

# XI COLOQUIO DE MATEMÁTICAS APLICADAS

Universidad Militar Nueva Granada, Colombia - 2025

## Desarrollo de un modelo de evaluación del riesgo cardiovascular basado en identificación de factores de riesgo: un estudio transversal retrospectivo

Sarmiento Alejandro<sup>1</sup>, Bueno Yilinet<sup>2</sup>, Ramírez Margy<sup>3</sup>, Moreno Esteban<sup>4</sup>, Socarras Rafael<sup>5</sup>

### Resumen:

Las enfermedades cardiovasculares (ECV) representan una de las principales causas de mortalidad en el mundo y en Colombia. Los modelos de predicción disponibles en el país son adaptaciones de calculadoras internacionales, diseñadas principalmente para mayores de 30 años y centradas en variables clínicas tradicionales, lo que limita su aplicabilidad y accesibilidad en contextos con barreras económicas o geográficas. Este estudio transversal utilizó datos retrospectivos recolectados entre 2022 y 2025, con inclusión de variables sociodemográficas, antropométricas, de composición corporal y clínicas. Se aplicaron análisis descriptivos, análisis factorial de datos mixto (FAMD), modelo de clusterización para la caracterización de la población y se desarrolló un modelo predictivo de riesgo cardiovascular mediante machine learning (LightGBM). El desempeño del modelo se evaluó con métricas de sensibilidad, especificidad, F1 score, matriz de confusión y AUC-ROC. Se analizaron 52.473 registros (edad promedio  $39,1 \pm 13,8$  años), de los cuales el 48 % correspondía a personas inactivas físicamente, con alta prevalencia de exceso de grasa corporal y obesidad abdominal. Entre las variables más relevantes en la predicción se identificaron la circunferencia de cintura, la tensión arterial sistólica, la ocupación y la inactividad física. El modelo preliminar obtuvo un AUC-ROC del 76 %, lo que refleja un rendimiento aceptable y el potencial del uso de técnicas de machine learning para la predicción temprana del riesgo cardiovascular en poblaciones colombianas.

**Palabras Claves:** Riesgo cardiovascular, Enfermedad cardiovascular, Clusterización, Calculadora de riesgo, Modelo predictivo.

### Introducción

El Ministerio de salud y protección social de Colombia (Minsalud) afirma que 100,5 de cada 100.000 habitantes de entre 30 a 70 años, fallecieron por enfermedades cardiovasculares en el país en el 2022 (Minsalud conmemora el día mundial del Corazón, s. f.). Estas enfermedades se asocian principalmente con factores de riesgo ampliamente estudiados; sin embargo, en los últimos años se han identificado otros determinantes relevantes, que aún no han sido incorporados en las calculadoras de predicción de riesgo disponibles (Lloyd-Jones et al., 2022).

Las más utilizadas en Colombia son la Framingham score y Hearts, estas herramientas evalúan el riesgo a población mayor de 40 años, dejando de lado la población adulta joven, lo cual es relevante debido a que en los últimos años se ha evidenciado un aumento en la proporción de esta

<sup>1</sup>Bodytech S.A. , Colombia. alejandro.sarmiento@bodytechcorp.com

<sup>2</sup>Bodytech S.A. , Colombia. angelica.bueno@bodytechcorp.com

<sup>3</sup>Bodytech S.A. , Colombia. margy.ramirez@bodytechcorp.com

<sup>4</sup>Bodytech S.A. , Colombia. esteban.moreno@bodytechcorp.com

<sup>5</sup>Bodytech S.A. , Colombia. consultor.tecnologia.vp@bodytechcorp.com

población que han experimentado infarto agudos de miocardio y accidente isquémico, estas enfermedades se han relacionado con factores de riesgo que comienzan en la adolescencia temprana, como cambios en el estilo de vida, independencia financiera, nuevas responsabilidades, disminución del transporte activo y participación en deportes, aumento de comportamientos sedentarios, contexto social en el que se desarrollan incluyendo relaciones inter personales e influencia de pares y cambios en los horarios rutinarios (Scott et al., 2025). A pesar de la información existente sobre los factores de riesgo, se desconoce la identificación de nuevos factores que podrían proporcionar una precisión predictiva en la aparición de ECV.

**Objetivo General:** Diseñar un nuevo modelo de evaluación de riesgo cardiovascular basado en la identificación de factores de riesgo significativos que podrían estar relacionados con el desarrollo de enfermedad cardiovascular.

**Metodología:** Para este estudio transversal retrospectivo se utilizaron fuentes de base de datos entre 2022 y 2025. Se aplicaron criterios de inclusión y exclusión y se incluyeron variables sociodemográficas, de composición corporal, estilo de vida y clínicas. Se estima el riesgo cardiovascular con calculadora existente (HEARTS) para identificar el nivel de riesgo en el que se encontraban los usuarios. Posteriormente, se realizó Análisis Factorial de Datos Mixtos incorporando variables numéricas y categóricas; se realizó preprocesamiento de los datos, inspección exploratoria de las variables y armonización de los datos. Luego, se utilizó el algoritmo K-means y finalmente se desarrolló el modelo predictivo teniendo en cuenta la ingeniería de características correspondiente, para la selección de variables significativas aplicadas a un algoritmo de aprendizaje automático supervisado con la técnica Light Boosting Machine. El desempeño del modelo se evaluará con las métricas de precisión, sensibilidad, F1 Score y el área bajo la curva.

**Resultados preliminares:** Desde febrero de 2022 hasta mayo de 2025 se obtuvo información de 277.851 personas. De ellas, el 79 % fueron excluidos del estudio por criterios de inclusión y exclusión, así como datos faltantes, quedando una población de estudio de 52.473 personas.

La edad promedio de la población fue de  $39,1 \pm 13,8$  años, en donde el 30 % está entre los 18 y los 30 años. La mayoría de la población pertenecía al estrato 3 (44,6 %) y al estrato 4 (35,3 %). En cuanto a los factores de riesgo, el 48 % de la población eran inactivos físicamente, el 6,0 % eran fumadores, el 4,0 % presentaban hipertensión arterial y el 42,5 % reportaban antecedentes de enfermedad cardiovascular. Según las características antropométricas, el promedio de la población se encuentra en sobrepeso y en el límite de los valores normales de grasa visceral. Además, el porcentaje de grasa medido con bioimpedancia, estuvo por encima de los rangos normales. A pesar de presentar características de riesgo al evaluarlo con la calculadora Hearts de la OMS se obtuvo que solo el 15 % de la población fue calificada como riesgo alto.

Teniendo en cuenta los eigenvalores, una vez realizado el modelo de FAMD, se conservaron 18 componentes principales, los cuales explican en conjunto el 56,5 % de la varianza acumulada de los datos. Esto permitió realizar el modelo no supervisado de clusterización K-means en donde, según el índice de silueta de 0.59, se segmenta la población en 3 grupos que permitieron identificar los posibles factores de riesgo cardiovascular por clúster, no tenidos en cuenta tradicionalmente, como lo son el sedentarismo, la ocupación y el tipo de régimen de salud. Una vez caracterizada la población, se procede a realizar un modelo predictivo de machine learning (LightGBM) en donde se realiza una ingeniería de características para la selección de las variables más significativas y la obtención de un porcentaje de probabilidad de riesgo cardiovascular por individuo. Se realiza una clasificación binaria de 1 = con riesgo y 0 = sin riesgo, obteniendo un AUC-ROC del 76 % en donde el 42 % de la población fue clasificada con riesgo potencial.

## Conclusión

Las variables de composición corporal y sociodemográficas demostraron ser importantes en la evaluación de riesgo cardiovascular. El modelo de clusterización permitió identificar la importancia de variables que no son tenidas en cuenta en las calculadoras actuales, así como la descripción

de la población de estudio. Se demostró que el uso de un modelo de machine learning permite generar una predicción del riesgo, mostrando preliminarmente métricas aceptables en train y test para identificar a las personas con y sin riesgo, demostrando que puede ser usado para la identificación temprana y prevención de enfermedades cardiovasculares en poblaciones jóvenes, explicando su potencial impacto.

## Bibliografía

- D. M. Lloyd-Jones, N. B. Allen, C. A. Anderson, T. Black, L. C. Brewer, R. E. Foraker, M. A. Grandner, H. Lavretsky, A. M. Perak, G. Sharma, W. Rosamond, & en nombre de la American Heart Association, *Life's Essential 8: Updating and Enhancing the American Heart Association's Construct of Cardiovascular Health: A Presidential Advisory From the American Heart Association*, *Circulation*, **146**(5), e18–e43, 2022. <https://doi.org/10.1161/CIR.0000000000001078>
- Minsalud, *Minsalud conmemora el día mundial del Corazón*, (s.f.). Recuperado el 7 de febrero de 2025, de <https://www.minsalud.gov.co/Paginas/Minsalud-conmemora-el-dia-mundial-del-Corazon.aspx>
- J. Scott, A. Agarwala, C. M. Baker-Smith, M. J. Feinstein, K. Jakubowski, J. Kaar, N. Parekh, K. V. Patel, J. Stephens, & the American Heart Association Prevention Science Committee of the Council on Epidemiology and Prevention and Council on Cardiovascular and Stroke Nursing; Council on Lifelong Congenital Heart Disease and Heart Health in the Young; and Council on Lifestyle and Cardiometabolic Health, *Cardiovascular Health in the Transition From Adolescence to Emerging Adulthood: A Scientific Statement From the American Heart Association*, *Journal of the American Heart Association*, **14**(9), e039239, 2025. <https://doi.org/10.1161/JAHA.124.039239>