ejercicio), lo que muestra interdependencias más detalladas entre ciertos factores clínicos que afectan la predicción de enfermedades cardíacas.

- Otras conexiones, como las que involucran Oldpeak y ST_Slope, reflejan cómo las anomalías detectadas en pruebas de esfuerzo están vinculadas tanto con la variable objetivo como con otras características como ExerciseAngina.

Este modelo es más flexible que el Naïve Bayes tradicional porque permite dependencias adicionales entre los predictores, mejorando la capacidad del modelo para capturar interacciones importantes entre las variables sin volverse demasiado complejo. Esto lo convierte en una herramienta valiosa para predecir enfermedades cardiovasculares al modelar relaciones más cercanas entre los factores de riesgo sin dejar de ser interpretable.

Tabla 2.

Comparativa de los modelos

Method	Accuracy	Sensitivity	Specificity	F1
NB	0.8478261	0.8333333	0.8593750	0.8461538
TAN_cl	0.8521739	0.8333333	0.8671875	0.8499234
FSSJ	0.8565217	0.8333333	0.8750000	0.8536585
TAN_hcsp	0.8478261	0.8333333	0.8593750	0.8461538
BSEJ	0.8782609	0.8725490	0.8828125	0.8776508
Kdb	0.8608696	0.8725490	0.8515625	0.8619280

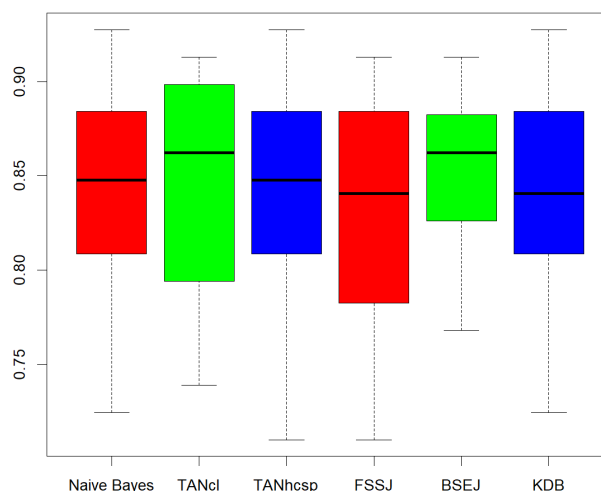
Fuente: Elaboración propia (2024).

En la Tabla 2 mostrada se comparan varios algoritmos de redes bayesianas en términos de las métricas de precisión (Accuracy), sensibilidad (Sensitivity), especificidad (Specificity), y F1 Score (F1). A partir de estas métricas, podemos analizar el rendimiento de cada modelo y seleccionar los tres mejores para el problema puntual de predicción de enfermedades cardiovasculares.

El diagrama de cajas y bigotes presentado en la Figura 8 compara el rendimiento de varios algoritmos utilizados para la predicción de enfermedades cardiovasculares, medido en términos de precisión (accuracy).

Figura 8.

Diagrama de cajas y bigotes



Fuente: Elaboración propia (2024).

A continuación, se ofrece un análisis de los tres mejores algoritmos, con base en la información visual del gráfico:

- BSEJ

Mediana: Es uno de los algoritmos con la mediana más alta, lo que indica que tiene un rendimiento consistentemente superior en la mayoría de las ejecuciones.

Rango intercuartil: Es pequeño, lo que muestra una baja variabilidad en los resultados. Esto significa que el modelo es estable y ofrece resultados confiables en la mayoría de las ejecuciones.

Bigotes: Aunque sus bigotes son más cortos que los de otros algoritmos, lo que sugiere menor variabilidad extrema, sigue siendo el algoritmo con el mejor desempeño en términos de precisión general.

- KDB

Mediana: Tiene una mediana alta, casi similar a la de BSEJ, lo que lo posiciona como uno de los algoritmos más precisos.

Rango intercuartil: Aunque el rango intercuartil es un poco mayor que en el caso de BSEJ, sigue siendo bastante ajustado, lo que sugiere que el modelo también es consistente y confiable.

Bigotes: Sus bigotes muestran una ligera variabilidad, pero su rendimiento en general es sólido, ubicándolo entre los mejores modelos.

- TAN_cl

Mediana: Tiene una mediana alta, superior a la de Naïve Bayes y TAN_hcsp, lo que lo coloca entre los mejores algoritmos.

Rango intercuartil: Presenta un rango intercuartil mayor que BSEJ y KDB, lo que implica mayor variabilidad en sus resultados. Sin embargo, su rendimiento sigue siendo fuerte en la mayoría de las ejecuciones.

Bigotes: Aunque muestra más variabilidad en los valores extremos que BSEJ, sigue ofreciendo un rendimiento consistente, lo que lo posiciona como el tercer mejor algoritmo.

Por lo tanto, podemos resumir diciendo que:

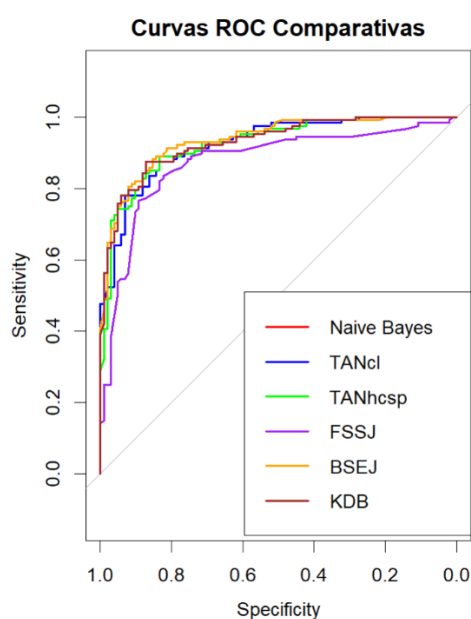
- BSEJ: Este es el mejor algoritmo, ya que tiene la mediana más alta y la menor variabilidad en su desempeño. Ofrece resultados confiables y consistentes, lo que lo convierte en la opción más sólida para este problema.
- KDB: El segundo mejor algoritmo, con una mediana alta y un rango intercuartil bajo, lo que sugiere que es un modelo preciso y estable.
- TAN_cl: Es el tercer mejor algoritmo, con una mediana elevada, aunque muestra algo más de variabilidad en comparación con los otros dos, sigue siendo una excelente opción para la predicción de enfermedades cardiovasculares.

Estos tres modelos destacan por su precisión y consistencia, lo que los convierte en las mejores opciones entre los algoritmos comparados.

El gráfico de la Figura 9, muestra las curvas ROC comparativas de diferentes algoritmos utilizados para la predicción de enfermedades cardiovasculares. La curva ROC (Receiver Operating Characteristic) es una herramienta útil para evaluar el rendimiento de los clasificadores, midiendo la relación entre la sensibilidad (verdaderos positivos) y la especificidad (falsos positivos) a través de diferentes umbrales de clasificación. El área bajo la curva (AUC) es una métrica clave que se deriva de las curvas ROC, y cuanto más cerca esté la curva del punto superior izquierdo (sensibilidad alta y especificidad alta), mejor será el rendimiento del modelo.

Figura 9.

Gráfica curvas ROC



Fuente: Elaboración propia (2024).

4. Discusión

Análisis de la comparación presentada en la Tabla 2:

BSEJ:

Accuracy: 0.8782609 (la más alta entre todos los modelos).

Sensitivity: 0.8725490 (la más alta junto con KDB).

Specificity: 0.8828125 (la más alta entre todos los modelos).

F1: 0.8776508 (también la más alta entre todos los modelos).

El algoritmo BSEJ presenta el mejor rendimiento global en todas las métricas, lo que lo convierte en el modelo más adecuado para este problema. Su capacidad de eliminar dependencias irrelevantes y simplificar la red, manteniendo las relaciones más significativas, parece haber optimizado tanto la precisión como las métricas clave de sensibilidad y especificidad.

KDB:

Accuracy: 0.8608696 (segunda más alta).

Sensitivity: 0.8725490 (la más alta junto con BSEJ).

Specificity: 0.8515625.

F1: 0.8619280 (segunda más alta).

KDB también ofrece un rendimiento sobresaliente, con una precisión muy alta y una excelente sensibilidad. Este modelo es capaz de capturar dependencias adicionales entre las variables predictoras, lo que mejora su capacidad para predecir correctamente tanto los casos positivos como negativos.

FSSJ:

Accuracy: 0.8565217 (tercera más alta).

Sensitivity: 0.8333333.

Specificity: 0.8750000 (segunda más alta).

F1: 0.8536585 (tercera más alta).

FSSJ también muestra un rendimiento sólido, con una precisión y especificidad relativamente altas. Este modelo es eficaz en la selección incremental de variables, lo que permite construir una red que optimiza las dependencias más relevantes, proporcionando un equilibrio entre las métricas clave.

Por lo tanto, podemos arribar a lo siguiente, BSEJ es el mejor algoritmo en términos de precisión, sensibilidad, especificidad y F1 Score, por lo que sería la opción más recomendada para este problema de predicción de enfermedades cardiovasculares. KDB se posiciona como el segundo mejor modelo debido a su alta precisión y sensibilidad, lo que lo convierte en una excelente alternativa cuando se busca capturar dependencias más complejas entre las variables. FSSJ es el tercer mejor modelo, destacando por su buen rendimiento en precisión y especificidad, lo que lo hace adecuado para situaciones en las que la especificidad es crucial. Estos tres modelos son los más recomendados debido a su balance entre las diferentes métricas, optimizando tanto la capacidad de detectar correctamente los casos de enfermedad cardiovascular como la minimización de falsos positivos.

Análisis de las curvas ROC:**BSEJ (Curva Naranja):**

Análisis: La curva del algoritmo BSEJ está muy cerca del eje superior izquierdo, lo que indica un rendimiento excelente en términos de sensibilidad y especificidad. Su área bajo la curva (AUC) es de las más altas, lo que confirma que el modelo es capaz de diferenciar bien entre los casos positivos y negativos.

BSEJ es el mejor algoritmo en términos de la curva ROC, con un excelente equilibrio entre sensibilidad y especificidad.

KDB (Curva Marrón):

Análisis: La curva del algoritmo KDB también está muy cerca del eje superior izquierdo, mostrando un rendimiento excelente. La AUC de este modelo es muy similar a la de BSEJ, lo que indica que también es muy eficiente en la clasificación.

KDB es el segundo mejor algoritmo, con un rendimiento muy alto en términos de la curva ROC, proporcionando un buen equilibrio entre sensibilidad y especificidad.

TAN_cl (Curva Azul):

Análisis: La curva de TAN_cl está muy cerca de las mejores, aunque se observa que es ligeramente inferior a la de BSEJ y KDB en algunas secciones. A pesar de esto, la curva ROC sigue mostrando un buen rendimiento, con una AUC alta.

TAN_cl es el tercer mejor algoritmo, ofreciendo un rendimiento competitivo, aunque ligeramente inferior a BSEJ y KDB.

Estos tres modelos destacan por su capacidad para clasificar con precisión y por mantener un buen balance entre falsos positivos y verdaderos positivos, como lo demuestra su desempeño en las curvas ROC.

La discusión ofrece una interpretación profunda de estos resultados, poniéndolos en contexto con el cuerpo existente de conocimiento. Analiza las implicaciones de los hallazgos, cómo contribuyen o desafían teorías previas, y su relevancia práctica o teórica para el campo. Además, esta sección aborda críticamente las limitaciones del estudio, discutiendo cómo éstas podrían afectar la interpretación de los resultados y sugiriendo áreas para futuras investigaciones que puedan superar estas limitaciones o explorar nuevas preguntas surgidas del estudio actual.

5. Conclusiones

Este estudio ha permitido la construcción, análisis y comparación de varios modelos de redes bayesianas, entre ellos Naive Bayes, TAN_cl, TAN_hcsp, FSSJ, BSEJ y KDB, para la predicción de enfermedades cardiovasculares. Se ha demostrado que los modelos de redes bayesianas, en particular BSEJ y KDB, son herramientas efectivas para identificar factores de riesgo y predecir la probabilidad de desarrollar una enfermedad cardiovascular, lo que es crucial para la detección temprana y la prevención de eventos cardíacos graves como infartos y accidentes cerebrovasculares.

Los resultados obtenidos indican que el modelo BSEJ fue el que mostró el mejor desempeño general, con las métricas más altas en precisión, sensibilidad, especificidad y F1 Score. Este modelo fue capaz de simplificar la red eliminando conexiones innecesarias, sin sacrificar la precisión, lo que lo convierte en una opción robusta y eficiente para la predicción de enfermedades cardiovasculares. Además, el modelo KDB también mostró un excelente rendimiento, posicionándose como una alternativa sólida para la clasificación.

El uso de técnicas de aprendizaje automático y redes bayesianas ha demostrado ser efectivo en el análisis de datos clínicos para la predicción de enfermedades. Estos modelos no solo permiten hacer predicciones con alta precisión, sino que también proporcionan una representación visual clara de las relaciones entre los factores de riesgo, lo que facilita su interpretación por parte de profesionales de la salud.

De cara al futuro, se recomienda implementar estos modelos en el software BayesFusion para permitir a los profesionales de la salud realizar inferencias en tiempo real basadas en los datos de los pacientes. Esto podría mejorar significativamente la toma de decisiones clínicas, al ofrecer a los médicos una herramienta para evaluar el riesgo cardiovascular de sus pacientes de manera personalizada y oportuna, lo que a su vez contribuiría a la prevención de enfermedades y a la reducción de la mortalidad por enfermedades cardiovasculares.

6. Referencias

- Bravo, F. P., Del Barrio García, A. A., Gago Veiga, A. B., Gallego de la Sacristana, M. M., Pinero, M. R., Peral, A. G., Dzeroski, S. y Ayala, J. L. (2019). SMURF: Systematic Methodology for Unveiling Relevant Factors in Retrospective Data on Chronic Disease Treatments. *IEEE Access*, 7, 92598–92614. <https://doi.org/10.1109/ACCESS.2019.2927429>
- Bravo, F. P., García, A. A. D. B., Russo, L. M. S. y Ayala, J. L. (2020). SOFIA: Selection of Medical Features by Induced Alterations in Numeric Labels. *Electronics*, 9(9), 1492. <https://doi.org/10.3390/electronics9091492>
- Cardiovascular diseases (CVDs). (n.d.). *Cardiovascular diseases*. <https://lc.cx/nFw6WA>
- Chicco, D. y Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Ecuador acumula pacientes con enfermedades cardiovasculares | CEAP :: Centro de Estudios Asia-Pacífico (n.d.). <https://lc.cx/x51AR1>
- Enfermedades cardiovasculares - OPS/OMS | Organización Panamericana de la Salud. (n.d.). <https://www.paho.org/es/temas/enfermedades-cardiovasculares>
- Heart Failure Prediction Dataset (n.d.). <https://lc.cx/D2AHi->
- Jahirul, M. (2010). *Investigating the Performance of Naïve- Bayes Classifiers and K- Nearest Neighbor Classifiers*. International Conference on Convergence Information Technology (ICCIT 2007), 1541-1546. <https://doi.org/10.1109/ICCIT.2007.148>
- Kumar, S. y Sahoo, G. (2015). *Classification of Heart Disease Using Naïve Bayes and Genetic Algorithm* (pp. 269-282). https://doi.org/10.1007/978-81-322-2208-8_25
- La Carga de Enfermedades Cardiovasculares - OPS/OMS | Organización Panamericana de la Salud. (n.d.). <https://www.paho.org/es/enlace/carga-enfermedades-cardiovasculares>
- Madden, M. G. (2009). On the classification performance of TAN and general Bayesian networks. *Knowledge-Based Systems*, 22(7), 489-495. <https://doi.org/10.1016/j.knosys.2008.10.006>
- Moscatelli, M., Manconi, A., Pessina, M., Fellegara, G., Rampoldi, S., Milanese, L., Casasco, A. y Gnocchi, M. (2018). An infrastructure for precision medicine through analysis of big data. *BMC Bioinformatics*, 19(S10), 351. <https://doi.org/10.1186/s12859-018-2300-5>
- Nagavelli, U., Samanta, D. y Thomas, B. (2023). *ML-Based Prediction Model for Cardiovascular Disease* (pp. 91-98). https://doi.org/10.1007/978-981-19-4052-1_11
- Tovey, C. A. (1985). Hill Climbing with Multiple Local Optima. *SIAM Journal on Algebraic Discrete Methods*, 6(3), 384-393. <https://doi.org/10.1137/0606040>
- Wang, X., Wang, F., Hu, J. y Sorrentino, R. (2014). *Exploring joint disease risk prediction*, 1180-1187.

Zhong, C., Pedrycz, W., Wang, D., Li, L. y Li, Z. (2016). Granular data imputation: A framework of Granular Computing. *Applied Soft Computing*, 46, 307-316. <https://doi.org/10.1016/j.asoc.2016.05.006>

CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN Y AGRADECIMIENTOS

Contribuciones de los/as autores/as:

Conceptualización: Rumbaut-Rangel, Dayron; **Software:** Rumbaut-Rangel, Dayron
Validación: Rumbaut-Rangel, Dayron **Análisis formal:** Maridueña Arroyave, Milton Rafael;
Curación de datos: Rumbaut-Rangel, Dayron; **Redacción-Preparación del borrador original:** Rumbaut-Rangel, Dayron **Redacción-Re- visión y Edición:** Maridueña Arroyave, Milton Rafael **Visualización:** Rumbaut-Rangel **Supervisión:** Maridueña Arroyave, Milton Rafael **Administración de proyectos:** Maridueña Arroyave, Milton Rafael **Todos los/as autores/as han leído y aceptado la versión publicada del manuscrito:** Rumbaut-Rangel, Dayron ;Maridueña Arroyave, Milton Rafael.

Financiación: Esta investigación no recibió financiamiento externo.

AUTORES:

Dayron Rumbaut Rangel

Universidad Bolivariana del Ecuador.

Doctorando en Ciencias de la Computación Aplicada, Magíster en Inteligencia Artificial y Magíster en Tecnología e Innovación Educativa. Con una sólida trayectoria profesional en la docencia e investigación en diversas instituciones educativas. Actualmente, docente e investigador en la Universidad Bolivariana del Ecuador, donde imparte clases en programas de pregrado y posgrado. Sus principales fortalezas profesionales se encuentran en las áreas de inteligencia artificial, analíticas del aprendizaje y tecnología educativa. Galardonado con la Mención de Honor Internacional METARED TIC en 2023 con el proyecto SMART EDUCATION TREE. Cuenta con competencias en gestión de plataformas virtuales y formación de formadores, así como con publicaciones en revistas académicas. Es coordinador de proyectos de investigación y director de tesis de maestría.

drumbautr@ube.edu.ec

Orcid ID: <https://orcid.org/0009-0001-9087-0979>

Google Scholar: <https://scholar.google.com/citations?user=8ZXmLmEAAAAJ&hl=es>

Milton Rafael Maridueña Arroyave

Universidad Bolivariana del Ecuador.

Ingeniero en Computación, Máster en Docencia Universitaria e Investigación Educativa, Máster en Investigación Matemática, Máster en Innovación y Transformación Digital, PhD en Ciencias Pedagógicas y, PhD en Ciencias Técnicas. Investigador Acreditado de la SENESCYT; Docente de la Facultad de Ciencias Matemáticas y Físicas de la Universidad de Guayaquil UG,

de la Facultad de Ciencias Naturales y Matemáticas de la ESPOL y del Instituto Tecnológico Universitario Bolivariano ITB. Docente de Posgrado UTEG-UPSE-UNEMI, autor de libros y artículos publicados en revistas indizadas. Ha sido Director General de Investigación Científica de la UG y Rector del Instituto Tecnológico Superior “Vicente Rocafuerte” y del Instituto Tecnológico Superior “Ana Paredes de Alfaro”. Actualmente es Par Evaluador y Consejero Académico del CACES.

mrmariduenaa@ube.edu.ec

Orcid ID: <https://orcid.org/0000-0002-8876-1896>