



PREDICCIÓN DE ATAQUES AL CORAZÓN

Proyecto Final
Equipo 1




TABLA DE CONTENIDOS

01	PROBLEMÁTICA
02	ANÁLIS DE PAPER
03	PROCESAMIENTO DE DATOS
04	MODELOS
05	RESULTADOS

INTRODUCCIÓN

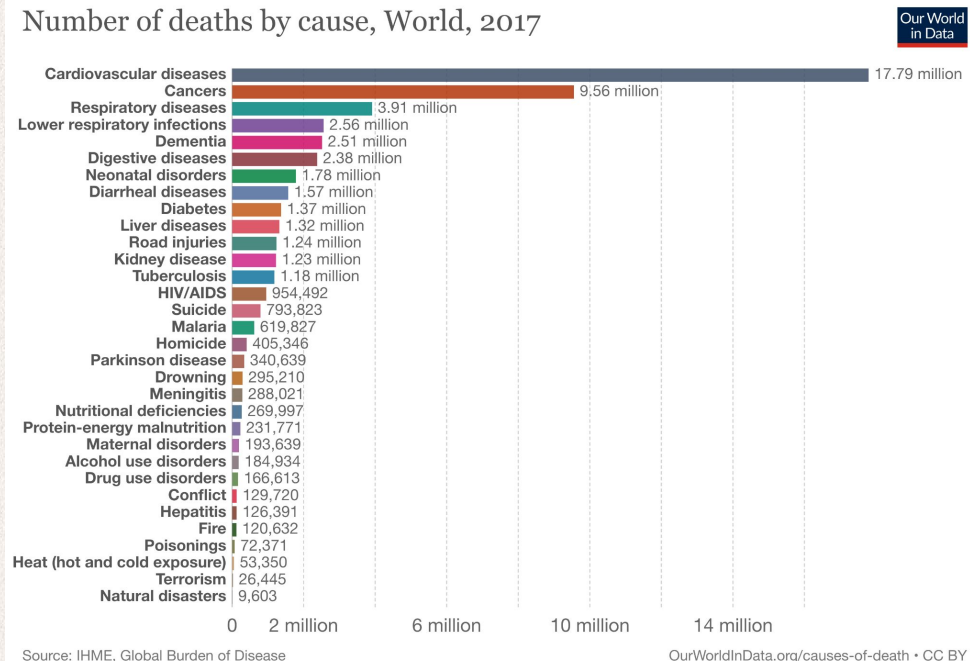
Las enfermedades cardiovasculares son una de las principales causas de muerte en todo el mundo y una amenaza para la salud pública que ha alcanzado proporciones epidémicas.

ENFERMEDADES CARDIOVASCULARES

Las enfermedades cardiovasculares son una de las principales causas de muerte y una amenaza para la salud pública.

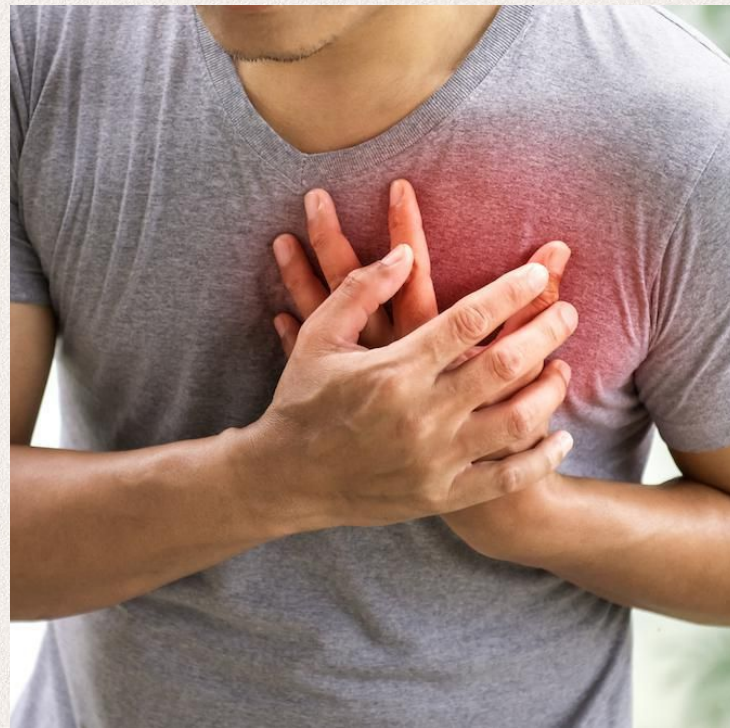
En México, han sido una de las principales causas de muerte desde 1990, con una tasa de mortalidad de 196 muertes por cada 100.000 de ese año.

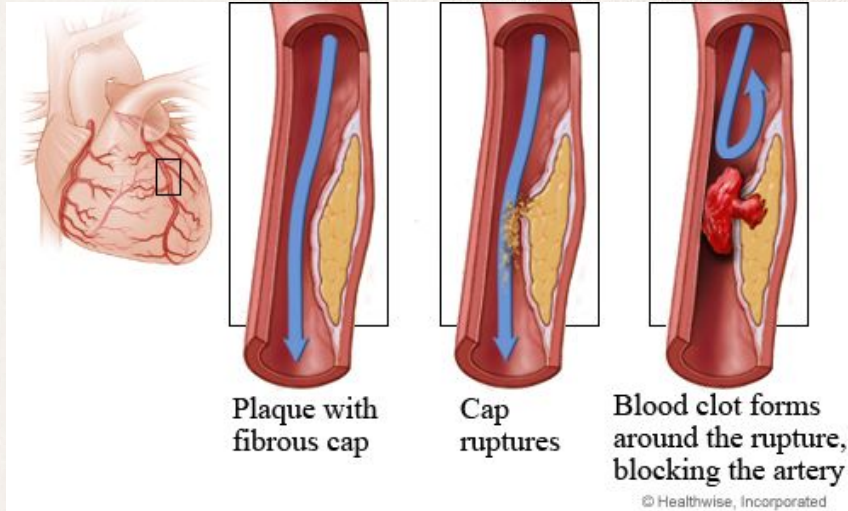
Number of deaths by cause, World, 2017



¿QUÉ ES UN ATAQUE AL CORAZÓN?

El músculo cardíaco necesita oxígeno para sobrevivir. Un ataque cardíaco ocurre cuando el flujo sanguíneo que lleva oxígeno al músculo cardíaco se reduce severamente o se corta por completo.





CÓMO OCURRE UN ATAQUE CARDÍACO

Esto sucede porque las arterias coronarias que suministran flujo sanguíneo al músculo cardíaco pueden estrecharse debido a la acumulación de grasa, colesterol y otras sustancias que en conjunto se denominan placa. A este lento proceso se conoce como aterosclerosis.



Análisis de Paper

Research Article | Open Access

Volume 2021 | Article ID 8387680 | <https://doi.org/10.1155/2021/8387680>

[Show citation](#)

Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning

Rohit Bharti ¹, Aditya Khamparia ², **Mohammad Shabaz**  ³, Gaurav Dhiman ⁴,
Sagar Pande ¹ and Parneet Singh ⁵

[Show more](#)

Academic Editor: Ahmed A. Abd El-Latif

Received
16 May 2021

Revised
15 Jun 2021

Accepted
21 Jun 2021

Published
01 Jul 2021

Set de Datos

El conjunto de datos consta de 14 atributos principales que se utilizan para realizar el análisis.

1. Edad: edad del paciente en años
2. Sexo: 1 = hombre; 0 = mujer
3. Cp: tipo de dolor en el pecho.
4. Trestbps: presión arterial en reposo.
5. Col: el colesterol sérico muestra la cantidad de triglicéridos presentes.
6. Fbs: azúcar en sangre en ayunas superior a 120 mg / dl (1 verdadero). Menos de 100 mg / dL (5,6 mmol / L) es normal y de 100 a 125 mg / dL (5,6 a 6,9 mmol / L) se considera prediabetes.
7. Restecg: resultados electrocardiográficos en reposo.
8. Thalach: frecuencia cardíaca máxima alcanzada.
9. Exang: angina inducida por el ejercicio (1 sí). La angina es un tipo de dolor en el pecho causado por la reducción del flujo sanguíneo al corazón.
10. Pico antiguo: depresión del ST inducida por el ejercicio en relación con el descanso.
11. Pendiente: la pendiente del segmento ST de ejercicio máximo.
12. Ca: número de vasos principales (0-3) coloreados por fluoroscopia.
13. Thal: no se proporcionó explicación, pero probablemente talasemia (3 normales; 6 defectos fijos; 7 defectos reversibles).
14. Objetivo (T): sin enfermedad = 0 y enfermedad = 1, (estado de enfermedad angiográfica).

Resultados

Classifiers	Accuracy (%)	Specificity	Sensitivity
Logistic regression	83.3	82.3	86.3
K neighbors	84.8	77.7	85.0
SVM	83.2	78.7	78.2
Random forest	80.3	78.7	78.2
Decision tree	82.3	78.9	78.5
DL	94.2	83.1	82.3



Procesamiento de los Datos

DataSet 'Heart Attack'

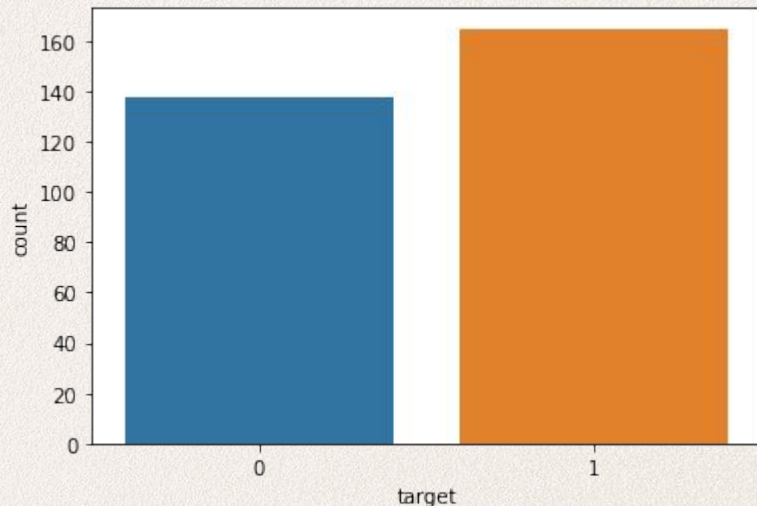
```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 303 entries, 0 to 302
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	age	303 non-null	int64
1	sex	303 non-null	int64
2	cp	303 non-null	int64
3	trestbps	303 non-null	int64
4	chol	303 non-null	int64
5	fbs	303 non-null	int64
6	restecg	303 non-null	int64
7	thalach	303 non-null	int64
8	exang	303 non-null	int64
9	oldpeak	303 non-null	float64
10	slope	303 non-null	int64
11	ca	303 non-null	int64
12	thal	303 non-null	int64
13	target	303 non-null	int64

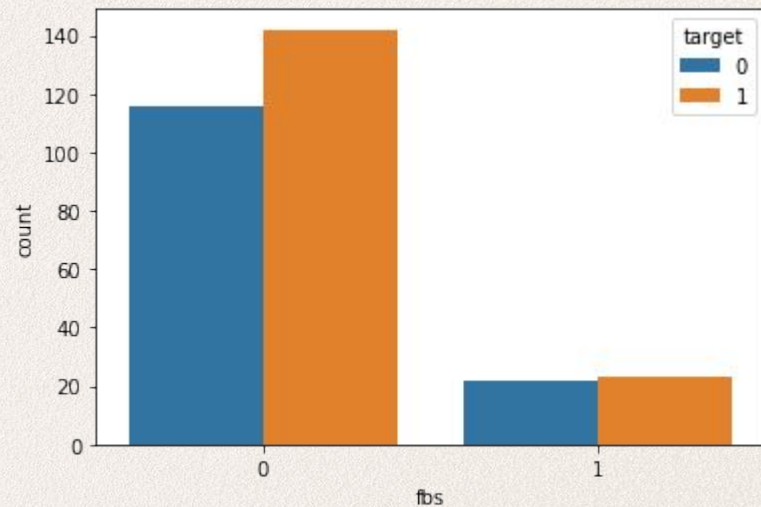
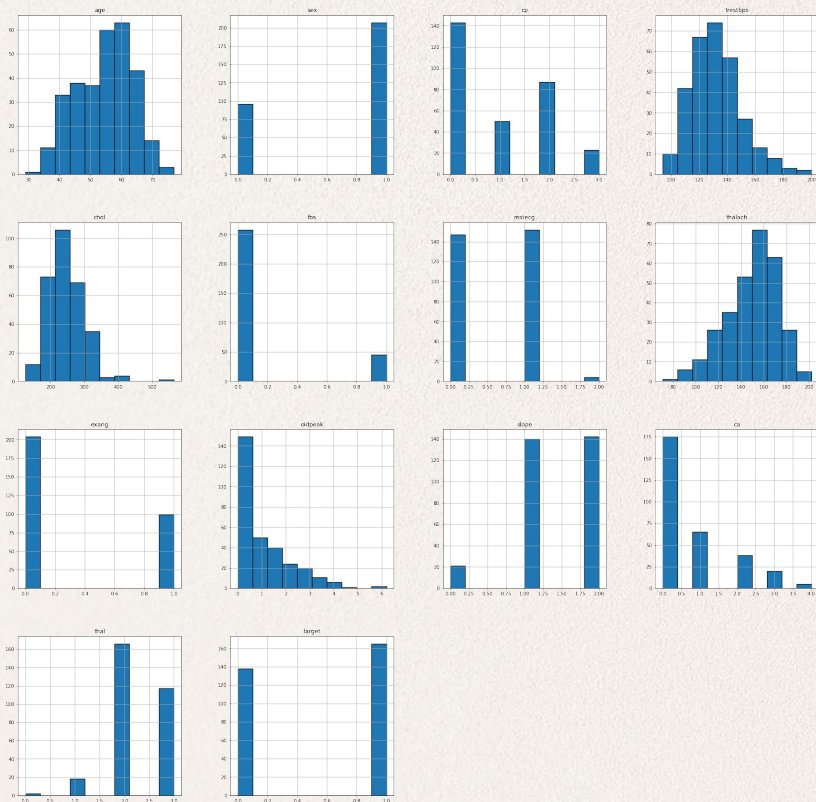
```
dtypes: float64(1), int64(13)
```



4 Características Categóricas

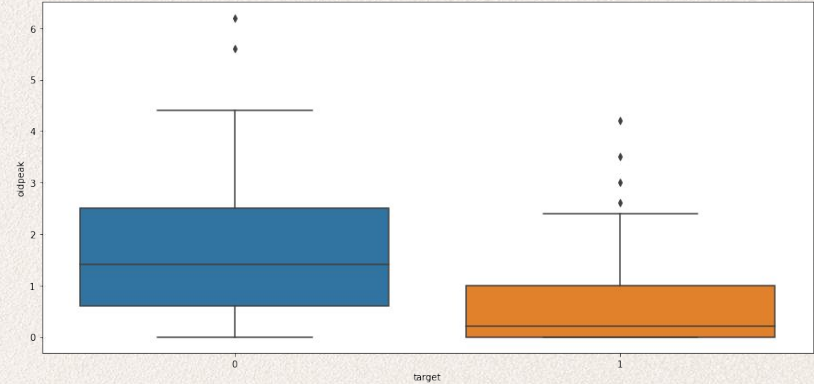
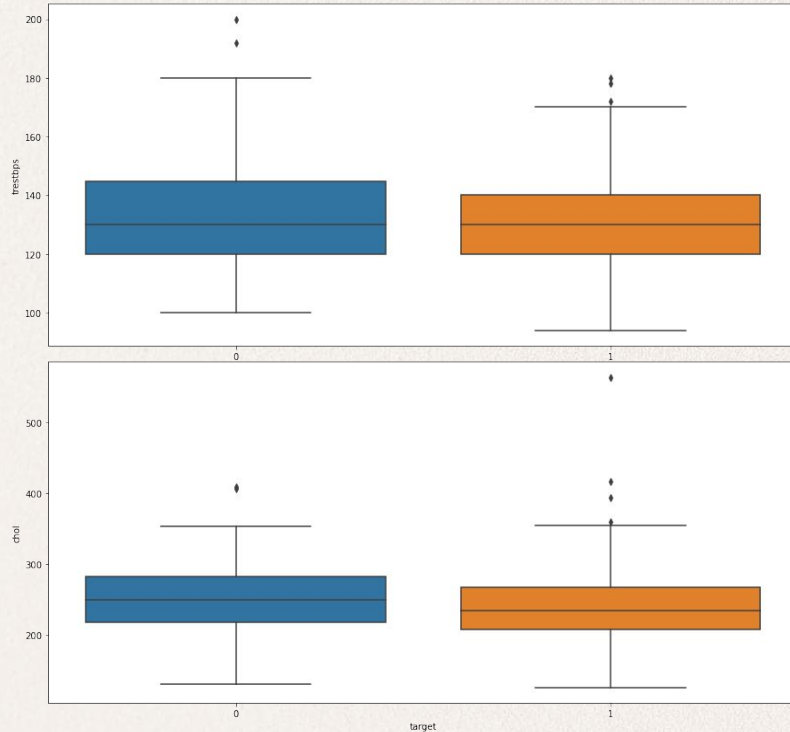
9 Características Numéricas

Analisis de las Distribuciones



Característica 'fbs' tiene el valor cero para el 87% de sus instancias (VarianceThreshold)

Eliminación de valores atípicos



BoxPlot de columnas **'trestbps'** , **'chol'** & **'oldpeak'**

Heat Map



Transformación de Variables Categóricas

RestECG	
normal	0
Onda ST-T abnormal	1
Hipertrofia ventricular	2

CP	
Angina típica	0
Angina atípica	1
Dolor no anginoso	2
Asintomático	3

Slope	
Pendiente ascendente	0
Plano	1
Pendiente descendente	2

THAL	
Normal	0
Defecto fijo	1
Defecto reversible	2

Transformador Utilizado:
OneHotEncoding

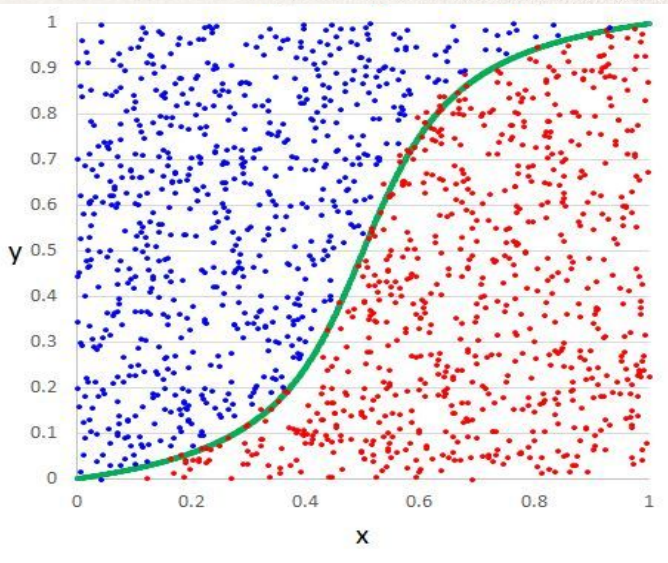
Estandarización de las características numéricas

StandardScaler()

The slide features a light gray background with decorative, thin, dark gray wavy lines in the corners. A large, empty rectangular box with a thin dark gray border is centered on the slide, containing the word "Modelos" in a large, black, serif font.

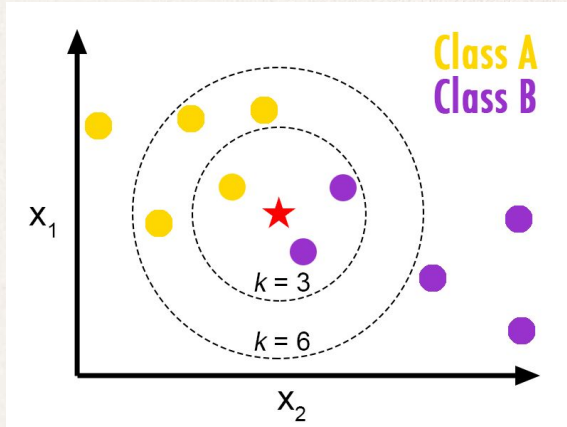
Modelos

Regresión Logística



```
class sklearn.linear_model.LogisticRegression(penalty='l2', *, dual=False,
tol=0.0001, C=1.0, fit_intercept=True, intercept_scaling=1,
class_weight=None, random_state=None, solver='lbfgs', max_iter=100,
multi_class='auto', verbose=0, warm_start=False, n_jobs=None,
l1_ratio=None)
```

Vecinos Más Cercanos (KNN)

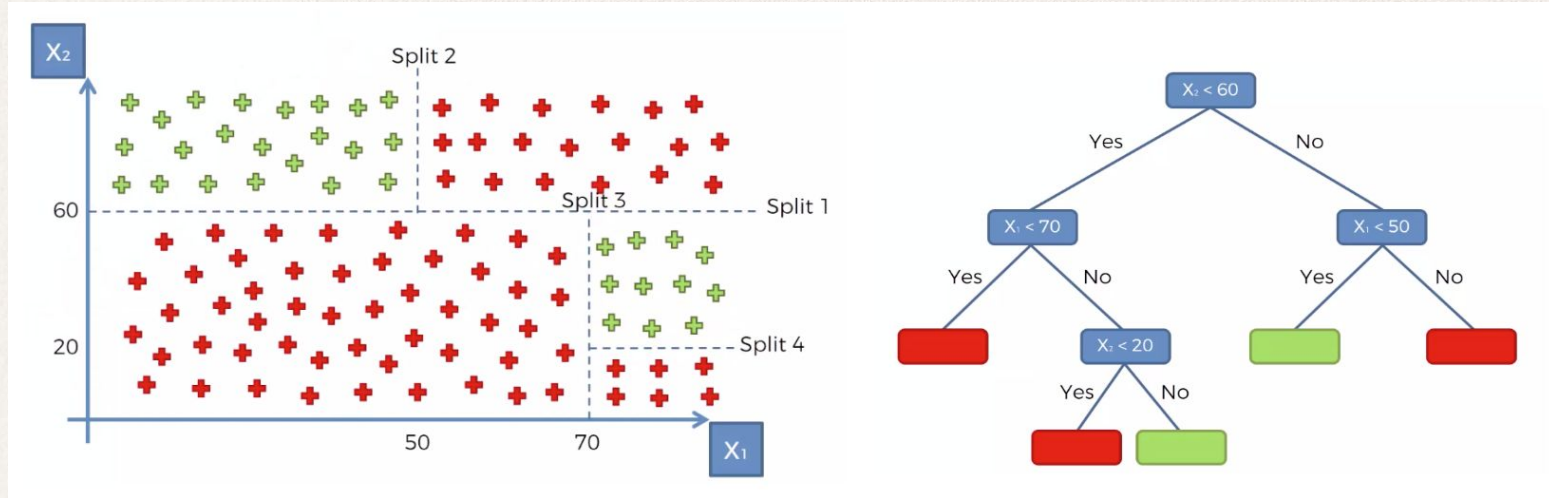


```
class sklearn.neighbors.KNeighborsClassifier(n_neighbors=5, *,  
weights='uniform', algorithm='auto', leaf_size=30, p=2, metric='minkowski',  
metric_params=None, n_jobs=None)
```

Hiperparámetros ingresados al GridSearch:
n_neighbors, leaf_size

```
KNeighborsClassifier(n_neighbors=39, leaf_size = 1)
```

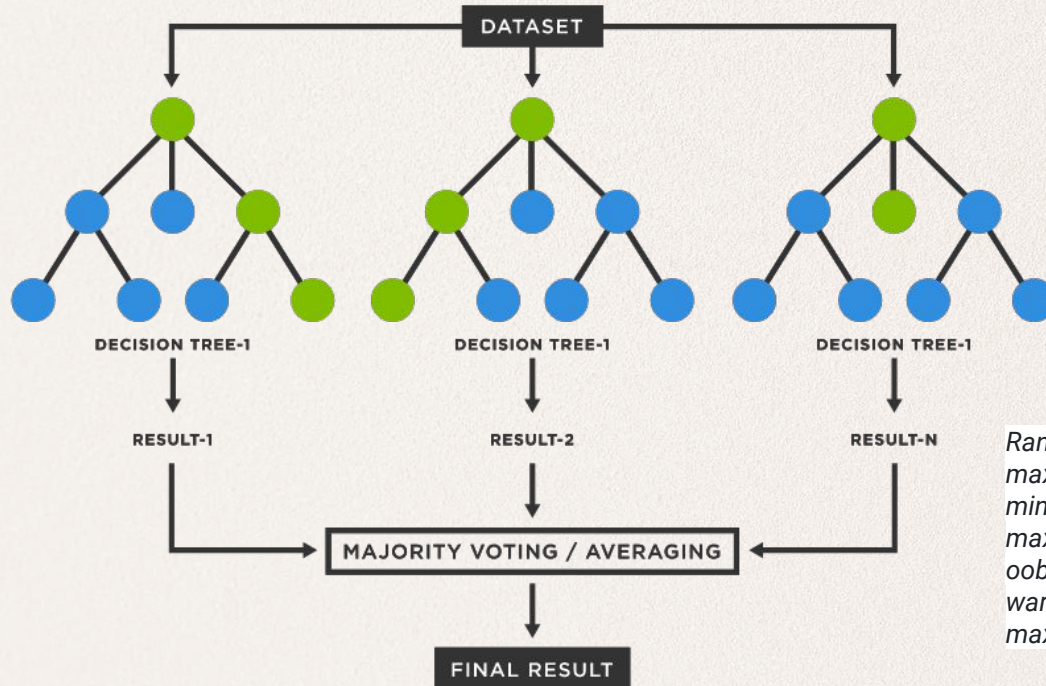

Decision Tree



El objetivo final de los árboles de decisión es ser capaz de entregar una respuesta óptima (teniendo en cuenta la mayor ganancia de información) al llegar al final de cada nodo.

```
DecisionTreeClassifier(criterion='entropy', splitter='random', max_depth=4,  
    random_state=2, class_weight={1:8}, min_samples_split=10,  
    min_samples_leaf=4)
```

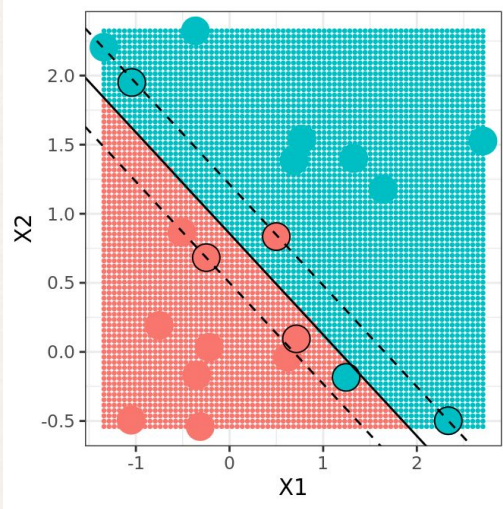
Random Forest



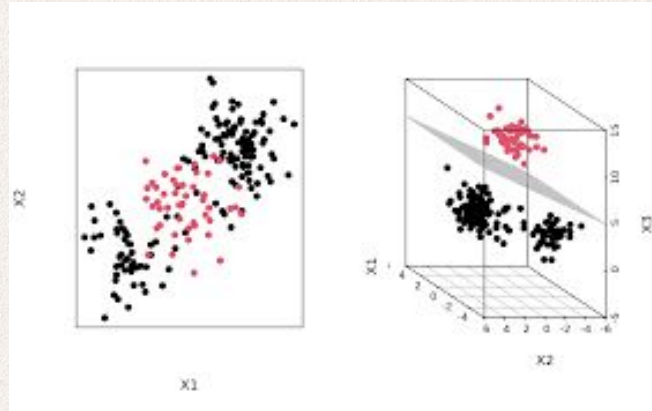
Modelo que ajusta un número de árboles de decisión en varias muestras del set de datos y obtiene el promedio para mejorar la precisión y controlar el sobreajuste.

```
RandomForestClassifier(n_estimators=100, *, criterion='gini',  
max_depth=None, min_samples_split=2, min_samples_leaf=1,  
min_weight_fraction_leaf=0.0, max_features='auto',  
max_leaf_nodes=None, min_impurity_decrease=0.0, bootstrap=True,  
oob_score=False, n_jobs=None, random_state=None, verbose=0,  
warm_start=False, class_weight=None, ccp_alpha=0.0,  
max_samples=None)
```


Máquinas de Vectores Soportes



Se utilizan planos o hiperplanos para separar clases en regiones, donde se busca obtener el máximo margen alrededor del hiperplano que logre separar mejor las clases.

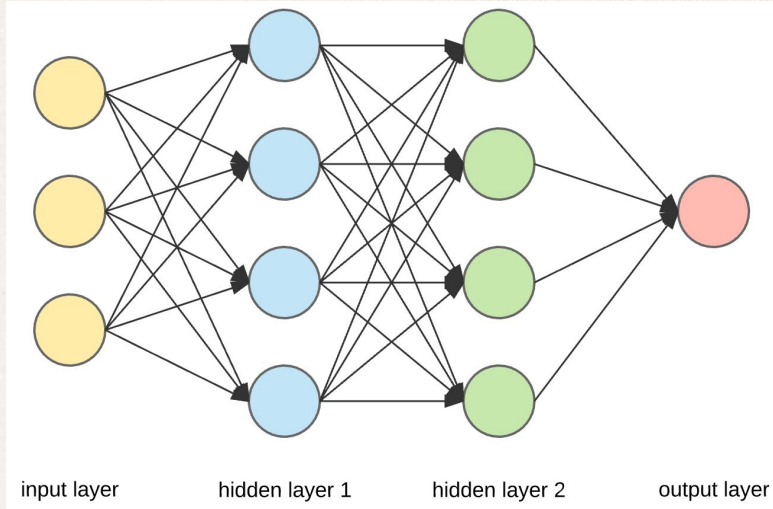


```
class sklearn.svm.SVC(*, C=1.0,  
kernel='rbf', degree=3, gamma='scale',  
coef0=0.0, shrinking=True,  
probability=False, tol=0.001,  
cache_size=200, class_weight=None,  
verbose=False, max_iter=-1,  
decision_function_shape='ovr',  
break_ties=False, random_state=None)
```

Hiperparámetros ingresados al GridSearch:
Kernel, C

SVC(kernel='linear', C=1)

Redes Neuronales (ANN)



```
tf.keras.wrappers.scikit_learn.KerasClassifier(  
    build_fn=None, **sk_params)
```

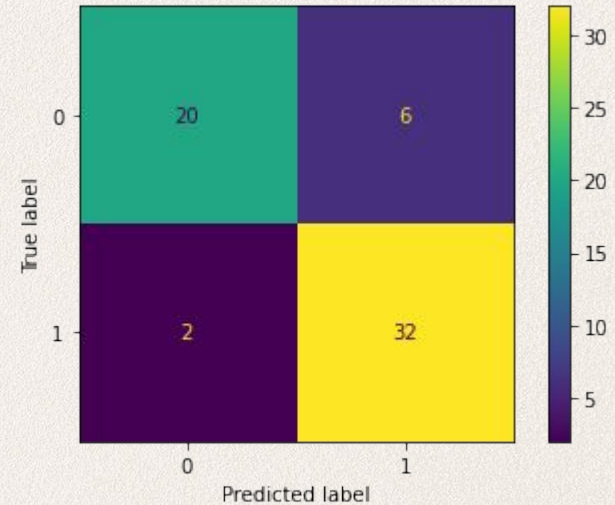
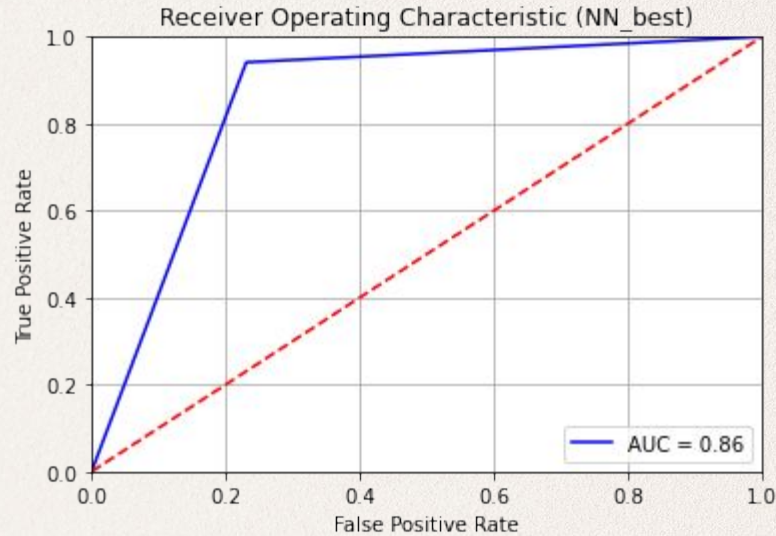
Hiperparámetros ingresados al GridSearch:
epochs, batch_size

```
KerasClassifier(build_fn= model, batch_size= 30, epochs =512)
```


The background of the slide is a light gray color. It is decorated with several thin, dark gray, hand-drawn style lines that flow across the page. These lines are most prominent in the top-left, bottom-left, and bottom-right corners, creating a sense of movement and elegance. A central rectangular box with a thin black border contains the main title.

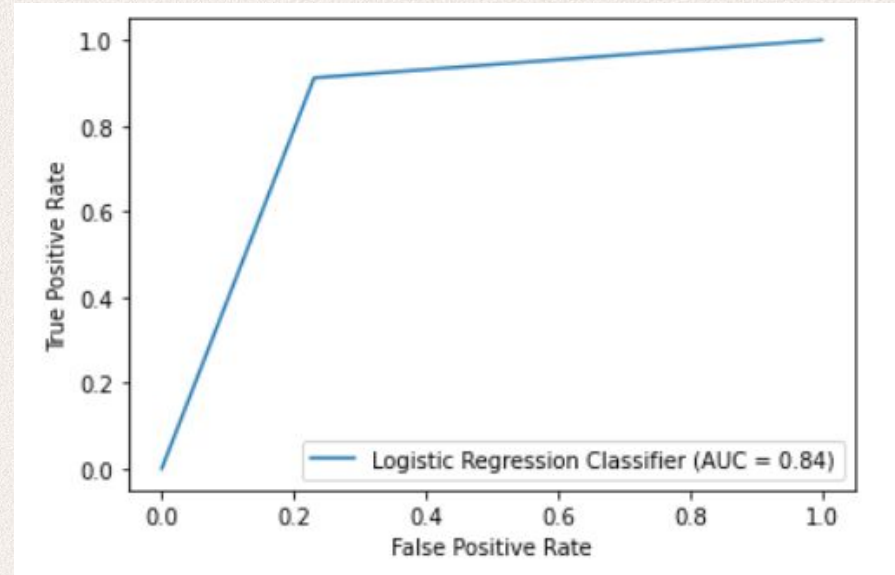
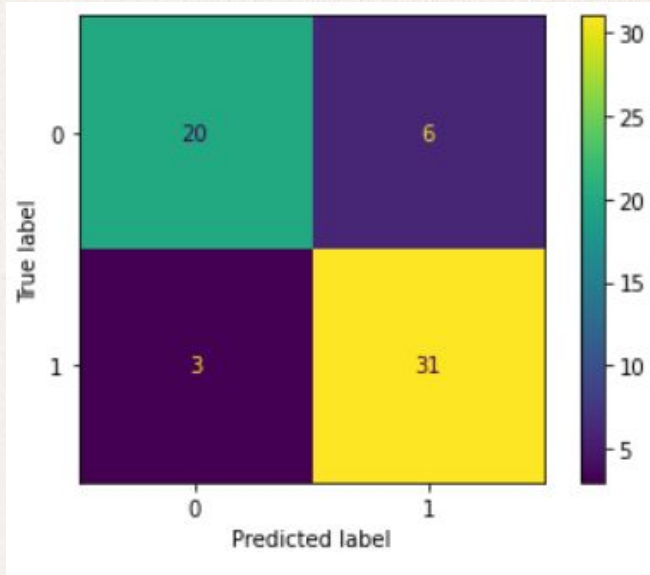
RESULTADOS

Vecinos Más Cercanos (KNN)



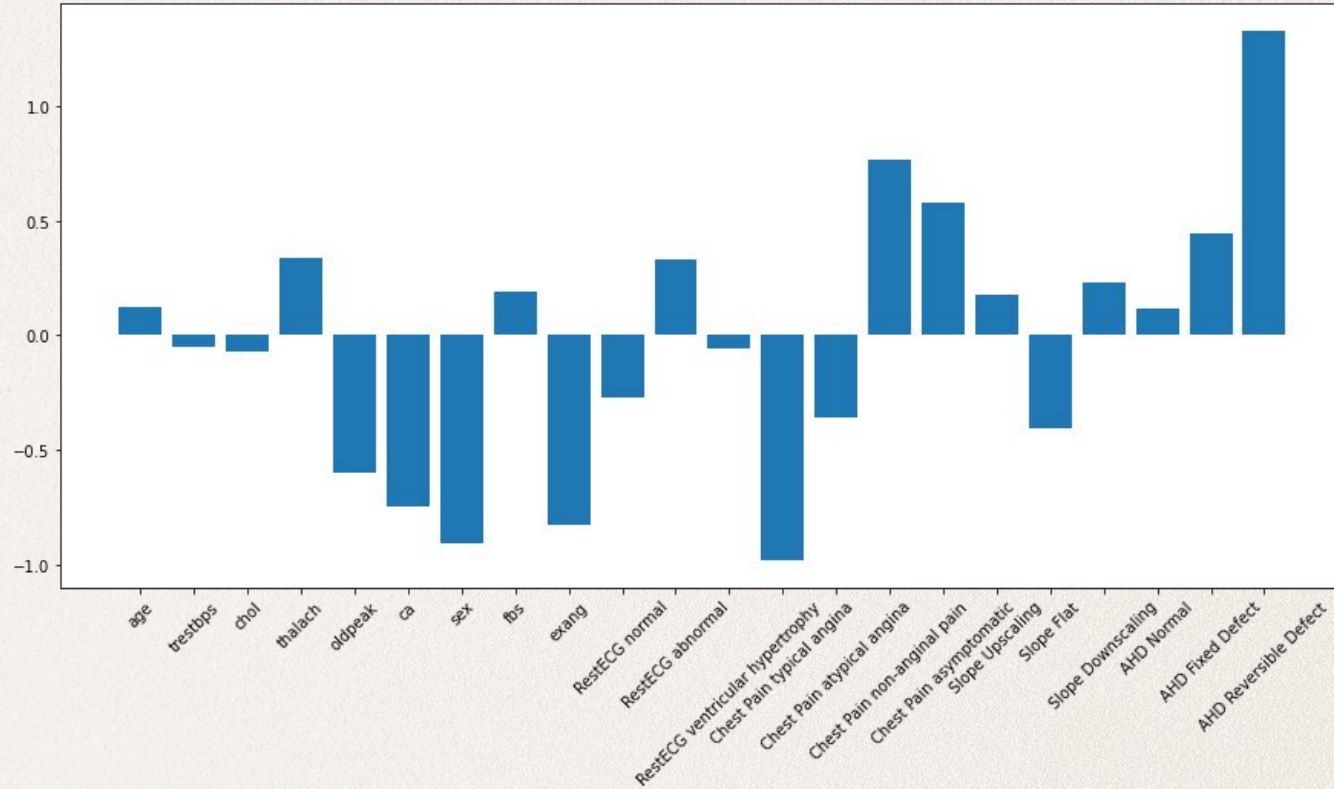
Puntaje de Precisión: 86.66 %

Regresión Logística

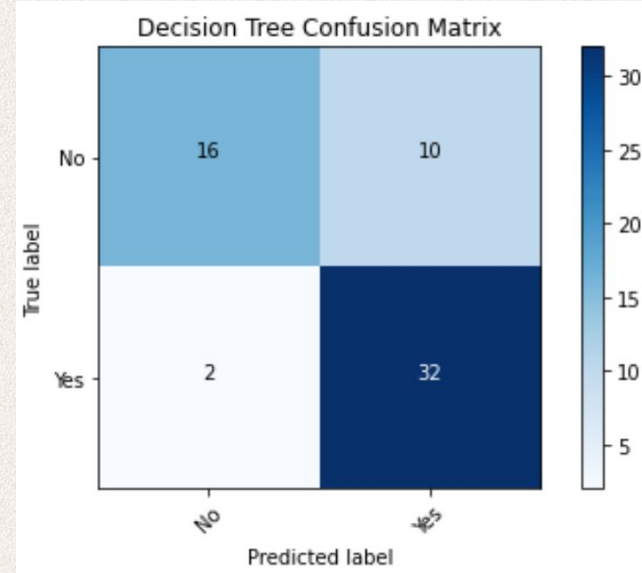
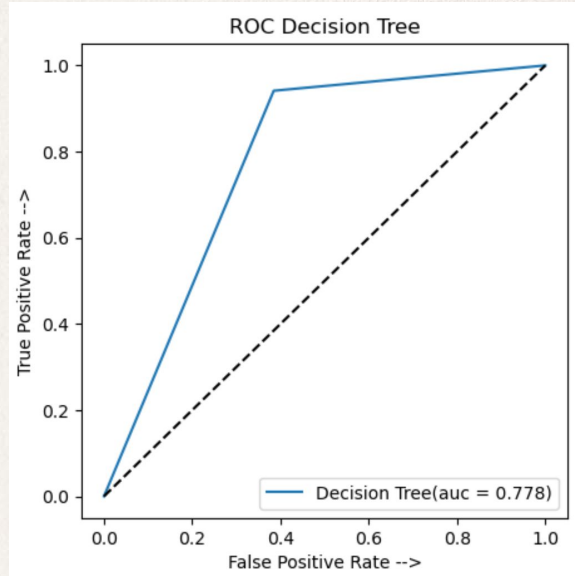


Puntaje de Precisión: 85 %

Regresión Logística

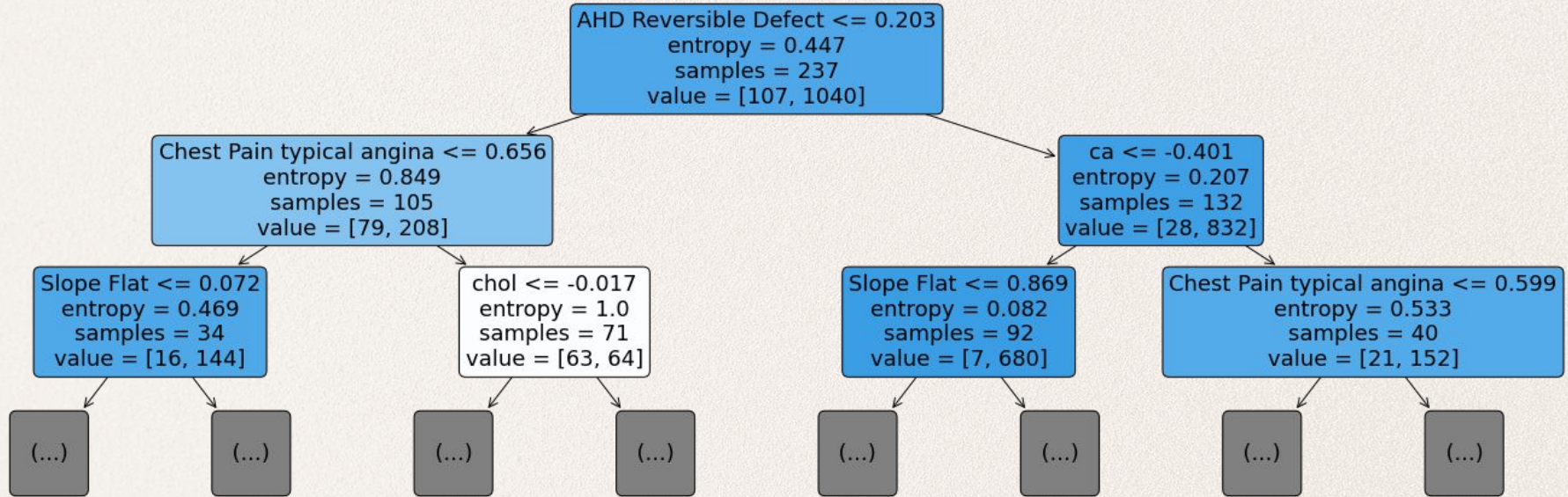


Decision Tree

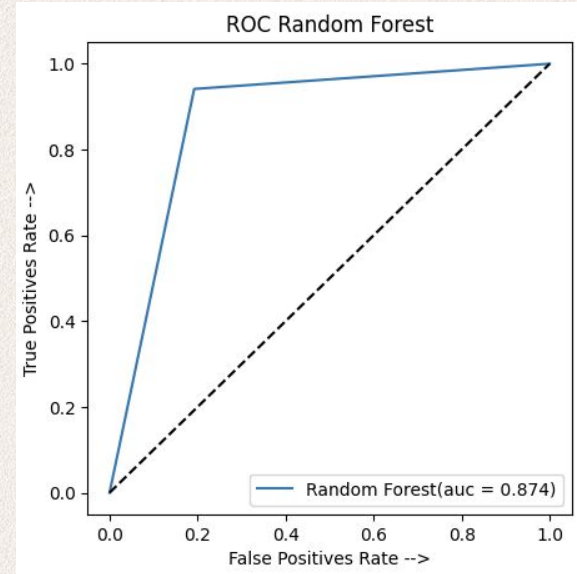
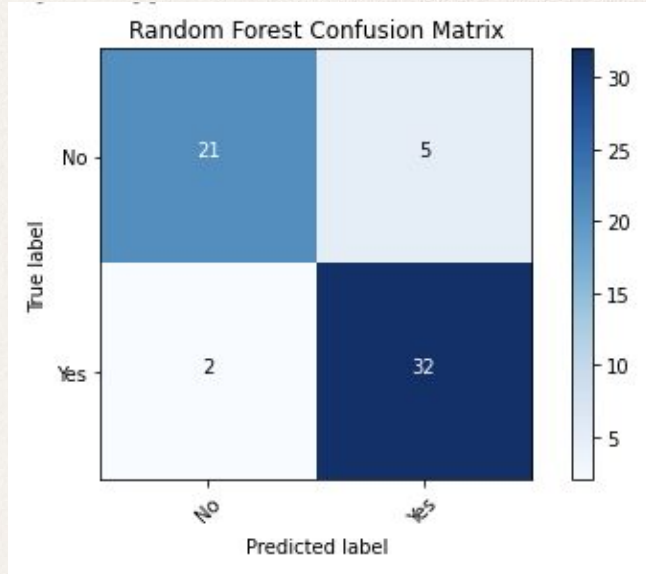


Puntaje de Precisión: 80 %

Decision Tree

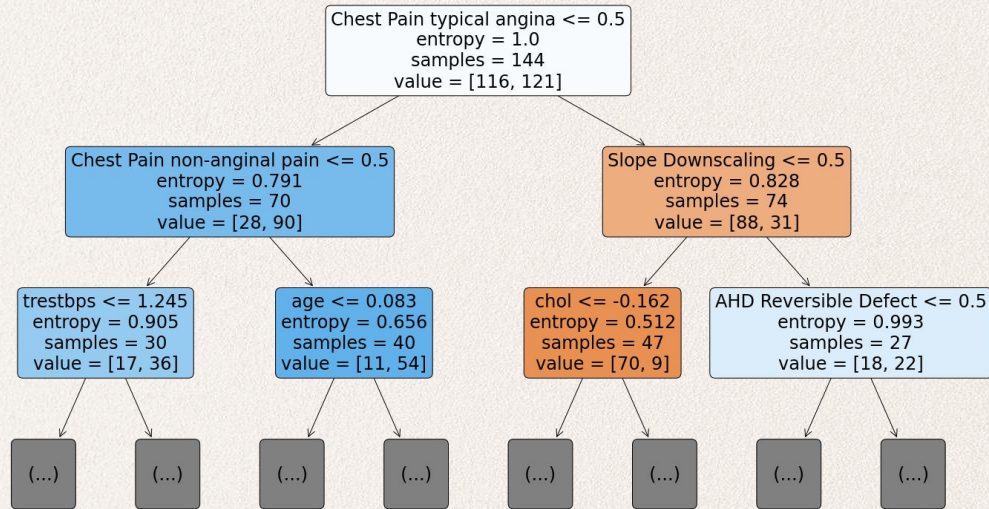


Random Forest



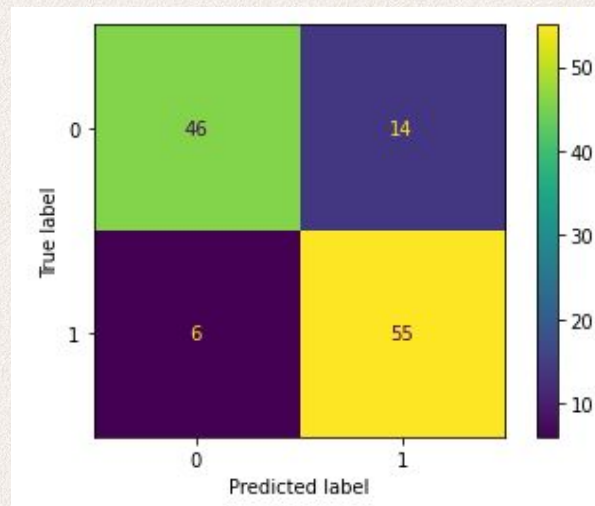
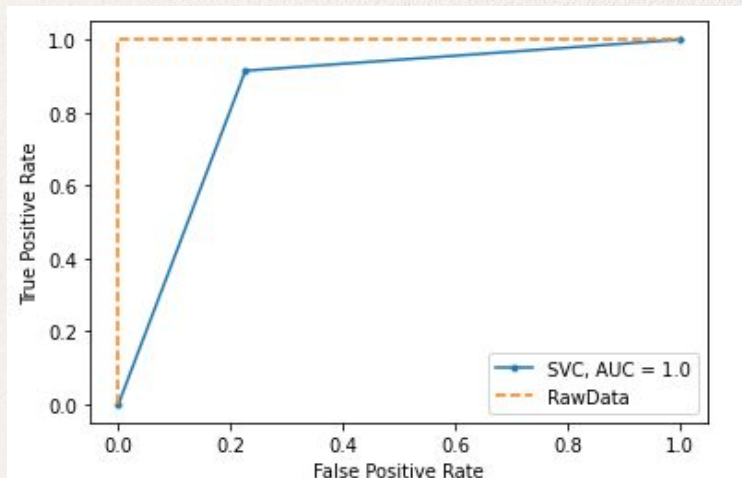
Puntaje de Precisión: 88.33%

Random Forest



Puntaje de Precisión: 88.33%

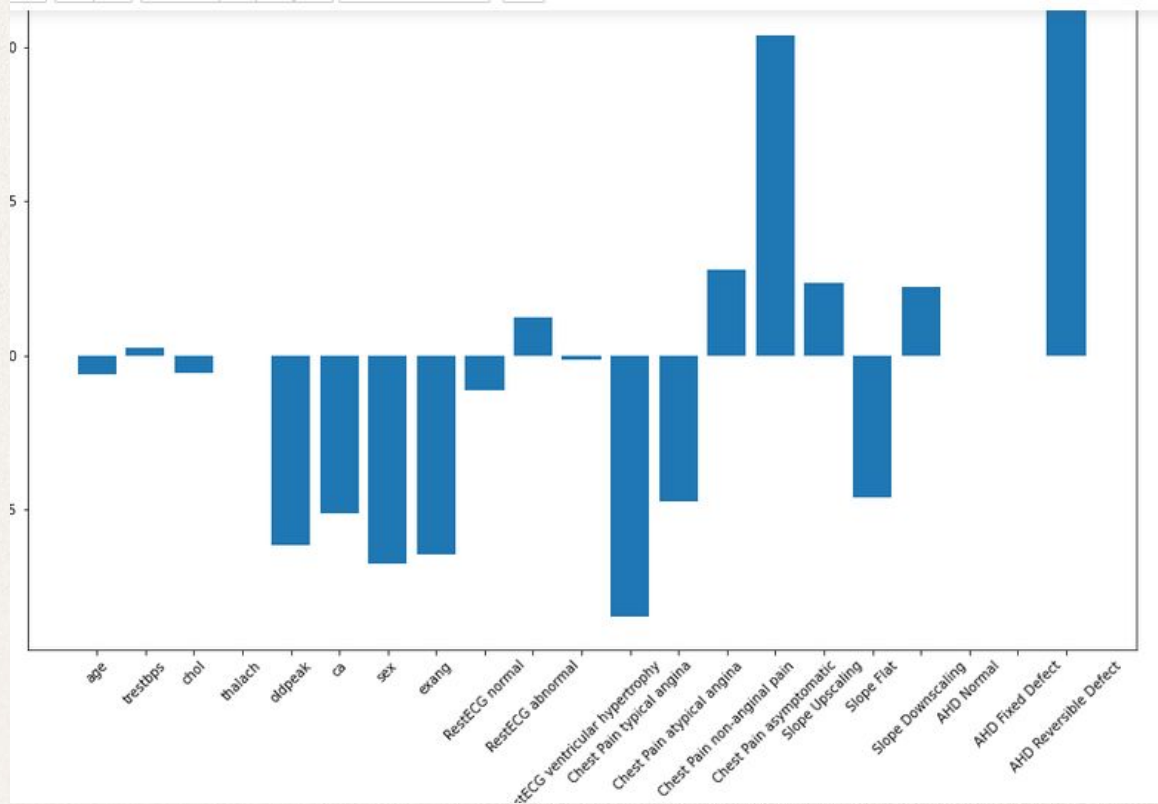
Máquinas de Vectores de Soporte (SVC)



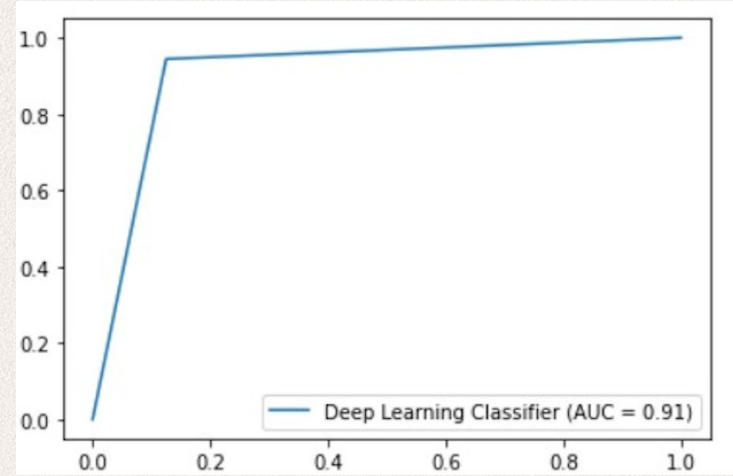
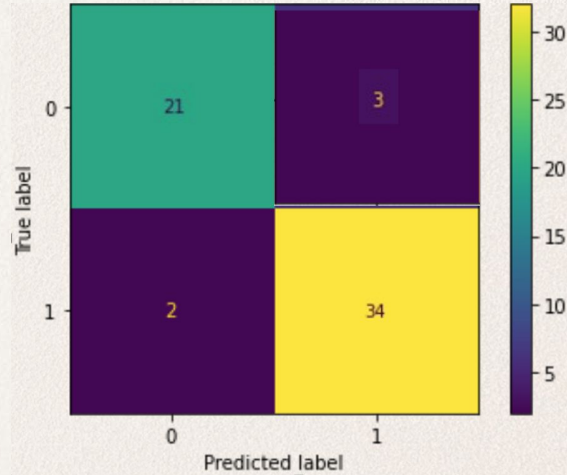
Puntaje de Precisión: 83.47%

Recall : 0.85, Precision: 0.844 --> F score: 0.847

Maquinas de Vectores de Soports



Red Neuronal (ANN)



Puntaje de Precisión: 91%

REFERENCIAS

- American Heart Association. What is a Heart Attack?
<https://www.heart.org/en/health-topics/heart-attack/about-heart-attacks>
 - American Heart Association. Lifestyle Changes for Heart Attack Prevention.
<https://www.heart.org/en/health-topics/heart-attack/life-after-a-heart-attack/lifestyle-changes-for-heart-attack-prevention>
 - Kaggle, Heart Attack Dataset
<https://www.kaggle.com/pritsheta/heart-attack>
 - Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning
<https://www.hindawi.com/journals/cin/2021/8387680/>
- 