Institut de Formació Contínua-IL3
UNIVERSITAT DE BARCELONA

UNIVERSITAT DE BARCELONA

Ciencia de los datos (Data Science), Aplicaciones en Biología y Medicina con Python y R

Predicción de ataques al corazón en una población a través de un análisis exploratorio, un análisis multivariante y un modelo de Machine Learning (Random Forest)

Luis Antelo Laguna, Weronika Balcerzak, Maya Carvalho-Evans, Oscar Rojas y Àlex Francisco Sigüencia

30 de Junio de 2025

ÍNDICE

1. INTRODUCCIÓN

Los infartos agudo de miocardo son un creciente desafío de salud global, caracterizado por la incapacidad del corazón para bombear sangre de manera efectiva para satisfacer las necesidades del cuerpo. Su prevalencia está aumentando, impulsada por el envejecimiento poblacional y la mejor supervivencia tras eventos cardiovasculares previos (Groenewegen et al., 2020). Se estima que 64.3 millones de personas viven con insuficiencia cardíaca en todo el mundo, lo que refleja una importante carga de enfermedad a nivel global (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018).

En la evaluación del riesgo cardiovascular, diversas variables clínicas y características individuales han demostrado tener un valor predictivo significativo tanto en el diagnóstico como en el pronóstico de enfermedades cardíacas. Estas variables pueden dividirse en factores de riesgo no modificables, como la edad, el sexo o los antecedentes familiares, y en factores modificables, cuya alteración puede prevenir o retrasar la aparición de eventos cardiovasculares. Una de las primeras variables consideradas es la frecuencia cardíaca, entendida como el número de latidos por minuto. La alteración de este parámetro, especialmente su variabilidad, se ha asociado con disfunciones autonómicas y con un incremento en la mortalidad cardiovascular, al reflejar un estado elevado de estrés fisiológico sobre el sistema cardíaco (Rajendra Acharya et al., 2006). La presión arterial sistólica y diastólica son también parámetros fundamentales. Niveles persistentemente elevados constituyen un marcador diagnóstico de hipertensión arterial, condición estrechamente relacionada con un mayor riesgo de eventos cardiovasculares mayores (Flint et al., 2019; Khera et al., 2021). En cuanto al metabolismo glucídico, se ha demostrado que niveles elevados de glucosa en sangre, incluso en ausencia de diagnóstico de diabetes, aumentan el riesgo de insuficiencia cardíaca y enfermedad coronaria (Nielson & Lange, 2005; American Diabetes Association, 2023).

En el ámbito de los biomarcadores cardíacos, la medición de CK-mb (isoenzima MB de la creatin cinasa) y troponina I resulta clave para la detección de daño miocárdico. Aunque tradicionalmente se emplean para confirmar el diagnóstico de infarto agudo de Miocardo, su elevación también se observa en contextos de insuficiencia cardíaca o isquemia crónica, siendo considerados indicadores de mal pronóstico (Yilmaz et al., 2006; Apple, 1999; Thygesen et al., 2018).

Por tanto, para este proyecto se realizará un estudio estadístico predictivo sobre una población de pacientes que han sufrido o no ataques al corazón (caso de resultado positivo o negativo).

2. OBJETIVOS

Para este proyecto se propusieron los siguientes tres objetivos:

Objetivo 1: Hacer un análisis descriptivo y exploratorio de los datos observados en la población de estudio, comprobar si las variables continuas siguen una distribución normal, comprobar si las medias y medianas de las variables bioquímicas entre casos de ataques al corazón positivos y negativos son significativamente diferentes o no.

Objetivo 2: Hacer un análisis multivariante de los datos observados. Al comparar todas las variables en el conjunto de datos, que incluye más de una variable dependiente, nuestro objetivo es entender las causas y las relaciones entre las variables. Además, debido a que nuestro conjunto de datos tiene más de un resultado (si el individuo tiene un ataque al corazón o no), podemos observar la influencia de las variables y su efecto combinado en el resultado.

Objetivo 3: Entrenar un modelo de Machine Learning (Random Forest) predictivo capaz de estimar la probabilidad de Infarto Agudo del Miocardio en función de las variables analizadas y previamente pre-procesadas. Evaluar el rendimiento de este modelo e interpretar los datos arrojados. Despliegue de un sitio web para la comprensión visual de manera interactiva de lo mencionado anteriormente.

3. MATERIAL Y MÉTODOS

El conjunto de datos de ataques al corazón se recopiló en el hospital Zheen en la ciudad de Erbil, Irak. Luego fueron depositados en las bases de datos Mendeley Data (Rashid & Hassan, 2022) y Kaggle (Heart Attack Dataset, n.d.). Se analizaron las variables fisiológicas de los pacientes como su edad [años], género [Masculino = 1; Femenino = 0], frecuencia cardíaca [lpm], presión arterial sistólica [mmHg] y presión arterial diastólica [mmHg]. También se ha analizado variables bioquímicas como el azúcar en sangre [mg/dL], concentración de CK-mb [ng/mL] y concentración de Troponina [ng/mL].

3.1 Metodologías para Objetivo 1

Para cumplir este objetivo se trabajó con los datos a través de herramientas como *RStudio* y *Google Colaboratory* con entorno de ejecución *R* y *Python*. Además, se usaron paquetes como *skimr*, *dplyr*, *ggplot2* y *DataExplorer.* También se aplicaron pruebas *t* de Student, como la prueba Shaphiro-Wilk, la prueba D'Agostino-Pearson y la prueba de Welch.

3.2 Metodologías para Objetivo 2

Para cumplir este objetivo de análisis multivariante se trabajó con los datos a través de herramientas como *Google Colaboratory* con entorno de ejecución *Python*. Además, se usaron paquetes como *pandas, numpy, matplotlib.pyplot, math, seaborn, sklearn (para preprocessing, decomposition, cluster), y statsmodels.api.* También se aplicaron pruebas de mapa de calor, diagrama de dispersión de matriz, clustering y regresión logística.

3.3 Metodologías para Objetivo 3

Para cumplir con este objetivo, se trabajó en un entorno local de desarrollo utilizando VSCode y Conda. Se empleó Jupyter Notebook para el preprocesamiento inicial del dataset. Posteriormente, se utilizaron scripts de Python para la preparación, entrenamiento y evaluación del modelo. Además, se implementó Streamlit para la interactividad web.

A lo largo del desarrollo, se utilizaron diversos paquetes como NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn y Joblib. Los procesos se automatizaron mediante un archivo Makefile y se exportaron los requisitos para asegurar la reproducibilidad futura del proyecto.

4. RESULTADOS

4.1 Objetivo 1. Obtención de resultados

4.1.1 Resultados del análisis exploratorio

Se realizó una descripción estadística de los datos de población y se obtuvo primeramente un resumen.

```
── Variable type: character ─────────────────────────────────────────────────
  skim_variable n_missing complete_rate min max empty n_unique whitespace
1 Gender                0             1   4   6     0        2          0
2 Result                0             1   8   8     0        2          0

── Variable type: numeric ───────────────────────────────────────────────────
  skim_variable            n_missing complete_rate    mean    sd     p0     p25
1 Age                              0             1  56.2   13.6  14      47
2 Heart.rate                       0             1  76.1   15.4  20      64
3 Systolic.blood.pressure          0             1 127.    26.1  42     110
4 Diastolic.blood.pressure         0             1  72.3   14.0  38      62
5 Blood.sugar                      0             1 147.    74.9  35      98
6 CK.MB                            0             1  15.3   46.3   0.321   1.65
7 Troponin                         0             1   0.361  1.15  0.001   0.006
```

```
── Data Summary ────────────────────
                           Values
Name                       Medical
Number of rows             1319
Number of columns          9
_____________________
Column type frequency:
  character                2
  numeric                  7
_____________________
Group variables            None
```

Figura 1. Resumen del conjunto de datos.

Se realizó otro resumen de los datos, pero las variables cuantitativas fueron agrupadas por el tipo de resultado obtenido de ataque al corazón. A más, se hizo una gráfica de barras sobre los contajes de las muestras clasificados por las variables categóricas.

Group variables Result

── Variable type: numeric ──

	skim_variable	Result	n_missing	complete_rate	mean	sd
1	Age	negative	0	1	52.1	13.7
2	Age	positive	0	1	58.8	13.0
3	Heart.rate	negative	0	1	75.9	14.9
4	Heart.rate	positive	0	1	76.2	15.6
5	Systolic.blood.pressure	negative	0	1	128.	27.0
6	Systolic.blood.pressure	positive	0	1	127.	25.5
7	Diastolic.blood.pressure	negative	0	1	72.4	14.3
8	Diastolic.blood.pressure	positive	0	1	72.2	13.9
9	Blood.sugar	negative	0	1	150.	78.4
10	Blood.sugar	positive	0	1	145.	72.6
11	CK.MB	negative	0	1	2.56	1.37
12	CK.MB	positive	0	1	23.3	57.7
13	Troponin	negative	0	1	0.0270	0.443
14	Troponin	positive	0	1	0.571	1.39

Figura 2. Datos de las variables cuantitativas agrupadas por tipo de resultado.

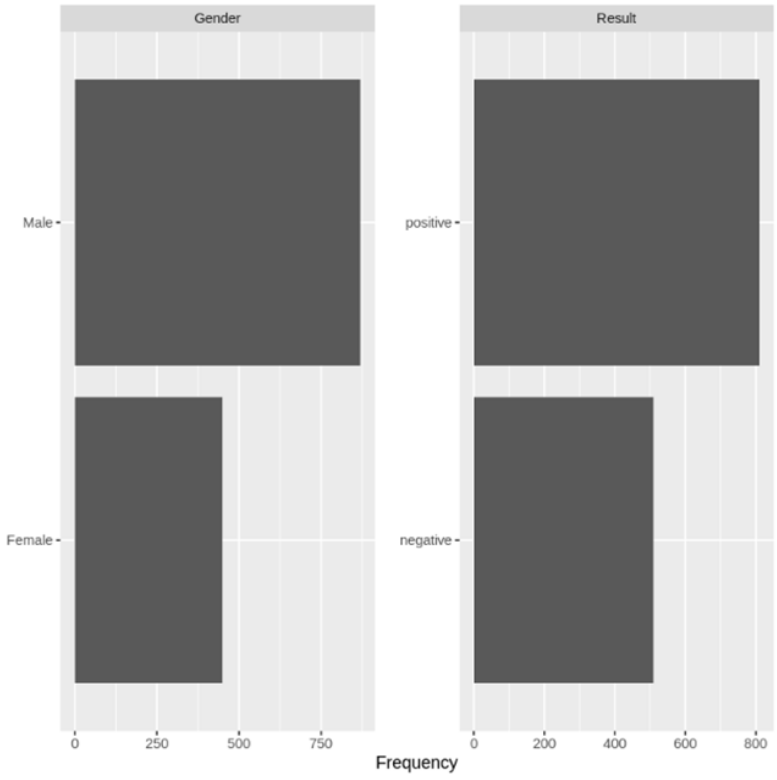


Figura 3. Datos de las variables cuantitativas agrupadas por tipo de resultado y género.

A continuación, se realizó otra vez un resumen de los datos, pero las variables cuantitativas fueron agrupadas por el tipo de resultado obtenido de ataque al corazón y el género de los pacientes.

```
Group variables            Result, Gender

── Variable type: numeric ──────────────────────────────────────────────────────
   skim_variable             Result   Gender n_missing complete_rate      mean          sd
 1 Age                       negative Female         0             1  53.8        1 14.2
 2 Age                       negative Male           0             1  51.0        2 13.3
 3 Age                       positive Female         0             1  61.3        3 13.3
 4 Age                       positive Male           0             1  57.6        4 12.6
 5 Heart.rate                negative Female         0             1  76.3        5 14.6
 6 Heart.rate                negative Male           0             1  75.7        6 15.1
 7 Heart.rate                positive Female         0             1  75.4        7 14.6
 8 Heart.rate                positive Male           0             1  76.5        8 16.1
 9 Systolic.blood.pressure   negative Female         0             1 128.         9 26.5
10 Systolic.blood.pressure   negative Male           0             1 128.        10 27.4
11 Systolic.blood.pressure   positive Female         0             1 126.        11 26.7
12 Systolic.blood.pressure   positive Male           0             1 127.        12 25.0
13 Diastolic.blood.pressure  negative Female         0             1  72.8       13 13.7
14 Diastolic.blood.pressure  negative Male           0             1  72.2       14 14.8
15 Diastolic.blood.pressure  positive Female         0             1  72.2       15 13.9
16 Diastolic.blood.pressure  positive Male           0             1  72.2       16 13.8
17 Blood.sugar               negative Female         0             1 142.        17 69.5
18 Blood.sugar               negative Male           0             1 155.        18 83.4
19 Blood.sugar               positive Female         0             1 150.        19 76.9
20 Blood.sugar               positive Male           0             1 143.        20 70.7
21 CK.MB                     negative Female         0             1   2.40      21  1.22
22 CK.MB                     negative Male           0             1   2.66      22  1.45
23 CK.MB                     positive Female         0             1  23.7       23 56.1
24 CK.MB                     positive Male           0             1  23.1       24 58.5
25 Troponin                  negative Female         0             1   0.00648   25  0.00337
26 Troponin                  negative Male           0             1   0.0405    26  0.571
27 Troponin                  positive Female         0             1   0.459     27  1.35
28 Troponin                  positive Male           0             1   0.620     28  1.41
```

Figura 4. Barplots sobre las frecuencias de contaje de las variables categóricas.

4.1.2 Resultados de comprobación de distribución normal.

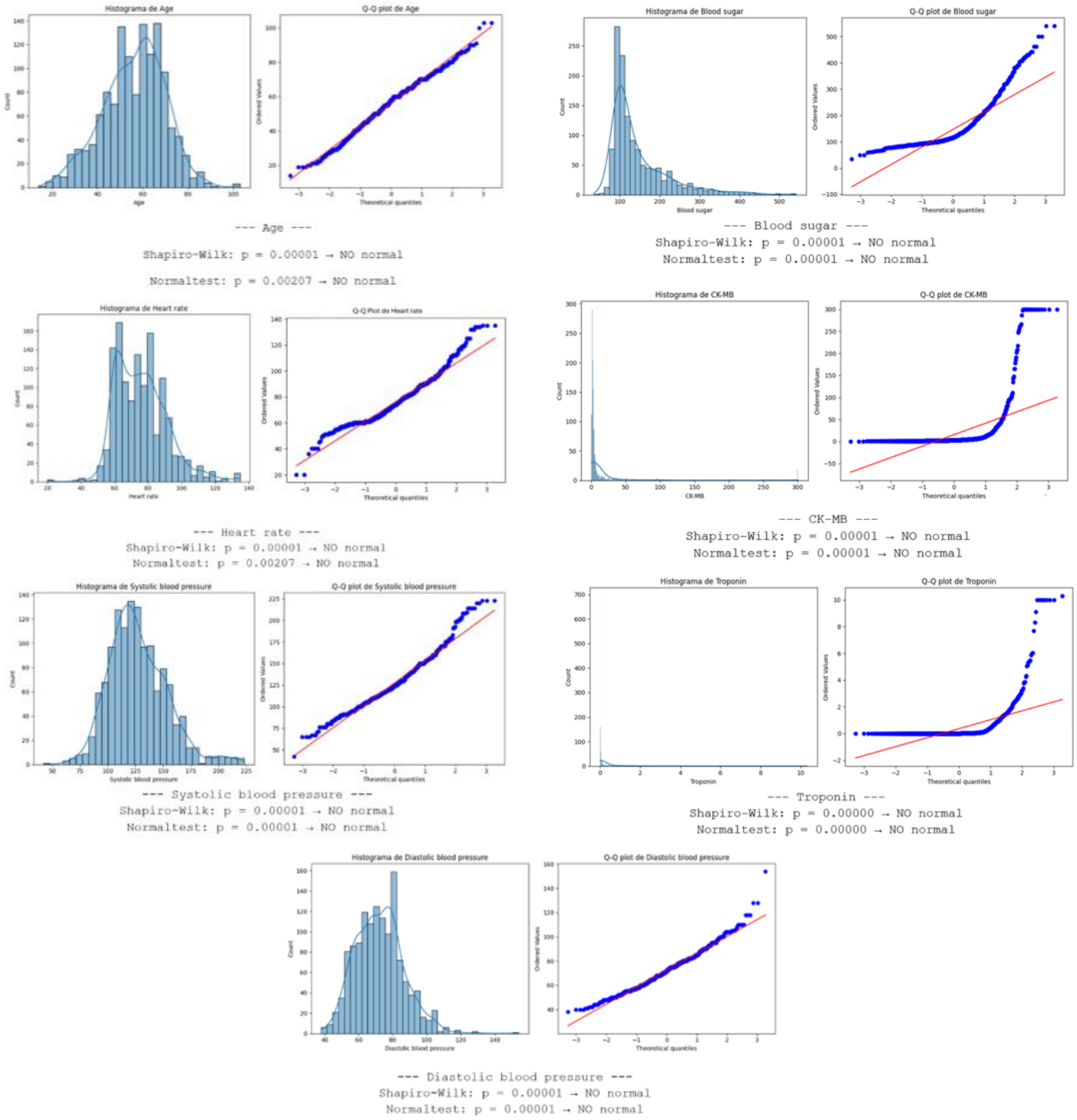
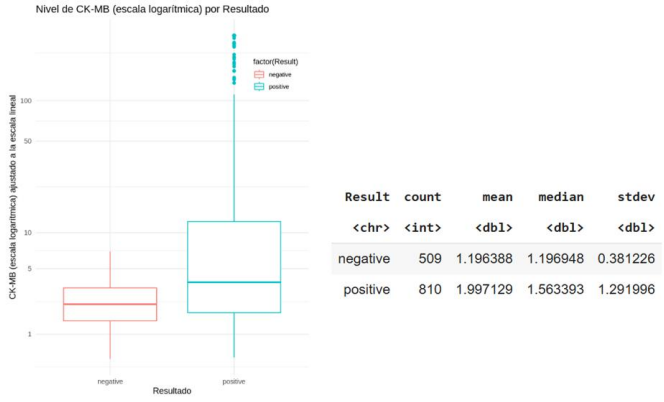


Figura 5. Histogramas y gráfica Q-Q plot sobre datos de la variable edad, frecuencia cardíaca, pressión arterial sistólica y pressión arterial diastólica, azúcar en sangre, CK-mb y Troponina.

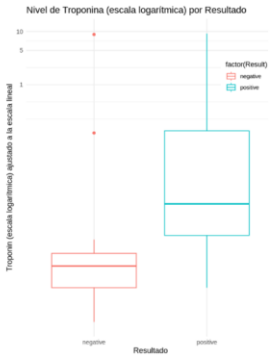
4.1.3 Resultados sobre las variables bioquímicas

Se analizó las medias, medianas y desviaciones estándares de los niveles de las tres variables bioquímicas, y fueron comparadas respecto los de caso de resultado positivo con los de resultado negativo. En las variables azúcar en sangre, CK-mb y Troponina se muestran resultados donde son presentados en escala logarítmica debido a que, si se presentan en escala lineal, ciertos valores se pueden visualizar y malinterpretar como valores outlayers (Magwene, 2025). Los resultados de medias y medianas son mostrados mediante cálculo logarítmico.



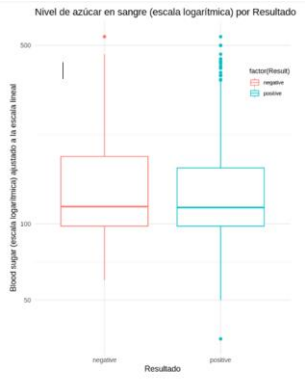
Result	count	mean	median	stdev
<chr>	<int>	<dbl>	<dbl>	<dbl>
negative	509	1.196388	1.196948	0.381226
positive	810	1.997129	1.563393	1.291996

```
        Welch Two Sample t-test

data:  CK.MB_log by Result
t = -16.531, df = 1017.6, p-value < 2.2e-16
alternative hypothesis: true difference in means between group negative and group positive is not equal to 0
95 percent confidence interval:
 -0.8957931 -0.7056896
sample estimates:
mean in group negative mean in group positive
              1.196388               1.997129
```



Result	count	mean	median	stdev
<chr>	<int>	<dbl>	<dbl>	<dbl>
negative	509	-5.148407	-5.115996	0.6560319
positive	810	-2.594000	-3.123566	2.1053561

```
        Welch Two Sample t-test

data:  Troponin_log by Result
t = -32.137, df = 1038.8, p-value < 2.2e-16
alternative hypothesis: true difference in means between group negative and group positive is not equal to 0
95 percent confidence interval:
 -2.710376 -2.398439
sample estimates:
mean in group negative mean in group positive
             -5.148407              -2.594000
```



A tibble: 2 × 5

Result	count	mean	median	stdev
<chr>	<int>	<dbl>	<dbl>	<dbl>
negative	509	4.914015	4.770685	0.4249508
positive	810	4.888383	4.762174	0.4072459

```
        Welch Two Sample t-test

data:  Blood.sugar_log by Result
t = 1.0836, df = 1045, p-value = 0.2788
alternative hypothesis: true difference in means between group negative and group positive is not equal to 0
95 percent confidence interval:
 -0.02078407  0.07204714
sample estimates:
mean in group negative mean in group positive
              4.914015               4.888383
```

Figura 6. Datos sobre los niveles de azúcar en sangre, CK-mb y Troponina de muestras de resultado positivo y de resultado negativo.

4.2 Objetivo 2. Obtención de resultados

La matriz de correlación de variables numéricas se visualizó con un mapa de calor, mostrando la fuerza y dirección de las relaciones lineales entre estas variables. Se trazó una matriz de correlación para entender qué tan estrechamente relacionadas están las variables entre sí, los valores más cercanos a 1 indican una correlación positiva más fuerte.

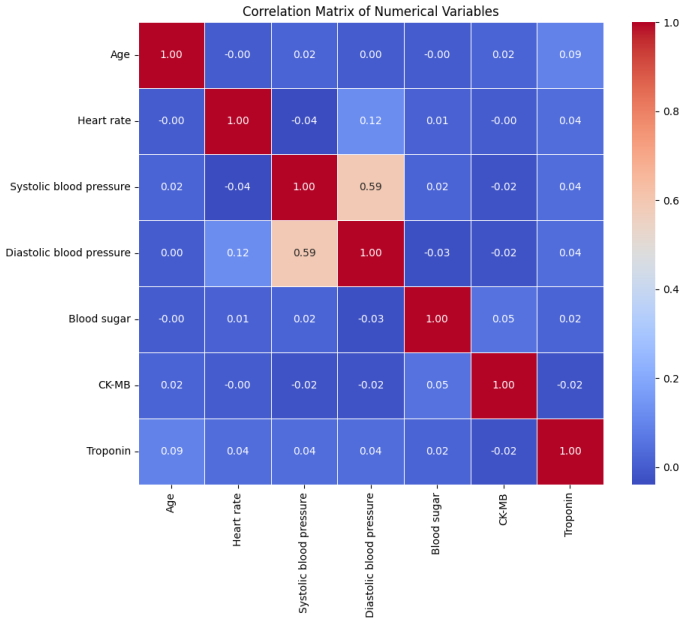


Figura 7. Mapa de calor de la matriz de correlación de las variables numéricas

Las correlaciones de las variables numéricas con el evento de infarto de miocardio se presentan en la Tabla 1.

Tabla 1. Correlaciones de las variables numéricas con el evento de infarto de miocardio

Variable	**Correlation with Result**
Result	1
Age	0,24
Troponin	0,23
CK-MB	0,22
Heart rate	0,01
Diastolic blood pressure	-0,01
Systolic blood pressure	-0,02
Blood sugar	-0,03

Matriz Diagrama explora visualmente los datos y sus patrones.

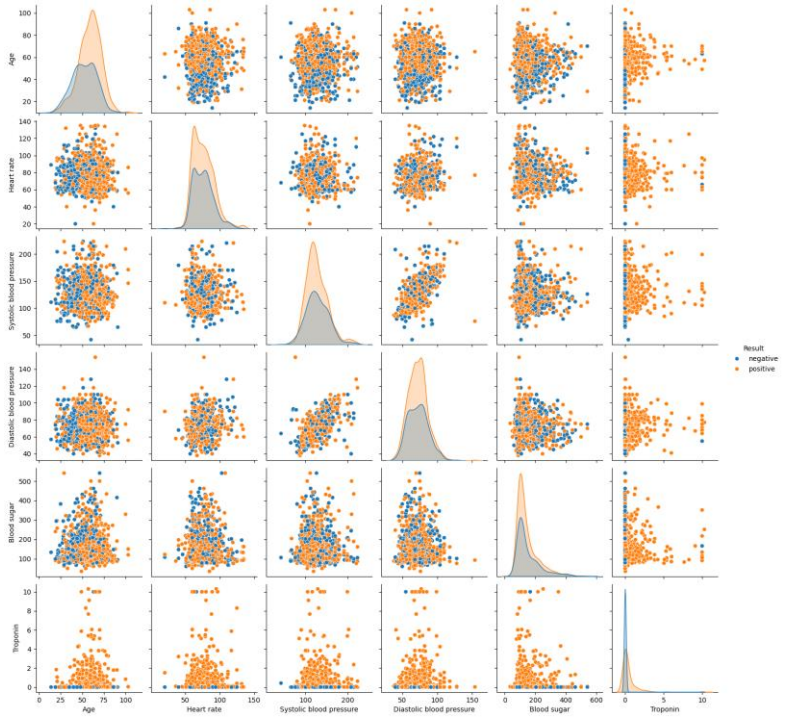


Figura 8. Matriz Diagrama de dispersión de variables biológicas basado en resultados positivos (naranja) y negativos (azul).

Se exploraron estadísticas comparativas para resultados positivos y negativos tanto para CK-MB como para troponina.

Tabla 2. Estadísticas comparativas de troponina y CK-MB para resultados positivos (con infarto) y negativos (sin infarto).

	Positive Results (n = 810)		**Negative Results (n = 509)**	
	Troponin	**CK-MB**	**Troponin**	**CK-MB**
Count	810	810	509	509
Mean	0.571	23.3	0.027	2.55
Standard Deviation	1.39	57.7	0.443	1.37
Minimum	0.003	0.353	0.001	0.321
25th Percentile (Q1)	0.016	1.87	0.003	1.5
Median (Q2)	0.044	3.78	0.006	2.31
75th Percentile (Q3)	0.456	12.2	0.009	3.35
Maximum	10.3	300	10	7.02

Los componentes principales (CP) capturan la varianza máxima del conjunto de datos y combinan los efectos de cada variable calculando su ponderación. Al crear los gráficos de componentes principales, los datos se capturan en dos dimensiones:

- PC1 (x-axis) = capta la influencia combinada de las variables responsables de la mayor variación.
- PC2 (y-axis) = capta la segunda mayor variación.

$$\text{PC1} = a_1 \cdot \text{Age} + a_2 \cdot \text{Heart Rate} + a_3 \cdot \text{Troponin} + \ldots$$

$$\text{PC2} = b_1 \cdot \text{Age} + b_2 \cdot \text{Heart Rate} + b_3 \cdot \text{Troponin} + \ldots$$

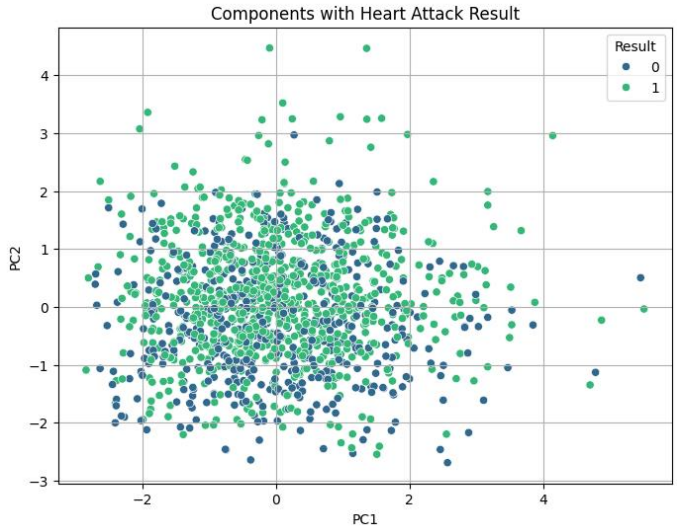


Figura 9. Observe la agrupación entre los resultados positivos y negativos de ataques cardíacos según dos componentes principales.

Cuando las variables no presentan una alta correlación, el uso de k-medias permite aprovechar con mayor eficacia las características únicas de cada variable para segmentar a los pacientes en conglomerados (González-Franco et al., 2025). Se aplicó el método del codo para determinar las k-medias óptimas. Posteriormente, se agruparon los datos.

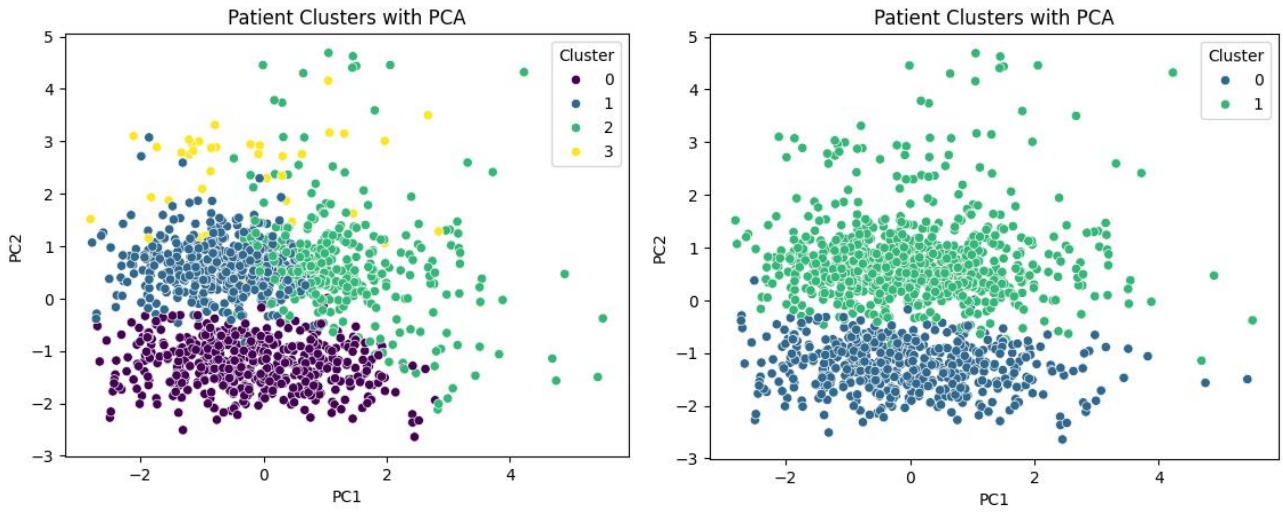


Figura 10. Agrupamiento de n_componentes=4 (izquierda) y n_componentes=2 (derecha) de las variables biológicas de un individuo en función de dos componentes principales.

Tabla 3 muestra el promedio de cada característica por clúster. De esta manera, se comprende la diferenciación de cada clúster según las variables individuales. Caracterización del género como masculino: 1 y femenino: 0, y resultado como positivo: 1 y negativo: 0.

Tabla 3. Características promedio por clúster cuando los componentes son n=4 y n=2

n_components=4

Cluster	Age	Gender	Heart rate (lpm)	Systolic blood pressure (mmHG)	Diastolic blood pressure (mmHg)	Blood sugar (mg/dL)	CK-MB (ng/mL)	Troponin (ng/mL)	Result
0	51.7	0.59	75.0	125.8	71.2	149.8	2.55	0.01	0.00
1	58.9	0.65	71.8	114.0	64.6	138.9	12.5	0.30	0.99
2	58.8	0.76	83.6	147.5	84.6	149.8	9.49	0.96	0.93
3	56.4	0.71	75.2	127.1	71.8	174.2	261.9	0.27	1.00

n_components=2

Cluster	Age	Gender	Heart rate (lpm)	Systolic blood pressure (mmHG)	Diastolic blood pressure (mmHg)	Blood sugar (mg/dL)	CK-MB (ng/mL)	Troponin (ng/mL)	Result
0	52.1	0.6	75.9	127.8	72.5	149.7	2.56	0.01	0.0
1	58.8	0.7	76.1	126.8	72.1	144.7	23.2	0.58	1.0

Se calcularon el logaritmo de Troponina y CK-MB para la regresión logística. Estos dos biomarcadores estaban sesgados a la derecha y necesitaban ser linealizados para mejorar el rendimiento del modelo. Posteriormente se realizó la regresión logística presentados en Figura 11.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                 Result   No. Observations:                 1319
Model:                          Logit   Df Residuals:                     1310
Method:                           MLE   Df Model:                            8
Date:                Sun, 29 Jun 2025   Pseudo R-squ.:                 0.4418
Time:                        14:17:24   Log-Likelihood:                -491.01
converged:                       True   LL-Null:                       -879.61
Covariance Type:            nonrobust   LLR p-value:                1.691e-162
============================================================================================
                               coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                        -5.4376      0.771     -7.057      0.000      -6.948      -3.927
Age                           0.0394      0.006      6.331      0.000       0.027       0.052
Gender                        0.2548      0.165      1.541      0.123      -0.069       0.579
Heart rate                   -0.0003      0.005     -0.050      0.960      -0.010       0.010
Systolic blood pressure      -0.0031      0.004     -0.826      0.409      -0.011       0.004
Diastolic blood pressure      0.0037      0.007      0.528      0.598      -0.010       0.017
Blood sugar                  -0.0008      0.001     -0.765      0.445      -0.003       0.001
CK-MB_log                     1.8445      0.151     12.205      0.000       1.548       2.141
Troponin_log                 33.5507      4.111      8.162      0.000      25.494      41.607
============================================================================================
```

Figura 11. Modelo de regresión logística.

A continuación, se realizó un análisis de regresión logística para examinar la asociación entre varias variables clínicas y demográficas y la probabilidad del resultado. Los resultados se presentan en términos de razones de momios (OR) y los correspondientes valores *p* en la Tabla 4.

Tabla 4. Modelo de regresión logística

Variable	Odds Ratio (OR)	P-value
Age	1.04	<0.001
Gender	1.29	0.12
Heart rate	1.00	0.96
Systolic blood pressure	1.00	0.41
Diastolic blood pressure	1.00	0.60
Blood sugar	1.00	0.44
CK-MB (log-transformed)	6.32	<0.001
Troponin (log-transformed)	372000000000000	<0.001

4.3 Objetivo 3. Obtención de resultados

4.3.1 Preprocesamiento y Análisis

Se realizó un análisis y preprocesamiento inicial de datos con miras al futuro entrenamiento de un modelo de Machine Learning, específicamente un Random Forest. Para ello, se utilizó un Jupyter Notebook, lo que facilitó la evaluación inicial del dataset. Este proceso incluyó la observación de valores faltantes, la forma del dataset y la identificación de posibles valores atípicos (outliers).

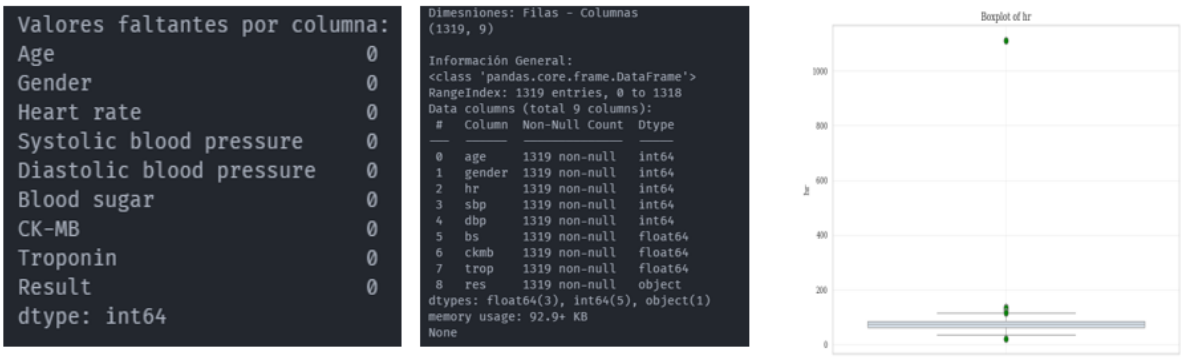


Figura 12. Listado izquierda: Valores faltantes en los registros; Listado derecha: Información general del dataset; Gráfica: Outlier en el campo Heart Rate de un registro.

Se renombraron las variables para hacerlas más consistentes, eliminando espacios en blanco y estableciendo su nomenclatura en minúsculas. Además, se convirtió el campo 'resultados' de texto a valores numéricos. Finalmente, se exportó una copia del dataset procesado al directorio 'data/processed/'.

Figura 13. Izquierda: Renombrado de variables; Derecha arriba: Conversion a valores numéricos de los registros en el campo resultados; Derecha abajo: Guardado del dataset procesado

4.3.2 Entrenamiento y Evaluación del Modelo

El proceso comenzó con la separación del dataset original en cuatro archivos diferentes: 'X_train', 'X_test', 'y_train' y 'y_test', los cuales fueron guardados como archivos .csv separados (ver preprocess.py). A continuación, se inició el entrenamiento del modelo utilizando sklearn con los siguientes parámetros: n_estimators=150, max_depth=10, max_features='sqrt', class_weight="balanced", random_state=42 y oob_score=True. Finalmente, el modelo entrenado fue exportado y guardado en el directorio 'models/'.

```python
model = RandomForestClassifier(
    n_estimators=150,
    max_depth=10,
    max_features= 'sqrt',
    class_weight="balanced",
    random_state=42,
    oob_score=True,
)
model.fit(X_train, y_train)
```

```python
joblib.dump(
    model,
    os.path.join(
        project_dir,
        "models/rf_model.joblib"
        )
        )
```

Figura 14. Izquierda: parámetros del modelo; Derecha: guardado del modelo entrenado

Posteriormente, se realizó la evaluación del modelo arrojando los siguientes resultados almacenados en el directorio ‘reports/’.

	precision	recall	f1-score	support
0	1.00	0.98	0.99	102
1	0.99	1.00	0.99	162
accuracy			0.99	264
macro avg	0.99	0.99	0.99	264
weighted avg	0.99	0.99	0.99	264

```
Feature Importances
age: 0.0475
gender: 0.0090
hr: 0.0180
sbp: 0.0185
dbp: 0.0185
bs: 0.0254
ckmb: 0.2689
trop: 0.5943
```

Figura 15. Resultados de la evaluación del modelo

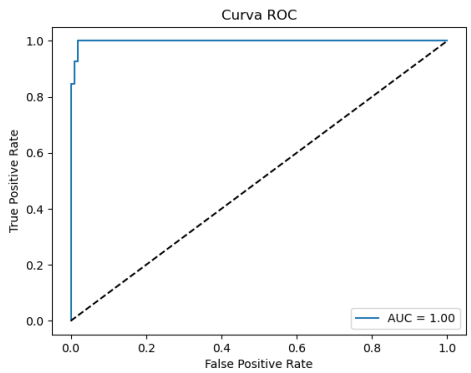


Figura 16. Grafica de evaluación del modelo (Curva ROC)

4.3.3 Visualización Web

Se desarrolló una aplicación web utilizando Streamlit, lo que permite una interactividad fluida con el usuario. La aplicación consta de tres secciones principales:

1. **Formulario Interactivo**: En la primera sección, se creó un formulario que permite a los usuarios introducir datos de manera interactiva. Este formulario utiliza el modelo previamente entrenado para predecir la probabilidad de Infarto Agudo del Miocardio.
2. **Estadísticas y Evaluaciones**: La segunda sección está dedicada a mostrar las estadísticas y evaluaciones del funcionamiento interno del modelo. Aquí, los usuarios pueden ver métricas de rendimiento y otras evaluaciones relevantes que demuestran la eficacia del modelo.
3. **Visualización de Árboles**: En la tercera sección, se muestran gráficamente los primeros nueve árboles generados por el modelo. Esta visualización ayuda a entender cómo el modelo está tomando decisiones basadas en los datos de entrada.
4. **Créditos**: En esta última sección se encuentran los créditos con los nombres de los autores y un link al repositorio del proyecto en GitHub.

Esta estructura de la aplicación web facilita tanto la interacción con el modelo como la comprensión de su funcionamiento y rendimiento. Link de App Web es encontrado en *8. ANNEXO*.

Figura 17. Sección 1 de la app web - predicción a partir de los datos de un paciente haciendo uso del modelo.

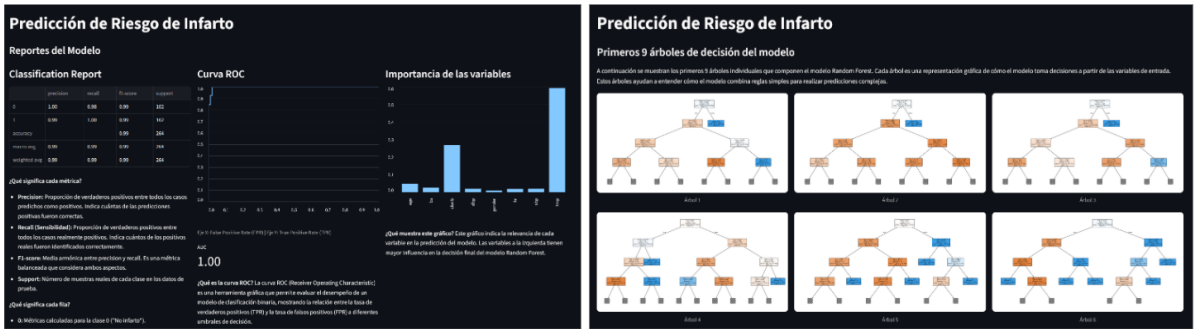


Figura 18. Izquierda: Sección 2 de la aplicación web - información interactiva de evaluación del modelo; Derecha: Sección 2 de la aplicación web - gráfica de los primeros 9 árboles de decisiones.

5. DISCUSIÓN DE RESULTADOS

5.1 Objetivo 1

Respecto a los tres análisis exploratorios realizados, en todos los análisis la dimensión de los datos es de 1319 muestras por 9 variables. De éstas 9 variables siete son cuantitativas y dos son categóricas: el género de los pacientes y el tipo de resultado de ataque al corazón. De ésta última variable en concreto tan solo hay dos categorías: positiva y negativa; y en cuanto a la variable de género también hay dos categorías: masculino y femenino. No se ha encontrado algún valor perdido (“n_missing”). En el primer análisis no hay ninguna variable de agrupamiento, pero en el segundo sí aparece uno: tipo de resultado. Y en el tercer análisis aparecen dos variables de agrupamiento: tipo de resultado y género.

En el primer análisis, Figura 1, se puede ver que la media de la presión sistólica es mayor que la presión diastólica, pero la desviación estándar de ésta es menor. La media de la frecuencia cardíaca es casi de 80 lpm pero con una desviación de casi 50 lpm. La edad media es de aproximadamente 56 años con aproximadamente 14 años de desviación estándar. En cuanto a las variables bioquímicas, la variable Troponina tiene mucha menos desviación estándar que las variables azúcar en sangre y CK-mb.En el segundo análisis, Figura 2, en cuanto a la edad la media en resultados positivos es algo mayor que los de resultados negativos con una desviación estándar muy similar. En relación con las variables fisiológicas como la frecuencia cardíaca, presión sistólica y presión diastólica sus medias estadísticas y desviaciones estándares son muy parecidas en ambos casos de resultado. Y en relación con las variables bioquímicas, la media y desviación estándar de la variable azúcar en sangre en los dos casos de resultado son muy similares. Pero en las variables Troponina y CK-mb sus medias y desviaciones estándares son muy distintas en los dos

casos de resultados. En el tercer análisis, Figura 4, ahora los datos agrupados por las variables género y tipo de resultado, se observa que las medias y desviaciones estándares de cinco variables cuantitativas son aproximadamente similares en los dos casos de género y de misma categoría de resultado a ataque al corazón. En contraste, en la variable Troponina las medias de las agrupaciones de género masculino y femenino son diferentes tanto en casos de resultado positivo como negativo. Pero en respecto a sus desviaciones estándares, solo en la de agrupación de resultado negativo la desviación del grupo masculino es diferente al del grupo femenino. También en la variable frecuencia cardíaca, las desviaciones estándares de la agrupación de resultado negativo son diferentes. En los barplots, Figura 3, se puede ver que abundan más contajes de pacientes con género masculino y contajes de resultado con caso positivo a ataques al corazón.

En referencia a las comprobaciones de distribución de normalidad, Figura 5, las variables "Blood Sugar", "CK-MB" y “Troponin” no siguen una distribución de normalidad como se observa tanto en los histogramas y gráficos Q-Q como en las pruebas estadísticas. Es recomendable tener esto en cuenta al diseñar un modelo de predicción de infartos y seleccionar métodos estadísticos apropiados. A primera vista, todas las pruebas dicen que ninguna variable continua sigue una distribución normal, pero se debe tener cuidado en no malinterpretar este resultado dado al tamaño del conjunto de datos o dataset. Ha sido importante usar la visualización para comprobar si la desviación de normalidad es relevante para el análisis (Gomes, 2025). Por tanto, ninguna de las pruebas estadísticas, *Shaphiro-Wilk* y *D’Agostino-Pearson*, indican que haya alguna variable que siga una distribución de normalidad (T-Test: p-value = 0.00001), pero visualmente se aproxima y será usada en modelos que no requieren normalidad estricta. Variables como la enzima CK-mb, el biomarcador Troponina, y azúcar en sangre claramente siguen una distribución no normal, y deben transformarse o usar modelos no paramétricos. En cambio, variables como edad, frecuencia cardíaca y los dos tipos de presión arterial tienen desviaciones muy pequeñas, probablemente no problemáticas para modelos paramétricos.

Respecto a las comparaciones de medias, medianas y desviaciones estándares de los niveles de las tres variables bioquímicas tanto en grupos de caso de resultado positivo como negativo, como se ve en la Figura 6, se observa que en la variable azúcar en sangre las dos medias y medianas no son significativamente diferentes (T-Test: p-value > 0.05), pero en la variable CK-mb las dos medias y medianas son significativamente diferentes (T-Test: p-value < 0.05). En la variable Troponina las dos medias y medianas aparecen con valores negativos debido a que la mayoría de valores de esta variable son decimales y, al ser tratados con

cálculo logarítmico, pues resultan de esta manera. De todos modos, se observa que sus medias y medianas son significativamente diferentes (T-Test: p-value < 0.05).

5.2 Objetivo 2

La matriz de correlación se utilizó para detectar posibles problemas de multicolinealidad (Schober et al., 2018). Solo la presión arterial sistólica y diastólica mostraron una correlación positiva moderada (r = 0.59), lo cual concuerda con su relación fisiológica esperada, ya que ambas se ven influenciadas por factores como la rigidez arterial y la resistencia vascular (Gonzalez-Franco et al., 2025). El resto de las variables mostraron correlaciones muy bajas o insignificantes, con coeficientes cercanos a cero, lo que sugiere relaciones lineales débiles o inexistentes entre ellas.

La edad, la troponina y la CK-MB presentaron correlaciones estadísticamente significativas (p-value < 0.05) con los eventos de infarto de miocardio (Tabla 1). En cambio, la presión arterial (sistólica y diastólica), la frecuencia cardíaca y la glucemia no mostraron una asociación significativa con dichos eventos.

La Figura 8 proporciona una representación visual de los patrones de las variables en función del resultado de infarto. Se observa una distribución aproximadamente normal para cada variable entre los grupos con y sin infarto, y una clara separación en los niveles de troponina entre ambos grupos. En contraste, otras variables no evidenciaron diferencias visuales claras, por lo que se requiere un análisis estadístico más profundo para evaluar su relevancia.

El análisis estadístico comparativo reveló que los niveles de troponina fueron significativamente más altos en los casos positivos de infarto (media = 0.571) en comparación con los negativos (media = 0.027), con una mayor variabilidad (DE = 1.39 vs. 0.443). Además, el percentil 75 en los casos positivos (0.456) superó ampliamente incluso el valor máximo observado en los casos negativos (0.009), lo que refuerza el papel de la troponina como marcador sensible de daño miocárdico (Tabla 2). De forma similar, los niveles de CK-MB también fueron considerablemente más altos en los casos positivos, con un percentil 75 de 23.2 frente a 2.56 en los negativos. La desviación estándar fue notablemente mayor en los casos positivos (DE = 57.7), lo cual sugiere una mayor heterogeneidad en la magnitud de la lesión miocárdica. En contraste, los niveles de CK-MB fueron más consistentes en los casos negativos (DE = 1.37), lo que sugiere una expresión estable en ausencia de infarto. Estos hallazgos respaldan el uso de la troponina y la CK-MB

como biomarcadores útiles para diferenciar entre pacientes con y sin infarto de miocardio, siendo la troponina especialmente destacada por su sensibilidad.

Dado que las variables de interés (por ejemplo, troponina, CK-MB, frecuencia cardíaca) son numéricas y continuas, el método del codo aplicado al algoritmo de K-medias fue apropiado para determinar el número óptimo de agrupamientos (k). Este análisis se visualizó en la Figura 10, donde cada punto representa un paciente agrupado según sus características biológicas. La Tabla 3 muestra los promedios de cada variable por agrupamiento. Se observó que la frecuencia cardíaca, la presión arterial (sistólica y diastólica) y los niveles de glucosa en sangre fueron similares entre los clúster, lo que sugiere que estas variables no fueron determinantes en la diferenciación de los perfiles de riesgo. En cambio, la troponina y la CK-MB sí mostraron diferencias notables entre clúster, lo que indica su relevancia clínica como marcadores discriminativos de infarto.

Estos resultados coinciden con la literatura, que resalta la utilidad diagnóstica y pronóstica de la troponina y la CK-MB en los síndromes coronarios agudos. Estudios previos han demostrado que incluso elevaciones moderadas de troponina se asocian con un mayor riesgo de mortalidad total y cardiovascular (Fan et al., 2018; Pavasini et al., 2015). La coelevación de troponina y CK-MB observada en los grupos con infarto sugiere una disfunción cardíaca más severa y una mayor desregulación metabólica (Wu et al., 1999).

La edad media de los pacientes fue de 56.2 años (DE ± 13.6). La literatura sugiere que aunque los infartos son menos frecuentes en pacientes jóvenes, cuando ocurren, los niveles elevados de troponina suelen deberse a infarto de miocardio en un porcentaje considerable de casos (58.8% en menores de 50 años), según (Wu et al., 2018).

En la Figura 10, se analiza la distribución de CK-MB y troponina considerando tanto 4 como 2 agrupamientos. Con cuatro clústers, ningún paciente del clúster 0 presentó infarto, mientras que la mayoría en los clústers 1, 2 y 3 sí lo hicieron (99%, 93% y 100%, respectivamente). Al dividir en dos clústers, el clúster 0 (sin infarto) presentó niveles promedio de CK-MB de 2.56 ng/mL y de troponina de 0.01 ng/mL, mientras que el clúster 1 (con infarto) presentó niveles promedio de CK-MB de 23.2 ng/mL y de troponina de 0.58 ng/mL.

En el modelo de regresión logística (Tabla 4.), la edad, la CK-MB (log) y la troponina (log) fueron predictores estadísticamente significativos del infarto de miocardio. Un incremento de un año en la edad se asoció con un aumento del 4% en la probabilidad del evento (OR = 1.04, $p < 0.001$). De manera similar, un aumento de una unidad logarítmica en CK-MB se

asoció con un aumento de más de seis veces en la probabilidad (OR = 6.32, $p < 0.001$). En el caso de la troponina logarítmica mostró una razón de momios extremadamente alta (OR $\approx 3.72 \times 10^{14}$, $p < 0.001$), indicando una fuerte asociación. La razón de probabilidades extremadamente alta para la troponina probablemente refleja su fuerte relevancia clínica como un biomarcador altamente específico y sensible para la lesión miocárdica. Los niveles de troponina permanecen indetectables en individuos sanos, pero aumentan drásticamente tras el daño al músculo cardíaco, creando un claro contraste binario entre los casos y los no casos (Antman et al., 2000; Apple & Collinson, 2012). Esta diferencia dramática mejora su poder predictivo en la regresión logística. Además, incluso pequeños aumentos en la troponina están fuertemente asociados con eventos cardíacos adversos, lo que puede explicar la razón de probabilidades amplificada a pesar de la transformación logarítmica (Giannitsis & Katus, 2013). No obstante, los OR extremadamente altos deben interpretarse con precaución debido a las posibles influencias de la distribución de los datos, el momento de la medición y la variabilidad clínica.

Las demás variables —género (OR = 1.29, $p = 0.12$), frecuencia cardíaca (OR = 1.00, $p = 0.96$), presión arterial sistólica (OR = 1.00, $p = 0.41$), presión arterial diastólica (OR = 1.00, $p = 0.60$) y glucemia (OR = 1.00, $p = 0.44$)— no fueron estadísticamente significativas.

En resumen, este análisis identificó a la edad, la CK-MB y la troponina como predictores significativos del infarto de miocardio. Estos resultados son consistentes con la literatura existente que respalda el valor pronóstico de estos biomarcadores en la evaluación del riesgo clínico (Antman et al., 2000; Wu et al., 1999). La asociación positiva entre la edad y el infarto concuerda con el riesgo creciente de eventos cardiovasculares asociado al envejecimiento (D'Agostino et al., 2008).

En general, los hallazgos subrayan la utilidad de la edad y los biomarcadores cardíacos en la predicción de resultados, al mismo tiempo que enfatizan la necesidad de un manejo cuidadoso de las transformaciones de variables y los diagnósticos de modelos en la modelización de regresión (Hosmer et al., 2013).

5.3 Objetivo 3

El algoritmo Random Forest para el entrenamiento de un modelo de Machine Learning es catalogado como uno de los más poderosos. En el caso de este trabajo, la elección se debió a su capacidad de prevención de overfitting, manejo de rangos amplios, robustez a valores atípicos y la no necesidad de normalización de los datos. Tanto las troponinas como CK-MB son biomarcadores de gran importancia para el diagnóstico de Infarto Agudo del

Miocardio. Sin embargo, estos biomarcadores tienen un rango de valores normales pequeño, mientras que, cuando se elevan, el rango puede ser muy amplio. Por lo tanto, este modelo se hizo idóneo para este caso.

Inicialmente, el dataset estaba exento de valores faltantes, lo que, aunque no supone un reto para el algoritmo elegido (capaz de manejar estos registros), denota que la confección del mismo se realizó meticulosamente. Las variables presentaban una nomenclatura que podría dificultar un trabajo ágil con los datos, por lo que fueron renombradas para facilitar su manipulación. Durante el análisis, se encontró un valor atípico en el campo de 'heart rate', específicamente un valor de 1111, que posiblemente fue un error humano al introducirlo. Este registro fue eliminado, ya que no afectaría significativamente el trabajo futuro.

A la hora de entrenar el modelo se decide establecer los hiperparámetros de la forma siguiente.

- n_estimators=150: Esta cantidad de árboles es suficiente para el tamaño de la muestra que se manejrá.
- max_depth=10: al limitar la profundidad de los árboles se previene el sobre entrenamiento.
- max_features='sqrt': ayuda con la diversidad de los árboles y reduce la varianza.
- class_weight="balanced": importante en este caso para ajustar el desbalance entre casos positivos y negativos.
- random_state=42: al establecer una semilla fija ayuda a la reproducibilidad de los resultados.
- oob_score=True: Importante para la estimación del error del modelo al seleccionar las muestras no seleccionadas en la construcción de cada árbol.

Los valores obtenidos tras la evaluación reflejan una alta eficiencia, rendimiento y robustez del modelo. Los parámetros de precisión, recall y F1-score son superiores a 0.98 para ambas clases a clasificar, lo cual indica un equilibrio, una buena generalización y una alta fiabilidad. El área bajo la curva (AUC) en la gráfica de la característica de operación del receptor (ROC) es igual a 1.0, lo que denota la perfección del modelo, considerándose completamente superior al azar. Al observar los primeros nueve árboles creados, se destaca la variedad de los mismos y su relación con todo lo mencionado previamente.

La evaluación de la importancia de las variables muestra que las troponinas y la CK-MB son las que más influyen en la toma de decisiones del modelo. Esto no es sorprendente, dado que la evidencia científica respalda la relevancia de estos biomarcadores en la fisiopatología de tales eventos cardíacos.

El desarrollo de la aplicación web con Streamlit, un paquete de Python, permite una interacción dinámica y fluida con el modelo generado. Al desplegar la aplicación en Streamlit Cloud, se ofrece acceso gratuito a cualquier persona con un dispositivo inteligente, facilitando así su uso generalizado. Además, el código fuente de la aplicación web, así como el de todo el proyecto, está disponible públicamente. Esto permite que cualquier persona interesada pueda acceder, revisar, utilizar y contribuir al código, fomentando la transparencia y la colaboración.

6. CONCLUSIONES

6.1 Objetivos 1

- Los pacientes con resultado positivo tienden a ser ligeramente mayores.
- En general, las medias y desviaciones de las variables cuantitativas son similares entre hombres y mujeres para un mismo resultado.
- Las medias y desviaciones estándares en las variables fisiológicas (frecuencia cardíaca, presión sistólica y diastólica) no presentan grandes diferencias entre los grupos de resultado positivo y negativo.
- Las variables azúcar en sangre, CK-mb y Troponina no siguen una distribución de normalidad, pero las variables edad, frecuencia cardíaca, presión arterial sistólica y presión arterial diastólica sí.
- Las medias y medianas de las variables bioquímicas (azúcar en sangre, CK-mb y Troponina) entre los grupos de resultado positivo y negativo, en la variable azúcar en sangre éstas no son significativamente diferentes pero las de las variables CK-mb y Troponina sí son.

6.2 Objetivos 2

- El clústers y modelo de regresión logística identifica la troponina, CK-MB y la edad como predictores significativos de infarto de miocardio, con la troponina mostrando una asociación extremadamente alta que probablemente refleja su fuerte relevancia clínica, pero también sugiere posibles artefactos de datos o de modelado.
- Otras variables como el género, la frecuencia cardíaca, la presión arterial y el nivel de azúcar en sangre no mostraron valores predictivos significativos.

6.3 Objetivos 3

- El dataset elegido estaba en óptimas condiciones para el entrenamiento de un modelo utilizando el algoritmo Random Forest.
- Los resultados obtenidos tras la evaluación muestran que el modelo es robusto, eficiente y con un alto rendimiento.
- Se confirma la idea inicial que tanto las troponinas como la CK-MB son las variables que más influirían en la toma de decisiones del modelo, lo cual está en concordancia con la evidencia científica sobre esta patología.
- La elección de este algoritmo de clasificación fue acertada debido a las ventajas que ofrece para la muestra de datos analizada.
- Finalmente, la creación de la aplicación web permitirá una interacción visual y cómoda con el modelo, facilitando su uso y accesibilidad.

7. BIBLIOGRAFÍA

7.1 Artículos académicos

Antman, E. M., Tanasijevic, M. J., Thompson, B., Schactman, M., McCabe, C. H., Cannon, C. P., & Braunwald, E. (2000). Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. *New England Journal of Medicine, 335*(18), 1342–1349. https://doi.org/10.1056/NEJM200010263431802

Apple**,** F. S. (1999). Tissue specificity of cardiac troponin I, cardiac troponin T and creatine kinase-MB. *Clinica Chimica Acta, 284*(2), 151–159.

Apple, F. S., & Collinson, P. O. (2012). Analytical characteristics of high-sensitivity cardiac troponin assays. *Clinical Chemistry, 58*(1), 54–61. https://doi.org/10.1373/clinchem.2011.165795

D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation, 117*(6), 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579

Fan, Y., Jiang, M., Gong, D., Man, C., & Chen, Y. (2018). Cardiac troponin for predicting all-cause mortality in patients with acute ischemic stroke: a meta-analysis. *Bioscience Reports, 38*(2), BSR20171178.

Flint, A. C., et al. (2019). Effect of systolic and diastolic blood pressure on cardiovascular outcomes. *New England Journal of Medicine, 381*(3), 243–251.

Giannitsis, E., & Katus, H. A. (2013). Cardiac troponin level elevations not related to acute coronary syndromes. *Nature Reviews Cardiology, 10*(11), 623–634. https://doi.org/10.1038/nrcardio.2013.161

Gonzalez-Franco, J. D., Galaviz-Mosqueda, A., Villarreal-Reyes, S., Lozano-Rizk, J. E., Rivera-Rodriguez, R., Gonzalez-Trejo, J. E., ... & Ibarra-Flores, E. A. (2025). Revolutionizing cardiac risk assessment: AI-powered patient segmentation using advanced machine learning techniques. *Machine Learning and Knowledge Extraction, 7*(2), 46.

Groenewegen, A., Rutten, F. H., Mosterd, A., & Hoes, A. W. (2020). Epidemiology of heart failure. *European Journal of Heart Failure*, *22*(8), 1342–1356. https://doi.org/10.1002/ejhf.1858

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., . . . Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Khera, R., Lu, Y., Wang, K., Gupta, A. K., Ayers, C. R., DeFilippis, A. P., ... & Virani, S. S. (2021). Contemporary evaluation of blood pressure thresholds and cardiovascular risk. *Circulation, 143*(14), 1364–1374.

Nielson, C., & Lange, T. (2005). Blood glucose and heart failure in nondiabetic patients. *Diabetes Care, 28*(3), 607–611.

Pavasini, R., d'Ascenzo, F., Campo, G., Biscaglia, S., Ferri, A., Contoli, M., ... & Ferrari, R. (2015). Cardiac troponin elevation predicts all-cause mortality in patients with acute

exacerbation of chronic obstructive pulmonary disease: systematic review and meta-analysis. *International Journal of Cardiology, 191*, 187–193.

Rajendra Acharya, U., et al. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing, 44*, 1031–1051.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763–1768.

Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., & White, H. D. (2018). Fourth universal definition of myocardial infarction (2018). *Journal of the American College of Cardiology, 72*(18), 2231–2264.

Wu, A. H. B., Apple, F. S., Gibler, W. B., Jesse, R. L., Warshaw, M. M., & Valdes, R. (1999). National Academy of Clinical Biochemistry standards of laboratory practice: Recommendations for the use of cardiac markers in coronary artery diseases. *Clinical Chemistry, 45*(7), 1104–1121. https://doi.org/10.1093/clinchem/45.7.1104

Wu, A. H., Christenson, R. H., Greene, D. N., Jaffe, A. S., Kavsak, P. A., Ordonez-Llanos, J., & Apple, F. S. (2018). Clinical laboratory practice recommendations for the use of cardiac troponin in acute coronary syndrome. *Clinical Chemistry, 64*(4), 645–655.

Wu, A. H., Panteghini, M., Apple, F. S., Christenson, R. H., Dati, F., & Mair, J. (1999). Biochemical markers of cardiac damage: From traditional enzymes to cardiac-specific proteins. *Scandinavian Journal of Clinical and Laboratory Investigation, 59*(sup230), 74–82.

Wu, C., Singh, A., Collins, B., Fatima, A., Qamar, A., Gupta, A., ... & Blankstein, R. (2018). Causes of troponin elevation and associated mortality in young patients. *The American Journal of Medicine, 131*(3), 284–292.

Yilmaz, A., Yalta, K., Turgut, O. O., Yilmaz, M. B., Ozyol, A., Kendirlioglu, O., ... & Tandogan, I. (2006). Clinical importance of elevated CK-MB and troponin I levels in congestive heart failure. *Advances in Therapy, 23*, 1060–1067.

7.2 Páginas web

American Diabetes Association. (2023). 10. Cardiovascular Disease and Risk Management: Standards of Care in Diabetes—2023. *Diabetes Care* , 46(Supplement_1), S137–S152. https://doi.org/10.2337/dc23-S010

Gomes, G. G. (2025, January 18). *Descriptive Statistics: Expectations vs. Reality (Exploratory Data Analysis – EDA)*. Towards Data Science. https://towardsdatascience.com/descriptive-statistics-expectations-vs-reality-exploratory-data-analysis-eda-8336b1d0c60b/

Heart attack dataset. (n.d.). https://www.kaggle.com/datasets/fatemehmohammadinia/heart-attack-dataset-tarik-a-rashid/data

Bakker, Jonathan D. "Applied Multivariate Statistics in R." *University of Washington*, 2024. https://uw.pressbooks.pub/appliedmultivariatestatistics/

Magwene, P. M. (2025, June 25). *Biology 304: Biological Data Analysis*. https://bio304-class.github.io/bio304-book/index.html

Rashid, T. A., & Hassan, B. (2022). Heart attack dataset. *Mendeley Data*. https://doi.org/10.17632/wmhctcrt5v.1

8. ANEXO

Link de Github de proyecto: https://github.com/IL3-DS-Trabajo-Final-Heart-Attack/IL3_DS_Trabajo_Final_Heart_Attack

Link de App Web: https://heart-attack-ds-rforest.streamlit.app/