

Entrega Final - Factores de Riesgo en Enfermedades Cardíacas

Grupo 8. Diana Bayona, Juan Diego Pérez, Jorge Rodríguez y Santiago Gutiérrez
GitHub: https://github.com/JuanD13Perez/Proyecto_Final_Despliegue_Gr8

1. Problema

Las enfermedades cardiovasculares son la principal causa de muerte a nivel mundial, se estima que al año mueren alrededor de 17.9 millones de personas (OMS). La mayoría de estas muertes ocurren en países de ingresos medianos y bajos. Estas enfermedades son trastornos del corazón y vasos sanguíneos, de esta manera se relaciona con el hecho de que el corazón no bombea la suficiente sangre para el cuerpo y esto puede provocar ataques cardíacos, enfermedades cerebrovasculares, insuficiencia cardíaca, entre otras.

De acuerdo con la Organización Mundial de la Salud, las enfermedades cardiovasculares están relacionadas con el hecho de que una persona padezca de hipertensión arterial, diabetes, colesterol alto, obesidad, o incluso por la falta de actividad física y el consumo excesivo de alcohol y tabaco. Se estima que la exposición a productos derivados del tabaco es responsable del 10% de todas las muertes ocasionadas por enfermedades cardiovasculares (Organización Panamericana de Salud).

Dado el significativo rol del corazón como un órgano vital, anticipar la insuficiencia cardíaca se ha vuelto una prioridad para profesionales médicos; sin embargo, hasta el momento, predecir eventos asociados con esta condición en la práctica clínica suele no alcanzar niveles de precisión satisfactorios dado que sólo se logra una precisión modesta con una interpretabilidad limitada de las variables predictivas.

Teniendo en cuenta lo anterior, se analiza un conjunto de datos que contiene los registros médicos de 299 pacientes con insuficiencia cardíaca recopilados en el Instituto de Cardiología de Faisalabad y en el Hospital Allied de Faisalabad (Punjab, Pakistán), entre abril y diciembre de 2015 (Ahmad et al., 2018).

Este conjunto de datos contiene los registros médicos entre abril y diciembre de 2015. El conjunto de datos contiene 13 variables, que brindan información clínica, corporal y de estilo de vida que se describirán brevemente en el siguiente apartado. Es imperante anotar que los 299 pacientes tenían disfunción sistólica del ventrículo izquierdo y tenían insuficiencia cardíaca previa que los ubicaba en las clases III o IV de la clasificación de etapas de insuficiencia cardíaca de la New York Heart Association (Bredy et. al., 2017). En este sentido, el ejercicio apunta a responder la pregunta para aquellos pacientes que cumplen estas condiciones, no para aquellos que se encuentran en condiciones normales.

A partir de los datos, y pensando más en un problema de predicción se busca responder a la pregunta: **¿Qué tan probable es que un paciente con insuficiencia cardíaca en clases III o IV de la clasificación de la NYHA sufra un evento fatal relacionado con un ataque al corazón, basándose en sus características clínicas, corporales y de estilo de vida?** Al identificar los factores de mayor riesgo se pueden generar estrategias de detección temprana para un manejo oportuno de pacientes con mayor probabilidad de sufrir una muerte por un ataque al corazón.

2. Desarrollo de modelos

Durante el proceso de tratamiento de datos, se confirmó que la base de datos no presentaba datos faltantes ni nulos, lo cual facilitó la preparación. Las variables categóricas ya estaban definidas como booleanos, lo que simplificó su manejo en el modelado. Dado que las variables se encontraban en diferentes escalas, se procedió a estandarizarlas para asegurar un modelado apropiado y una mejor comparación entre ellas. No se identificó la necesidad de realizar tratamientos adicionales. Por último, los datos se dividieron en dos subconjuntos: un

subconjunto de entrenamiento, que comprendió el 67% de los datos, y un subconjunto de prueba, destinado a evaluar el desempeño de los modelos.

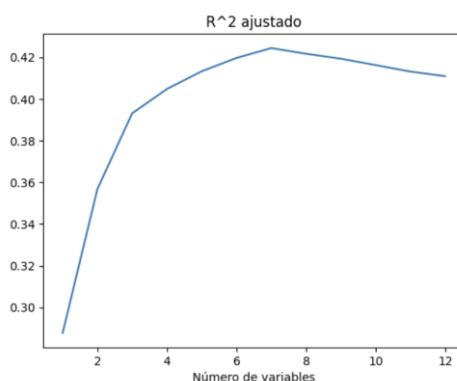
Dada la complejidad y la multitud de factores que contribuyen al desarrollo de enfermedades cardiovasculares — tales como hipertensión arterial, diabetes, colesterol alto, obesidad, falta de actividad física, y el consumo de tabaco y alcohol— se requiere de modelos predictivos capaces de manejar interacciones lineales y no lineales y la heterogeneidad en los datos de pacientes. Por lo tanto, se consideran varios candidatos para la modelización predictiva, cada uno con sus fortalezas en términos de precisión, capacidad de manejar grandes volúmenes de datos, habilidad para interpretar la importancia de las variables predictoras, así como su capacidad para trabajar con datos desbalanceados y muestras pequeñas.

Métodos de Ensamble (Bagging, Random Forest, Xtreme Gradient Boosting): Estos modelos capturan no linealidades y relaciones complejas entre los factores de riesgo, ofreciendo al mismo tiempo una interpretación visual intuitiva de las decisiones de clasificación.

La selección del modelo más adecuado dependerá de un balance entre precisión predictiva y la capacidad de interpretar qué factores son más significativos para el riesgo de enfermedad cardiovascular, considerando la naturaleza y calidad de los datos disponibles del Instituto de Cardiología de Faisalabad y el Hospital Allied. Este enfoque no solo busca responder a la pregunta de investigación planteada, sino también contribuir a una estrategia de intervención más informada y efectiva, que pueda reducir la prevalencia de muertes por enfermedades cardiovasculares a nivel mundial.

2.1. Selección de variables

Para la selección de variables en nuestro análisis, empleamos el método forward, una técnica progresiva cuyo criterio de elección se basó en identificar el modelo con el mejor coeficiente de determinación ajustado (R^2 ajustado). Este enfoque nos permitió evaluar y seleccionar, de manera incremental, las variables que contribuían significativamente al poder predictivo del modelo.



A pesar de la rigurosidad en la selección, los modelos generados no alcanzaron valores de R^2 ajustado particularmente altos. De hecho, el modelo con el mejor desempeño exhibió un R^2 ajustado ligeramente superior a 0.42, lo cual, aunque no es excepcionalmente alto, sí representa el mejor balance encontrado entre complejidad y capacidad explicativa dentro del conjunto de datos analizado.

Dicho modelo óptimo incorporó 7 de las 12 variables iniciales, descartando específicamente las variables 2 (Anemia), 4 (Diabetes), 6 (hipertensión), 7 (plaquetas) y 11 (Si es fumador). La exclusión de estas variables se justificó por su contribución marginal o nula al incremento del R^2 ajustado, lo que indica que su presencia no mejoraba, y en algunos casos podía incluso reducir, la eficacia predictiva del modelo. Este proceso selectivo

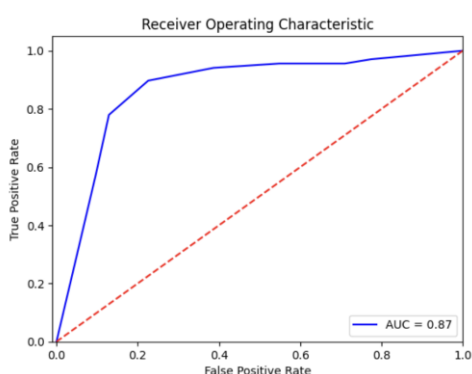
subraya la importancia de una elección cuidadosa de variables, priorizando aquellas que realmente aportan al entendimiento y predicción del fenómeno estudiado, mientras se minimiza el riesgo de sobreajuste y se preserva la generalidad del modelo.

2.2. Modelos

Se desarrollan modelos de ensamble dado que al ensamblar algoritmos se combinan con alguna estrategia que hace que se corrijan los vacíos de cada predictor individualmente. Generalmente, las estrategias buscan disminuir el sesgo o la varianza.

a. Bagging (Bootstrap aggregation):

Crea múltiples versiones del conjunto de datos original mediante muestreo con reemplazo (Bootstrap) y entrena un modelo base en cada submuestra. La predicción final se obtiene usando la votación mayoritaria, lo que ayuda a reducir la varianza y mejorar la estabilidad del modelo.



El error de clasificación para Bagging en la muestra de test es:
0.16161616161616163

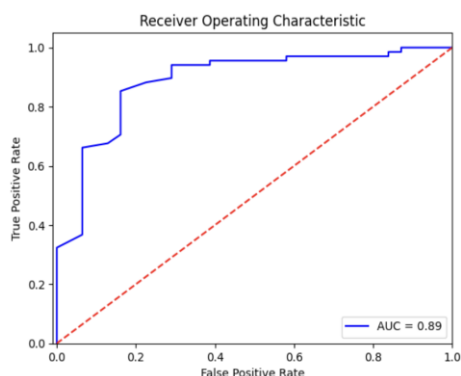
La matriz de confusión para Bagging en la muestra de test es:
[[64 4]
[12 19]]

El accuracy para Bagging en la muestra de test es:
0.8383838383838383

La precisión para Bagging en la muestra de test es:
0.8260869565217391

b. Random Forest:

Es una adaptación del modelo Bagging, en donde utiliza árboles de decisión como modelos base y agrega aleatoriedad al seleccionar un subconjunto aleatorio de características en cada división de nodo. Esto con el fin de crear mayor diversidad y que pueda disminuir la varianza.



El error de clasificación para Random Forest en la muestra de test es:
0.1313131313131313

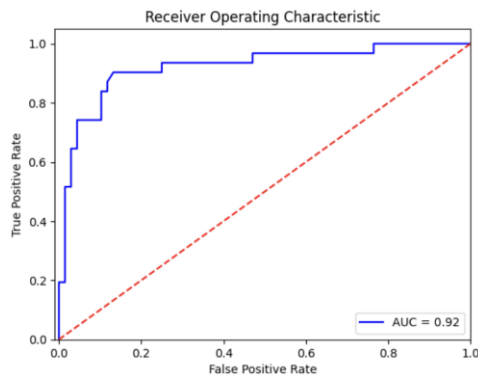
La matriz de confusión para Random Forest en la muestra de test es:
[[64 4]
[9 22]]

El accuracy para Random Forest en la muestra de test es:
0.8686868686868687

La precisión para Random Forest en la muestra de test es:
0.8461538461538461

c. Xtreme Gradient Boosting (XGBoost)

Es una implementación optimizada del algoritmo de Gradient Boosting, que entrena modelos de forma secuencial. Para este modelo cada árbol intenta corregir los errores del árbol anterior, optimizando una función de pérdida. En este caso se utiliza la función de pérdida “logloss”.



El error de clasificación para XGBoost en la muestra de test es:
0.1414141414141414

La matriz de confusión para XGBoost en la muestra de test es:
[[65 3]
[11 20]]

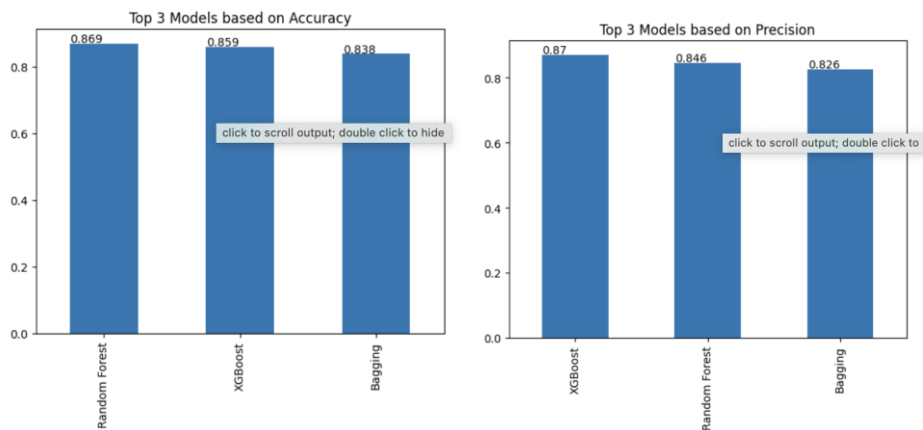
El accuracy para XGBoost en la muestra de test es:
0.8585858585858586

La precisión para XGBoost en la muestra de test es:
0.8695652173913043

Los modelos de ensamble como Bagging, Random Forest y XGBoost muestran un rendimiento superior, con el Random Forest alcanzando una precisión del 84.61% y un AUC de 0.89, y el XGBoost destacando con una precisión del 86.9565% y el AUC más alto de todos los modelos evaluados, un impresionante 0.92. Estos resultados subrayan la eficacia de los métodos de ensamble en la mejora de la precisión y la capacidad predictiva a través de la combinación de múltiples modelos para reducir la varianza y el sesgo.

En conclusión, estos modelos no lineales demuestran una capacidad robusta para abordar la predicción de eventos adversos relacionados con enfermedades cardiovasculares, con los métodos de Random Forest y XGBoost mostrando el mayor potencial.

3. Evaluación



El modelo Random Forest mostró una exactitud del 86.9%, una precisión del 84.6% y un AUC de 0.89, evidenciando su habilidad para manejar con eficacia la complejidad de los datos médicos. El XGBoost por su parte, registró una exactitud del 85.9%, la mayor precisión entre los modelos evaluados con un 86.9565% y el AUC más alto de 0.92, resaltando su capacidad para mejorar iterativamente las predicciones. Finalmente, el Bagging presentó una exactitud del 83.8%, una precisión del 82.6% y un AUC de 0.87, demostrando su efectividad en la reducción de la varianza y el mejoramiento de la estabilidad de las predicciones mediante la agregación de múltiples modelos.

Al considerar todas estas métricas de manera conjunta, se seleccionó el modelo que ofreció el mejor equilibrio entre minimizar el riesgo de clasificaciones incorrectas y maximizar la confiabilidad y precisión global de las

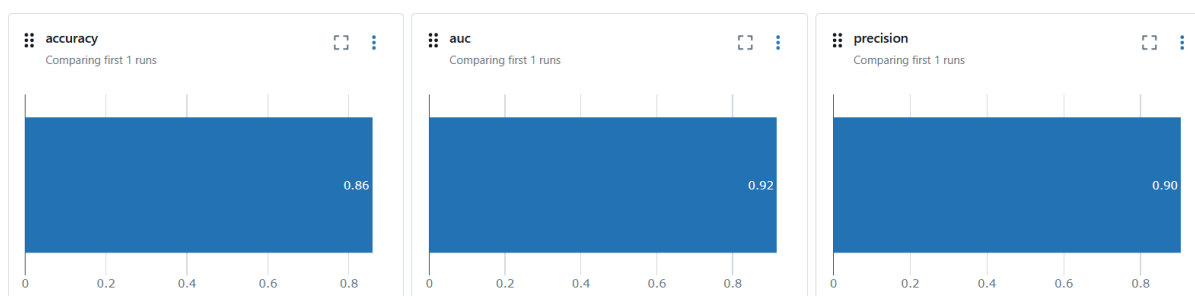
predicciones. Aunque los tres modelos demostraron ser altamente competentes, el XGBoost hasta este punto se presenta como el modelo óptimo debido a su excepcional precisión y el AUC más alto, lo que indica su superioridad en la identificación precisa de pacientes con alto riesgo de ataques cardíacos. De igual modo, se realizó una serie de experimentos cambiando diferentes hiperparámetros para identificar el mejor modelo posible.

3.1. Selección final del modelo con MLFlow

Se realizaron una serie de experimentos utilizando MLflow en DataBricks con el objetivo de identificar el modelo de machine learning con el mejor desempeño. En estos experimentos se implementaron los modelos Random Forest y XGBoost, seleccionados tras una evaluación preliminar que los destacó como las opciones más prometedoras en términos de rendimiento (ver anexos).

Al comparar los dos mejores modelos de cada técnica, los resultados mostraron que XGBoost superó a Random Forest en dos métricas clave: AUC y precisión. Esta selección se fundamentó en un análisis exhaustivo que ponderó todas las medidas de desempeño, posicionando a XGBoost como la opción más eficaz para su implementación en entornos clínicos, especialmente para optimizar la detección temprana y la prevención de enfermedades cardiovasculares.

El modelo final seleccionado fue XGBoost, configurado con una profundidad máxima de 10, un número de estimadores igual a 7 y una tasa de aprendizaje de 0.1. Esta configuración alcanzó un desempeño sobresaliente, logrando la mayor exactitud (0.8585), un AUC de 0.9160 y una precisión de 0.9047, lo que refuerza su capacidad de discriminación y su idoneidad para este propósito.



4. Observaciones y conclusiones del modelamiento

La implementación y selección meticulosa del modelo XGBoost, tras comparar diversas metodologías predictivas y diferentes experimentos, resalta un avance significativo en la capacidad de anticipar eventos cardíacos adversos. Este modelo, distinguido por su alta precisión, exactitud, y un AUC sobresaliente, demuestra una notable eficacia en identificar pacientes en alto riesgo de sufrir ataques cardíacos. La relevancia de esta solución radica no solo en su potencial para salvar vidas mediante la detección temprana y la intervención preventiva, sino también en su capacidad para optimizar los recursos médicos al focalizar esfuerzos en individuos de mayor riesgo, lo cual es de inestimable valor en contextos organizacionales y de sistemas de salud donde los recursos son limitados.

Este enfoque predictivo abre puertas a estrategias de manejo personalizadas, permitiendo a las organizaciones de salud implementar medidas preventivas más efectivas y mejorar la calidad del cuidado ofrecido a los pacientes. Al mismo tiempo, la adopción de este modelo subraya la importancia de la analítica avanzada y la ciencia de datos en la toma de decisiones médicas, marcando un paso adelante hacia la medicina predictiva y la salud preventiva.

No obstante, es crucial reconocer las posibles limitaciones de esta implementación, incluyendo la dependencia de la calidad y la representatividad de los datos disponibles, así como la necesidad de interpretar cuidadosamente los resultados para evitar sesgos en la toma de decisiones clínicas. Además, el modelo, aunque robusto, puede

requerir ajustes y validaciones continuas para adaptarse a nuevas poblaciones o cambios en los patrones de enfermedad.

A partir de los resultados de este proyecto, se plantea la oportunidad de realizar análisis más profundos y estudios posteriores que exploren la aplicabilidad del modelo en distintos contextos demográficos o geográficos, la integración de nuevas variables predictivas que puedan enriquecer el modelo, y el desarrollo de sistemas de alerta temprana que faciliten intervenciones aún más proactivas. Asimismo, sería beneficioso investigar la interacción entre los factores de riesgo identificados y otros determinantes sociales de la salud, para comprender más ampliamente cómo abordar de manera integral la prevención de enfermedades cardiovasculares. La solución desarrollada, por lo tanto, no solo representa un avance en la lucha contra las enfermedades cardiovasculares, sino que también establece un fundamento sólido para futuras investigaciones y mejoras en el campo de la salud predictiva y preventiva.

5. Dashboard – Funcionalidad

Link: <http://3.89.38.247:8050/>

Repo: https://github.com/JuanD13Perez/Proyecto_Final_Despliegue_Gr8

Video: [Funcionamiento](#)

El Dashboard “**Factores de Riesgo asociados a Enfermedades Cardíacas**” está diseñado para facilitar la exploración e interpretación de datos clínicos de pacientes con insuficiencia cardíaca. Contiene tres pestañas principales, cada una enfocada en diferentes aspectos del análisis. En la parte superior, se encuentra un panel con las tres pestañas de los diferentes datos que se pueden presentar (demográfico, antecedentes y factores), cuenta con un logotipo y unos indicadores superiores que presentan una información básica de la data.

El Dashboard cuenta con una sección de indicadores clave que presentan información resumida del conjunto de datos, como el número total de pacientes, la edad promedio, el tiempo de seguimiento y el número de factores de riesgo analizados. Adicionalmente, el tablero incluye tarjetas informativas que destacan estadísticas importantes y un formulario interactivo donde los usuarios pueden ingresar variables relacionadas con su salud. Al completar estas variables, el modelo predictivo evalúa la probabilidad de que la persona sufra un evento fatal asociado a un ataque cardíaco, ofreciendo un análisis personalizado y basado en los datos proporcionados.



- Demográfico:** Esta es la pestaña inicial que se muestra al acceder al enlace. En esta sección se ofrecen estadísticas descriptivas básicas, enfocadas en los datos demográficos y de seguimiento de los pacientes.

- **Distribución de la edad:** Histograma que muestra la frecuencia de pacientes por grupos etarios.
- **Pacientes fallecidos:** Gráfico de barras que compara la cantidad de pacientes fallecidos frente a los que sobrevivieron durante el periodo de estudio.
- **Distribución Género:** Diagrama circular que muestra la distribución de pacientes por género.

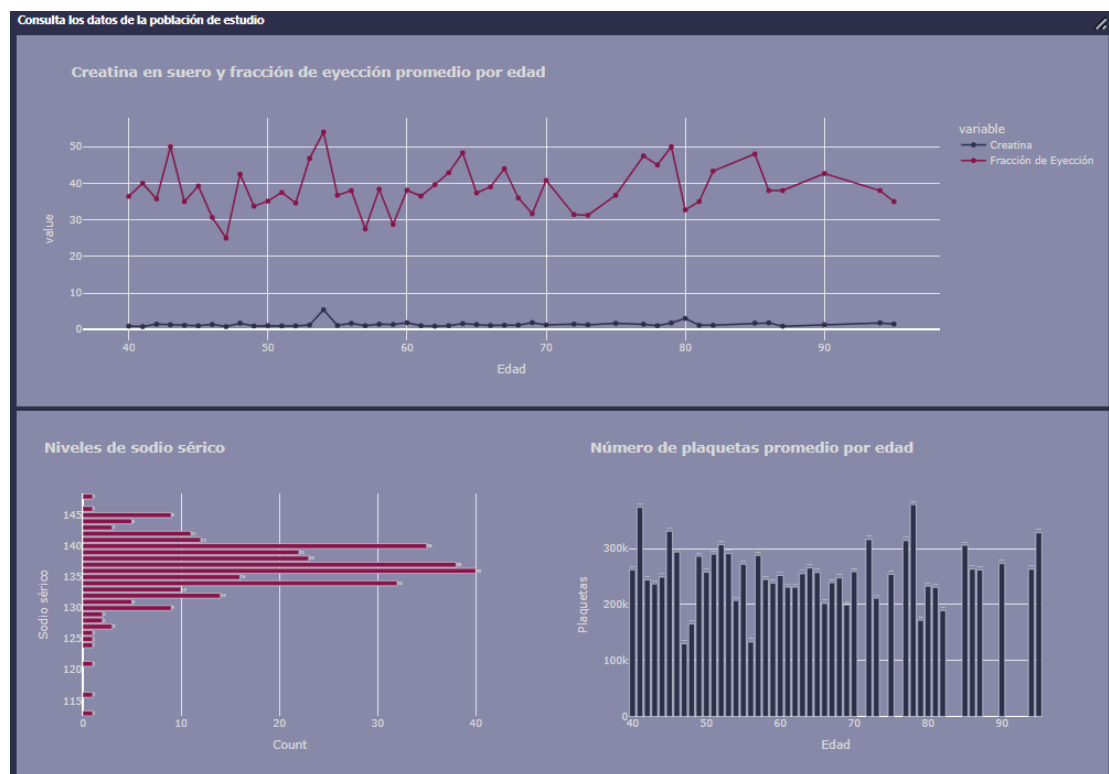


- b. **Antecedentes:** Aquí se presentan las visualizaciones con las cuales el usuario puede conocer sus antecedentes y el tiempo que tuvieron los pacientes.
- **Distribución de antecedentes:** Diagrama circular que muestra la proporción de pacientes con anemia, diabetes, hipertensión y si es fumador o no.
 - **Distribución de tiempo de seguimiento:** Gráfico de líneas que muestra la cantidad de días que estuvieron en seguimiento los pacientes.



c. **Factores:** En la sección de Factores, se analizan variables claves relacionadas con la salud cardiovascular. Esta pestaña incluye visualizaciones interactivas que permiten al usuario explorar y comprender los factores de riesgo más relevantes en pacientes con antecedentes de incidencias cardíacas. Estas visualizaciones facilitan la identificación de patrones, tendencias y relaciones entre los distintos factores, proporcionando una herramienta poderosa para el análisis y la toma de decisiones informadas.

- **Distribución de creatina en suero y fracción de eyección:** Gráfico de líneas que representa la relación entre la edad, los niveles de creatinina en suero y la fracción de eyección. En este análisis, una fracción de eyección inferior al 40% se asocia con insuficiencia cardíaca, mientras que niveles elevados de creatinina en suero indican una posible afectación renal, frecuentemente vinculada a insuficiencia cardíaca. Este gráfico permite identificar cómo estas variables varían con la edad y su impacto en la salud cardiovascular.
- **Distribución de sodio sérico:** Gráfico de barras que muestra la cantidad de pacientes distribuidos en distintos niveles de sodio sérico. Los pacientes con niveles inferiores a 135 mEq/L, indicativos de hiponatremia, suelen estar asociados con insuficiencia cardíaca, lo que resalta la importancia de este indicador en el monitoreo de esta condición.
- **Distribución de número de plaquetas:** Gráfico de barras con la distribución del número de plaquetas promedio por edad, en donde las alteraciones en el recuento de plaquetas pueden estar relacionadas con complicaciones asociadas a la insuficiencia cardíaca.



El dashboard “Factores de Riesgo Asociados a Enfermedades Cardíacas” proporciona una herramienta integral para analizar y comprender los datos clínicos de pacientes con insuficiencia cardíaca. A través de sus visualizaciones interactivas, permite identificar patrones clave en factores como la fracción de eyección, el sodio sérico y el recuento de plaquetas, destacando su relación con la progresión y las complicaciones de la enfermedad. Así mismo, permite conocer los antecedentes de los pacientes que tienen insuficiencia cardíaca. Además, su sección predictiva, basada en variables específicas de los

pacientes, facilita la evaluación del riesgo individual de sufrir eventos cardíacos fatales, lo que contribuye a una mejor toma de decisiones clínicas y estrategias de intervención personalizadas.

6. Conclusiones

El proyecto ha demostrado el potencial significativo del modelo XGBoost para anticipar eventos cardíacos adversos con alta precisión y exactitud, lo que representa un avance clave en la medicina predictiva. Este modelo alcanzó un desempeño sobresaliente, evidenciado por su exactitud de 0.8585, un AUC de 0.9160 y una precisión de 0.9047. Estos resultados destacan su capacidad para discriminar eficazmente entre pacientes con y sin riesgo de eventos cardíacos adversos, consolidando su idoneidad como herramienta predictiva en este contexto, permitiendo optimizar los recursos médicos, focalizando la atención en aquellos que más lo necesitan.

Este nivel de rendimiento no solo subraya su robustez y confiabilidad, sino que también refuerza su potencial para apoyar intervenciones preventivas oportunas y decisiones clínicas informadas. Sin embargo, es fundamental continuar validando y ajustando el modelo con datos actualizados y representativos para garantizar su efectividad en diversas poblaciones y contextos. Este enfoque abre la puerta a nuevas investigaciones, como la integración de variables adicionales que puedan mejorar la precisión del modelo y permitir una prevención más personalizada.

El Dashboard desarrollado complementa este avance ya que proporciona una herramienta interactiva esencial para analizar los factores de riesgo asociados a enfermedades cardíacas, facilitando la interpretación de datos clínicos mediante visualizaciones claras y accesibles. A través de sus distintas secciones, los usuarios pueden explorar los antecedentes y características demográficas de los pacientes, así como los factores de riesgo clave que afectan la salud cardiovascular. Esta plataforma no solo ayuda en la toma de decisiones informadas, sino que también ofrece un enfoque personalizado al evaluar el riesgo individual de sufrir eventos fatales, contribuyendo a una mejor gestión de la salud y promoviendo intervenciones más efectivas y preventivas.

En conjunto, tanto el modelo como el Dashboard abren nuevas oportunidades para avanzar en la prevención de enfermedades cardiovasculares y en la investigación de nuevos factores que puedan optimizar aún más la precisión del sistema

7. Avances de equipo

A continuación, se presentan las actividades en las que trabajo cada uno de los integrantes del equipo:

Nombre responsable	Actividad
Jorge Andrés Rodríguez	Planteamiento del problema
Juan Diego Pérez	Extracción y preprocesamiento de la información
Diana Bayona – Jorge Andrés Rodríguez – Juan Diego Pérez – Santiago Gutiérrez	Análisis preliminar de la información
Diana Bayona – Jorge Andrés Rodríguez – Juan Diego Pérez – Santiago Gutiérrez	Selección a priori de los modelos
Diana Bayona – Santiago Gutiérrez	Desarrollo de los modelos y experimentación
Diana Bayona – Santiago Gutiérrez	Diseño Dashboard prototipo
Juan Diego Pérez - Jorge Andrés Rodríguez	Desarrollo de Dashboard en Dash
Diana Bayona – Santiago Gutiérrez	Empaquetamiento del modelo
Juan Diego Pérez - Jorge Andrés Rodríguez	Despliegue de la API y el Dashboard
Diana Bayona - Jorge Andrés Rodríguez	Manual de usuario del tablero

Juan Diego Pérez - Santiago Gutiérrez	Manual de instalación del tablero.
Diana Bayona – Jorge Andrés Rodríguez – Juan Diego Pérez – Santiago Gutiérrez	Documento final

8. Bibliografía

- Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA. Survival analysis of heart failure patients: a case study. PLoS ONE. 2017; 12(7):0181001
- Bredy C, Ministeri M, Kempny A, Alonso-Gonzalez R, Swan L, Uebing A, Diller G-P, Gatzoulis MA, Dimopoulos K. New York Heart Association (NYHA) classification in adults with congenital heart disease: relation to objective measures of exercise and outcome. Eur Heart J – Qual Care Clin Outcomes. 2017; 4(1):51–8.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer
- Kuhn, M., & Johnson, K. (2013). Applied predictive modellingSpringer. New York Heidelberg Dordrecht London
- Plotly (2024). Plotly Express user guide. Plotly. <https://plotly.com/python/plotly-express/>

9. Anexos

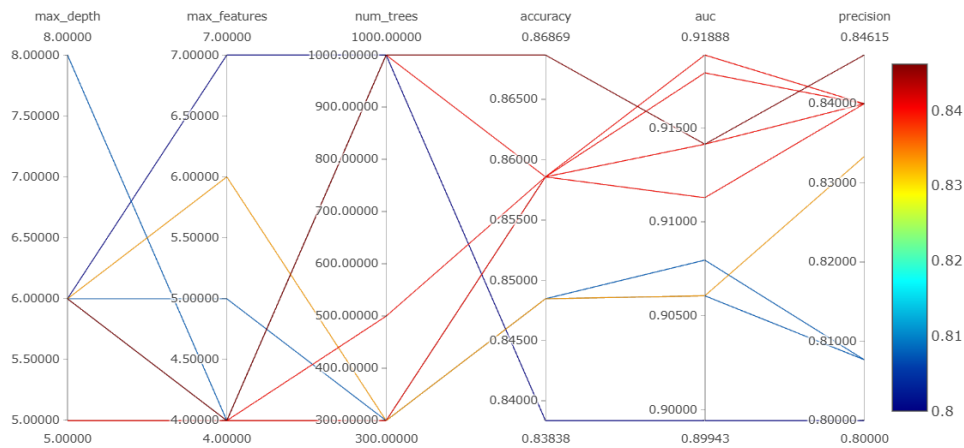
- Experimentos en Databricks**
Usuario de Databricks: d.bayonap@uniandes.edu.co
- Modelo de Random Forest**

Para el modelo de Random Forest, se llevaron a cabo 10 experimentos en los que se evaluaron diferentes configuraciones de hiperparámetros, incluyendo la profundidad máxima de los árboles (max_depth), el número de características consideradas en cada división (max_features) y el número de árboles (num_trees). Las métricas analizadas incluyeron la precisión, el AUC (Área Bajo la Curva ROC) y la exactitud (accuracy), con el objetivo de comparar y evaluar el rendimiento de los distintos modelos. A partir de este análisis, se seleccionó la configuración óptima que maximizó el desempeño del modelo.

Resultados de los experimentos – Random Forest

					Metrics			Parameters		
	Run Name	Created	Duration	Source	accuracy	auc	precision	max_depth	max_features	num_tr
<input type="checkbox"/>	overjoyed-hound-697	15 days ago	5.2s	Modelo ...	0.858585858...	0.917931688...	0.84	5	4	1000
<input type="checkbox"/>	bold-shad-159	15 days ago	5.4s	Modelo ...	0.838383838...	0.899430740...	0.8	6	7	1000
<input type="checkbox"/>	melodic-worm-962	15 days ago	5.3s	Modelo ...	0.868686868...	0.914136622...	0.846153846...	6	4	1000
<input type="checkbox"/>	exultant-moose-911	15 days ago	4.2s	Modelo ...	0.848484848...	0.906072106...	0.807692307...	8	4	300
<input type="checkbox"/>	capable-kit-600	15 days ago	4.4s	Modelo ...	0.858585858...	0.918880455...	0.84	5	4	300
<input type="checkbox"/>	resilient-newt-888	15 days ago	4.5s	Modelo ...	0.858585858...	0.911290322...	0.84	6	4	500
<input type="checkbox"/>	wistful-sponge-576	15 days ago	4.4s	Modelo ...	0.858585858...	0.914136622...	0.84	6	4	300
<input type="checkbox"/>	bedecked-goat-377	15 days ago	4.3s	Modelo ...	0.848484848...	0.906072106...	0.833333333...	6	6	300

Comparación de los experimentos – Random Forest



Los experimentos realizados con Random Forest revelaron que el mejor modelo es aquel configurado con una profundidad máxima de 6, un número de características de 4 y un total de 1000 árboles. Esta configuración mostró un rendimiento sobresaliente, obteniendo el valor más alto de exactitud (0.8686) y precisión (0.8461). Aunque en el AUC (0.9141) se posicionó como el segundo más alto, sigue siendo una métrica sólida que respalda su capacidad de discriminación. Estos resultados indican que el modelo es altamente eficaz para clasificar correctamente a los pacientes con insuficiencia cardíaca y predecir si tienen un alto riesgo de sufrir un evento fatal, como un ataque al corazón, utilizando características clínicas, corporales y relacionadas con su estilo de vida.

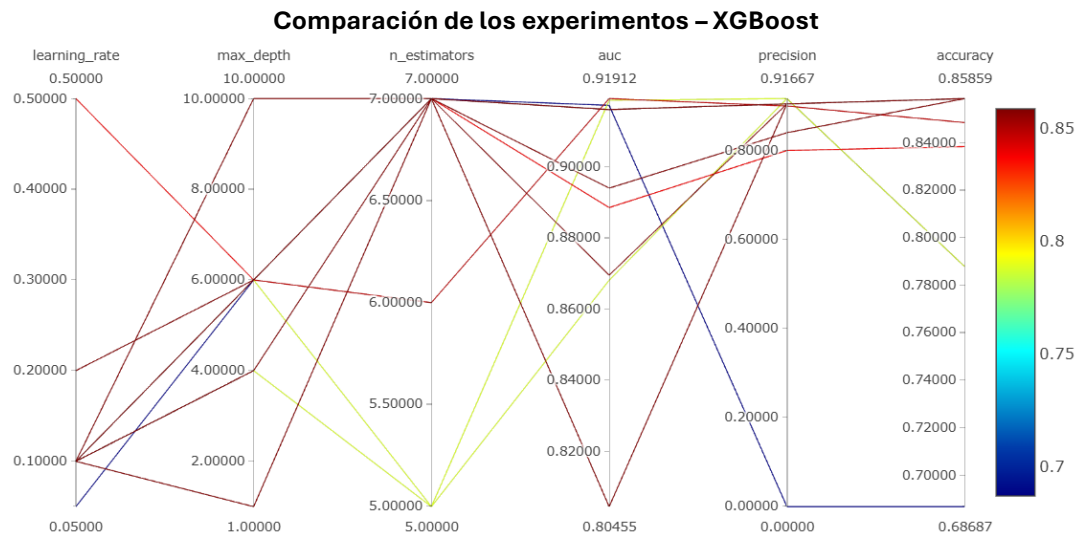
- ### Modelo de XGBoost

Resultados de los experimentos – XGBoost

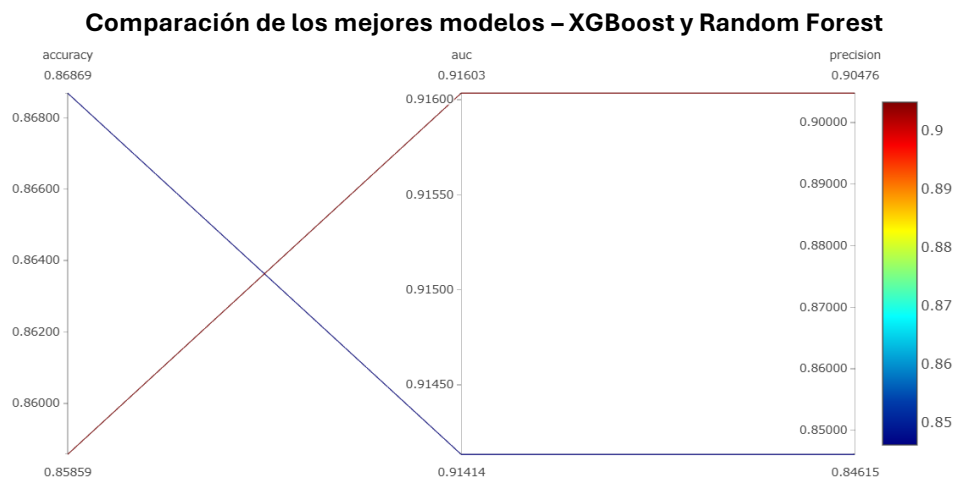
				Metrics			Parameters			
	Run Name	Created	ion	Source	accuracy	auc	precision	learning_rate	max_depth	n_estimators
	unique-sow-44	15 days ago		Xgboost	0.848484848...	0.919117647...	0.9	0.1	6	6
	spiffy-finch-482	15 days ago		Xgboost	0.858585858...	0.893975332...	0.84	0.2	6	7
	grandiose-shrew-875	15 days ago		Xgboost	0.838383838...	0.888519924...	0.8	0.5	6	7
	popular-skunk-858	15 days ago		Xgboost	0.858585858...	0.916034155...	0.904761904...	0.1	10	7
	ambitious-jay-804	15 days ago		Xgboost	0.858585858...	0.869544592...	0.904761904...	0.1	4	7
	stylish-donkey-188	15 days ago		Xgboost	0.787878787...	0.868121442...	0.916666666...	0.1	4	5
	peaceful-pig-220	15 days ago		Xgboost	0.787878787...	0.918643263...	0.916666666...	0.1	6	5
	dashing-ape-613	15 days ago		Xgboost	0.686868686...	0.917220113...	0	0.05	6	7

En el caso del modelo de XGBoost, se realizaron 10 experimentos donde se evaluaron distintas configuraciones de hiperparámetros para optimizar el rendimiento. Entre los hiperparámetros considerados se incluyeron la profundidad máxima de los árboles (max_depth), el número de estimadores (n_estimators) y la tasa de aprendizaje (learning_rate). Las métricas clave analizadas fueron la precisión, el AUC (Área Bajo la Curva ROC) y la exactitud (accuracy), lo que permitió una comparación detallada del rendimiento de las distintas configuraciones.

Los experimentos realizados con XGBoost demostraron que el mejor modelo es aquel configurado con una profundidad máxima de 10, un número de estimadores de 7 y una tasa de aprendizaje de 0.1. Esta configuración logró un rendimiento destacado, alcanzando la exactitud más alta (0.8585). Aunque en las métricas de AUC (0.9160) y precisión (0.9047) se posicionó en el segundo lugar, los valores obtenidos se acercaron mucho a los más altos, lo que refuerza la solidez de su capacidad de discriminación. Estos resultados sugieren que el modelo es muy eficaz para clasificar adecuadamente a los pacientes con insuficiencia cardíaca y predecir su riesgo de sufrir un evento fatal, como un ataque al corazón, basado en características clínicas, corporales y de estilo de vida.



- **Modelo final**



Finalmente, se compararon los dos mejores modelos de XGBoost y Random Forest, y los resultados mostraron que XGBoost superó a Random Forest en dos de las métricas clave: AUC y precisión. Esta elección se basó en un análisis comprensivo que ponderó todas las medidas de desempeño, destacando a XGBoost como la herramienta más eficaz para la implementación en entornos clínicos, con el fin de optimizar la detección temprana y la prevención de enfermedades cardiovasculares.