



2019 10 al 13 de septiembre - Cartagena de Indias, Colombia

RETOS EN LA FORMACIÓN
DE INGENIEROS EN LA
ERA DIGITAL



DESARROLLO DE UNA APLICACIÓN MÓVIL (APPTEROMA) QUE PERMITA ESTIMAR EL RIESGO DE PADECER ATEROSCLEROSIS UTILIZANDO ALGORITMOS DE MACHINE LEARNING

**María Alejandra Barajas Padrón, Deneý Ballesteros Ruiz, Rodrigo García Hoyos,
José Fernando Pineda Vergara, Andrés Buelvas Serpa**

**Universidad del Sinú
Montería, Colombia**

Resumen

La aterosclerosis es la principal causa de la enfermedad arterial coronaria, la cual hace parte de las enfermedades cardiovasculares, siendo estas últimas la principal causa de muerte a nivel mundial. En la presente investigación se evidencia el resultado del trabajo realizado acerca de una aplicación para dispositivos móviles, la cual permitirá estimar el riesgo de padecer aterosclerosis utilizando algoritmos de Machine Learning. El funcionamiento de la aplicación consiste en que el médico suministra información básica del paciente durante una consulta y posteriormente, la aplicación dará una predicción de manera porcentual, la cual corresponde al riesgo de padecer aterosclerosis en el futuro. El alcance de este proyecto está dado por la obtención de un diagnóstico rápido y eficaz a través de la utilización de las TIC (Tecnologías de la Información y Comunicación) en el área de las Ciencias de la Salud y de disciplinas como Machine Learning. En este documento se encuentran las etapas de análisis, diseño y desarrollo de la aplicación.

Palabras clave: aterosclerosis; aplicación móvil; machine learning

Abstract

Atherosclerosis is the main cause of coronary artery disease, which is part of cardiovascular diseases, the latter being the leading cause of death worldwide. In the present investigation, the result of the work done on an application for mobile devices is demonstrated, the quality estimating the risk of suffering atherosclerosis using automatic learning algorithms. The operation of the

application is that the doctor becomes the result. The scope of this project is given by obtaining a rapid and effective diagnosis through the use of ICT (Information and Communication Technologies) in the area of Health Sciences and disciplines such as Machine Learning. In this document are the stages of analysis, design and development of the application.

Keywords: *atherosclerosis; mobile app; machine learning*

1. Introducción

Las enfermedades cardiovasculares son la primera causa de muerte en el mundo, siendo la aterosclerosis la etiología más frecuente de estas. Por ejemplo, la Organización Mundial de la Salud calcula que en 2012 murieron por esta causa 17,5 millones de personas, lo cual representa un 31% de todas las muertes registradas en el mundo. La aterosclerosis es la formación de una placa de ateroma en los vasos sanguíneos, que disminuye la irrigación de un tejido principalmente y en este caso, del corazón. La presente investigación se refiere al desarrollo de una aplicación para dispositivos móviles (Appteroma) que permita estimar el riesgo de padecer aterosclerosis en el futuro, mediante algoritmos de Machine Learning, teniendo en cuenta los factores que participan en la formación de las placas ateroscleróticas.

2. Planteamiento del problema

¿Cómo el desarrollo de una aplicación móvil permite facilitar la estimación del riesgo de padecer aterosclerosis utilizando algoritmos de Machine Learning?

Actualmente la aterosclerosis es la principal causa de la enfermedad arterial coronaria, que pertenece a las enfermedades cardiovasculares, las cuales son responsable de la mayoría de muertes en todo el mundo, Los estudios epidemiológicos muestran que para 2020 la enfermedad cardiovascular será responsable de 25 millones de muertes al año según la Revista Española de Cardiología.

En esta patología influyen distintos factores de riesgo como el cigarrillo, dieta, alcohol, obesidad, sedentarismo, dislipidemias y la edad, de esta última encontramos que la aterosclerosis es un proceso crónico, que desde la juventud hay que controlar los factores de riesgo, para así evitar en las edades avanzadas la formación de ateromas que disminuyen el flujo sanguíneo a órganos importantes y que pueden llevar a la muerte. (Llapur Milián & González Sánchez, 2017). Muchos de los factores que influyen en la fisiopatología de esta, como los mencionados anteriormente, los podemos modificar, he aquí la importancia de calcular la probabilidad de adquirir una enfermedad aterosclerótica, ya que nos permite tomar esos factores, modificarlos de manera precoz para así disminuir el riesgo y mejorar calidad de vida.

3. Justificación

La aterosclerosis es una enfermedad crónica y progresiva, que afecta a arterias de diferentes localizaciones, tiende a asentarse en las arterias que irrigan el cerebro, las extremidades inferiores y el corazón (coronarias), siendo esta últimas las más afectadas (Lahoz & Mostaza;2007). Según la Revista Española de Cardiología, las enfermedades cardiovasculares es la principal causa de muerte en el mundo, entre ellas la enfermedad arterial coronaria que se asocia a alta morbimortalidad. La placa de ateroma está constituida por lípidos, colesterol, células inflamatorias, que provocan una estenosis de la luz de un vaso sanguíneo disminuyendo así el aporte de sangre y con ello el oxígeno hacia el tejido irrigado, provocando necrosis de este (Rev. Urug cardiol; 2014).

Calcular el riesgo de padecer aterosclerosis tiene gran importancia puesto que nos permite detectar la predisposición de padecer este evento, llevando así a evitar las complicaciones a las que conlleva, y disminuir la mortalidad, todo esto mediante el ingreso de datos y factores modificables y significativos que influyen en la fisiopatología de la formación de una placa de ateroma.

4. Objetivos

4.1. Objetivo general

Desarrollar de una aplicación móvil (Appteroma) que permita estimar el riesgo de padecer aterosclerosis utilizando algoritmos de Machine Learning.

4.2. Objetivos específicos

- Identificar las variables a tener en cuenta para el diagnóstico de aterosclerosis.
- Diseñar una aplicación móvil capaz de estimar el riesgo de padecer aterosclerosis.
- Implementar el algoritmo de Machine Learning adecuado para estimar el riesgo de padecer aterosclerosis.
- Realizar pruebas a pacientes con aterosclerosis y pacientes sanos, con el fin de medir el rendimiento de la aplicación y asimismo, del algoritmo.

5. Referente teórico

Aterosclerosis: presencia de una placa de ateroma en la luz de un vaso sanguíneo.

Placa de ateroma: Conjunto de grasa y otras sustancias que se acumulan en forma de placas en las arterias.

Síndrome Coronario Agudo: Reducción aguda o progresiva del flujo coronario que puede originar necrosis miocárdica.

Vasos sanguíneos: Estructura hueca y tubular que conduce la sangre impulsada por la acción del corazón.

Machine Learning: Es una disciplina de la Inteligencia Artificial¹ que permite crear sistemas que aprenden de forma automatizada, es decir que, para resolver problemas, toma decisiones basadas en la experiencia acumulada.

Python: Es un lenguaje de programación interpretado de alto nivel.

Scikit-Learn: Biblioteca de Python que contiene Herramientas simples y eficientes para la minería de datos y el análisis de datos

Seaborn: Biblioteca de visualización de datos de Python. Proporciona una interfaz de alto nivel para dibujar gráficos estadísticos atractivos e informativos.

ASCVD Risk Estimator: Aplicación que permite estimar los riesgos de enfermedad cardiovascular aterosclerótica para los próximos 10 años y de por vida.

6. Metodología

La investigación está enmarcada en un modelo cuantitativo y orientado hacia un enfoque correlacional.

La investigación cuantitativa usa la recolección de datos para probar hipótesis, con base en la medición numérica y el análisis estadístico, para establecer patrones de comportamiento y probar teorías. (Hernández Sampieri, Fernández Collado, Baptista Lucio, García Espejo & Limón Cano, 2010). Asimismo, la utilidad y el propósito principal de los estudios correlacionales son saber cómo se puede comportar un concepto o variable conociendo el comportamiento de otras variables relacionadas. (Hernández et al., 2010, p.63)

Se pretende desarrollar una aplicación que permita estimar el riesgo de padecer aterosclerosis en pacientes utilizando algoritmos de Machine Learning. La aplicación graficará un mapa de calor, indicando de manera porcentual la probabilidad de padecer aterosclerosis en los próximos años en base a información suministrada por el médico, que será comparada con un set de datos con información de personas que padecen aterosclerosis en búsqueda de similitudes. Este proyecto se divide en 7 fases:

1. Recolección de información: Consiste en estudiar las variables que intervienen en el diagnóstico de la aterosclerosis y cuáles de estas son las más relevantes.

Se tomó como referencia la aplicación ASCVD Risk Estimator desarrollada por American College of Cardiology. Al estudiar la aplicación, se pudo notar tres puntos importantes. En primera instancia, saltó a la vista el rango de edad de los pacientes que pueden utilizar la aplicación, el cual se encuentra entre los 40 y 79 años, lo que contradice en gran medida la afirmación que aunque los síntomas atribuidos a la enfermedad aterosclerótica aparecen en la edad mediana o tardíamente, los factores de riesgo están presentes desde edades tempranas, y tienden a persistir a lo largo del tiempo hasta la etapa adulta (Llapur Milián & González Sánchez, 2017).

¹ Actualmente la Inteligencia Artificial es un área de la ciencia de gran interés por ser un área multidisciplinaria donde se realizan sistemas que tratan de hacer tareas y resolver problemas como lo hace un humano

De acuerdo a lo expuesto anteriormente, se determinó ampliar el rango de edad entre 20 y 85 años. También, se destacó que en dicha aplicación se desprecian parámetros determinantes para el diagnóstico de aterosclerosis, entre estos tenemos: peso, talla, antecedentes familiares de enfermedades cardiovasculares, entre otros. Por otro lado, en el estudio de la aplicación se pudo concluir que utilizan el algoritmo de árbol de decisión², como se observa en la siguiente figura.

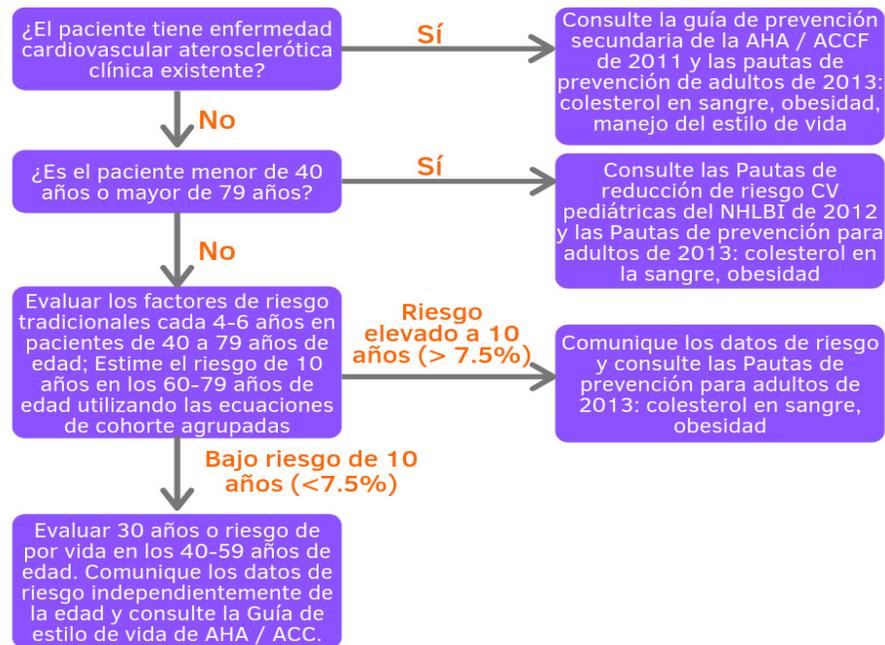


Figura 1. Arquitectura del algoritmo utilizado en ASCVD Risk Estimator. Elaboración propia

Dicho esto, se procedió a realizar correlaciones entre variables adicionando aquellas que se consideren importantes, teniendo en cuenta un set de datos con información de personas que padecen aterosclerosis, esto con el fin de comprobar estadísticamente cuáles de estas son realmente determinantes.

2. Diseño: Diseñar la aplicación teniendo en cuenta las variables faltantes, garantizando la alta comprensibilidad por el usuario final (médicos).

3. Selección del algoritmo: Probar algoritmos de Machine Learning con el fin de identificar el más apropiado para la aplicación. En esta etapa se realizaron pruebas con algoritmos de árbol de decisión, el cual utiliza ASCVD Risk Estimator, con el objetivo de examinar que tan óptimo es y la magnitud de su error, para de esta manera tener una referencia. También se probaron algoritmos de regresión lineal³, random forest⁴ y de clasificación.

² Dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

³ La regresión lineal o ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes Xi y un término aleatorio ϵ .

⁴ Es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos

4. Entrenamiento: Entrenar el algoritmo para que con la experiencia obtenida sea capaz de detectar patrones en la información suministrada por el médico que pueda significar que la persona padece o puede padecer de aterosclerosis.

5. Revisión: Verificar posibles errores teniendo en cuenta los resultados que arroja el algoritmo

6. Pruebas: Probar la aplicación con la finalidad de medir su rendimiento.

7. Implementación: Una vez esté desarrollada la aplicación, se procederá a implementarla.

7. Resultados

Con relación a los resultados obtenidos en las correlaciones entre variables tenemos que la obesidad tiene una alta correlación con el cáncer, las enfermedades cardiovasculares y respiratorias, la diabetes y el consumo de alcohol. De igual manera, estos parámetros están fuertemente correlacionados entre sí, como se observa en la siguiente figura:

	id	obesidad	cancer	cardiovascular	hipertension	diabetes	enf_respiratorias	alcohol
id	1.000000	-0.059603	-0.022216	-0.010639	-0.035143	0.004167	-0.020694	-0.039787
obesidad	-0.059603	1.000000	0.910811	0.921884	0.878661	0.936250	0.918346	0.967821
cancer	-0.022216	0.910811	1.000000	0.990866	0.981942	0.959140	0.993004	0.941462
cardiovascular	-0.010639	0.921884	0.990866	1.000000	0.969167	0.977319	0.989950	0.945324
hipertension	-0.035143	0.878661	0.981942	0.969167	1.000000	0.933451	0.982303	0.888750
diabetes	0.004167	0.936250	0.959140	0.977319	0.933451	1.000000	0.971411	0.948639
enf_respiratorias	-0.020694	0.918346	0.993004	0.989950	0.982303	0.971411	1.000000	0.935103
alcohol	-0.039787	0.967821	0.941462	0.945324	0.888750	0.948639	0.935103	1.000000

Figura 2. Correlaciones entre las variables de un set de datos con información de personas con aterosclerosis.

Considerando esto, se identificaron los puntos relevantes para el diagnóstico de la aterosclerosis. Seguidamente, se hicieron pruebas en algoritmos de Machine Learning como árbol de decisión, random forest y algoritmos de clasificación para así buscar la mejor predicción y posteriormente, seleccionar el algoritmo óptimo.

Finalmente, se diseñó la aplicación teniendo en cuenta los requerimientos identificados previamente.

DESARROLLO DE UNA APLICACIÓN MÓVIL (APPTEROMA) QUE PERMITA ESTIMAR EL RIESGO DE PADECER ATEROSCLEROSIS UTILIZANDO ALGORITMOS DE MACHINE LEARNING

Figura 3. Formulario de la aplicación Appteroma

Como trabajos futuros se espera elegir el algoritmo de Machine Learning con la respuesta más óptima y el menor porcentaje de error, implementar el algoritmo escogido en la aplicación previamente diseñada y realizar las respectivas pruebas a pacientes sanos y pacientes con aterosclerosis.

8. Referencias

Artículos de revistas

- Moderador, W. H., Bassand, J., Noruega, S. A., Bax, J., Bajos, P., Boersma, E., ... Unidos, E. (2012). *Guia2012_3*, 65(2).
- Raudales, J. C., Zago, A. C., Casco, M. F. De, Bortolini, M. A. G., & Flores, I. V. C. (2015). *Coronaria Aterosclerótica Vulnerable*, 83, 57–65.
- Lahoz, C., Mostaza, J. M., & Lahoz, C. (2007). La aterosclerosis como enfermedad sistémica PUESTA AL DÍA. *Rev Esp Cardiol*, 60(2), 184–95.
- Núñez García, M. V., Ferrer Arrocha, M., Meneau Peña, T. X., Cabalé Vilariño, B., Gómez Nario, O., & Miguelez Nodarse, R. (2007). Factores de riesgo aterogénico en la población de 19 a 39 años de 2 consultorios del médico de familia. *Revista Cubana de Investigaciones Biomedicas*, 26(2), 1–7.
- Milián, R. L., & Sánchez, R. G. (2017). La enfermedad cardiovascular aterosclerótica desde la niñez a la adultez *Atherosclerotic cardiovascular disease going from childhood to adulthood*, 89(3), 271–277.

Libros

- Moreno, A., Armengol, E., Béjar, J., Belanche, L., Cortes, U., & Gavaldá, R. et al. (2004). *Aprendizaje Automático* (p. 7). Edicions UPC.

- Hernández Sampieri, R., Fernández Collado, C., Baptista Lucio, P., García Espejo, M., & Limón Cano, S. (2010). *Fundamentos de metodología de la investigación*. Madrid: McGraw-Hill.
- Ponce Gallegos, J., Torres Soto, A., Quezada Aguilera, F., Silva Sprock, A., Martínez Flor, E., Casali, A., Scheihing, E., Túpac Valdivia, Y., Torres Soto, M., Ornelas Zapata, F., Hernández A., J., Zavala D., C., Vakhnia, N. and Pedreño, O. (2014). *Inteligencia Artificial*. 1st ed. Iniciativa Latinoamericana de Libros de Texto Abiertos.

Fuentes electrónicas

- Álvarez, M. (2019). Qué es Python. Recuperado de <https://desarrolloweb.com/articulos/1325.php>
- Scikit-Learn: machine learning in Python — scikit-learn 0.20.3 documentation. (2019). Retrieved from <https://scikit-learn.org/stable/>
- Seaborn: statistical data visualization — seaborn 0.9.0 documentation. (2019). Retrieved from <https://seaborn.pydata.org>

Sobre los autores

- **María Alejandra Barajas Padrón:** Estudiante de Ingeniería de Sistemas. alejabarajas46@gmail.com
- **Deney Ballesteros Ruiz:** Estudiante de Medicina. dalal1106.sbr@gmail.com
- **Rodrigo García Hoyos:** Docente de Ingeniería de Sistemas. rodrigogarciahoyos@gmail.com
- **José Fernando Pineda Vergara:** Docente de Ingeniería industrial. josepineda@unisnu.edu.co
- **Andrés Buelvas Serpa** Estudiante Ingeniería Ambiental

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2019 Asociación Colombiana de Facultades de Ingeniería (ACOFI)