

PREDICIÓN COLESTEROL EN LAS PERSONAS CON MODELOS SUPERVISADOS

Introducción	Data Set	Preparación de los Datos
--------------	----------	--------------------------

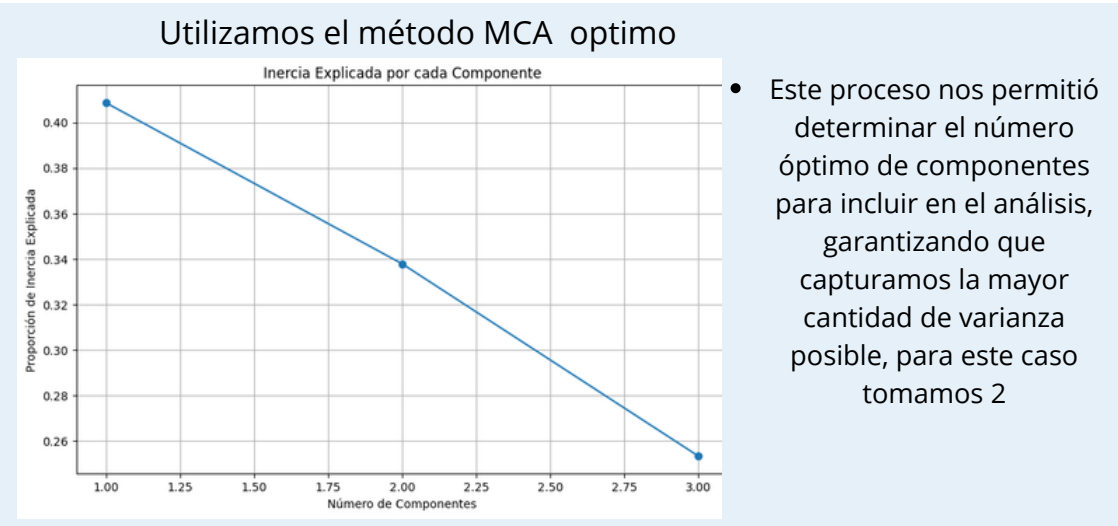
Este informe analiza un modelo de clasificación para predecir los niveles de colesterol utilizando un conjunto de datos que incluye diversas características. El objetivo principal es identificar las variables más importantes que influyen en los niveles de colesterol y evaluar el rendimiento del modelo de clasificación.

El dataset utilizado en este estudio incluye varias características relevantes para la predicción del colesterol alto, tales como:

- Edad: La edad de la persona.
- Sexo: El género de la persona .
- PS: Presión sistólica
- Na: Nivel de sodio en la sangre.
- K: Nivel de potasio en la sangre.
- Droga: medicamentos específicos
- Colesterol: Nivel de colesterol en la sangre

Carga de Datos y Preprocesamiento:

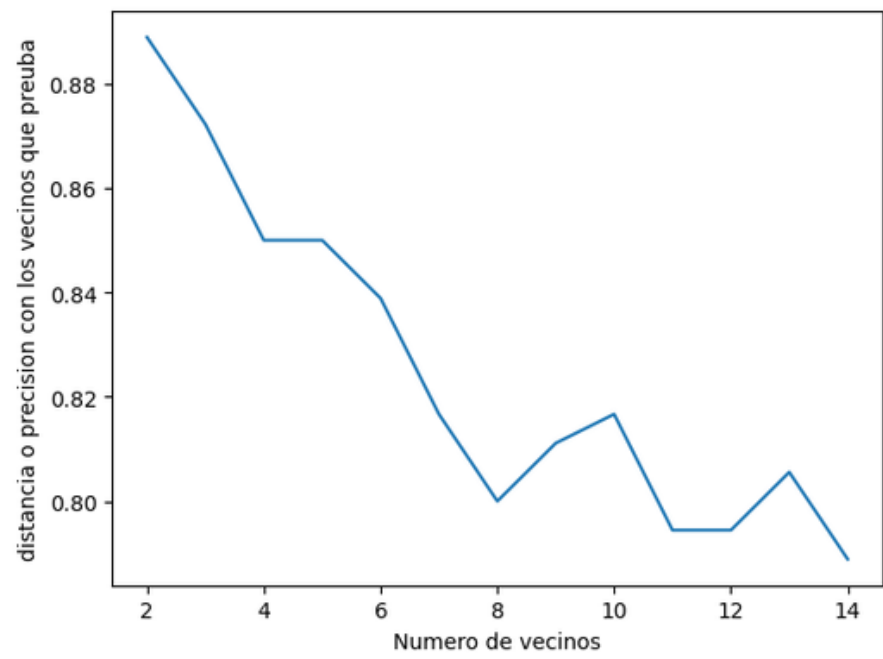
- Los datos se cargaron y limpiaron para eliminar valores faltantes y asegurar que las variables estuvieran en el formato correcto.
- Se realizó una exploración inicial para entender la distribución de los datos y detectar posibles valores atípicos.
- Las variables de tipo objeto fueron identificadas y convertidas en variables categóricas.
- Esta conversión permite la codificación MCA (Análisis de Correspondencias Múltiples)), lo que ayuda a obtener modelos de predicción más precisos y optimizados.



Modelos de Predicción

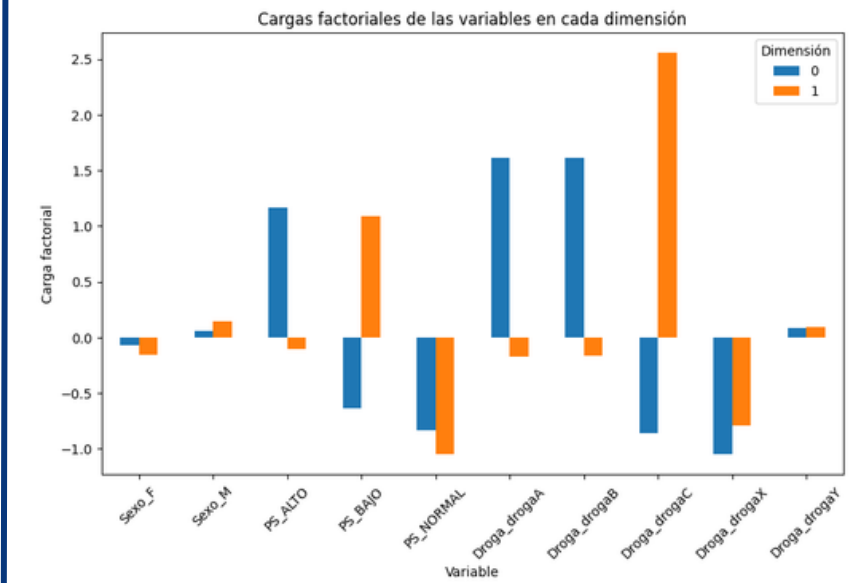
1. Selección de Modelos:
- Se utilizaron los siguientes algoritmos de predicción:
 - K-Nearest Neighbors (KNN)
 - Regresión Logística
 - Naive Bayes
 - Redes Neuronales
 - Random Forest

Modelo KNN(%precisión)

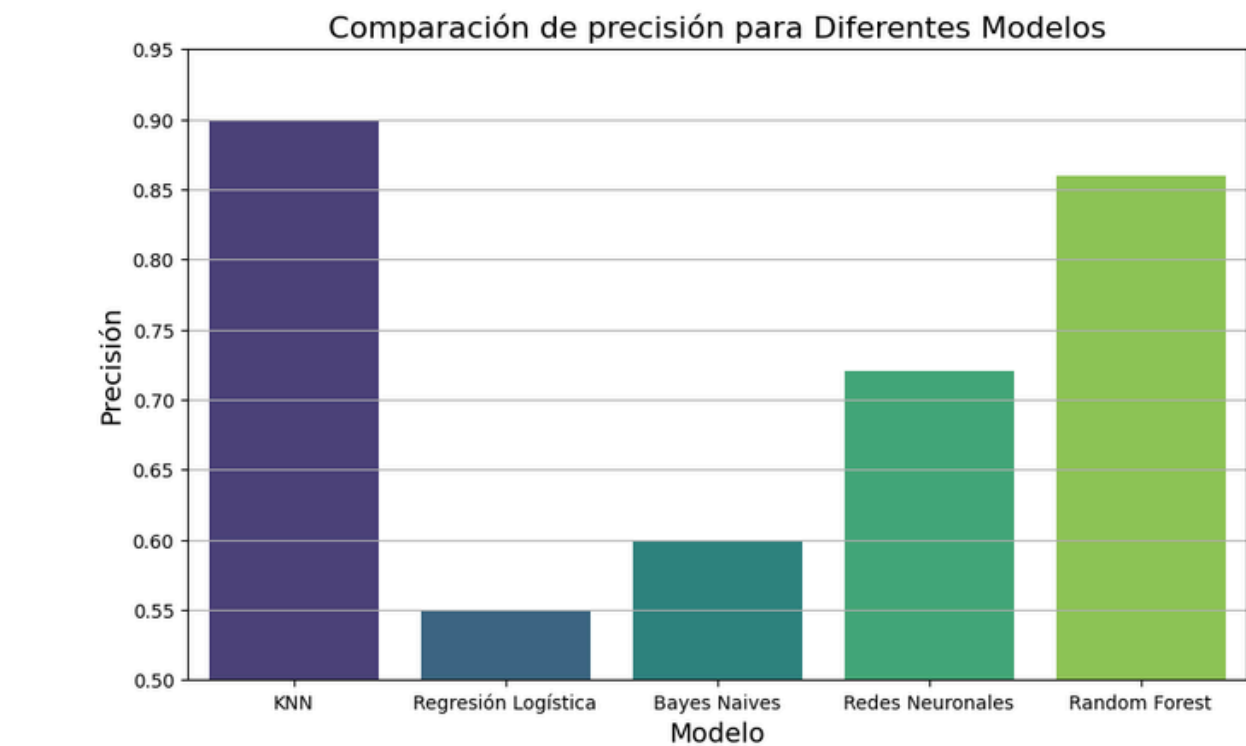


Cargas Factoriales de las Variables

Para entender la influencia de cada variable en las dimensiones obtenidas del AMC, se calcularon y graficaron las cargas factoriales de las variables en cada dimensión:



Resultados



Conclusión

El modelo K-Nearest Neighbors (KNN) fue el mejor modelo con una precisión del 90%. Este modelo optimizado proporciona una herramienta valiosa para la predicción del colesterol alto, permitiendo intervenciones tempranas y mejorando la salud cardiovascular de las personas.

Referencias

Estudios previos y fuentes de datos utilizada