

Dominando el Entrenamiento de Modelos de Machine Learning

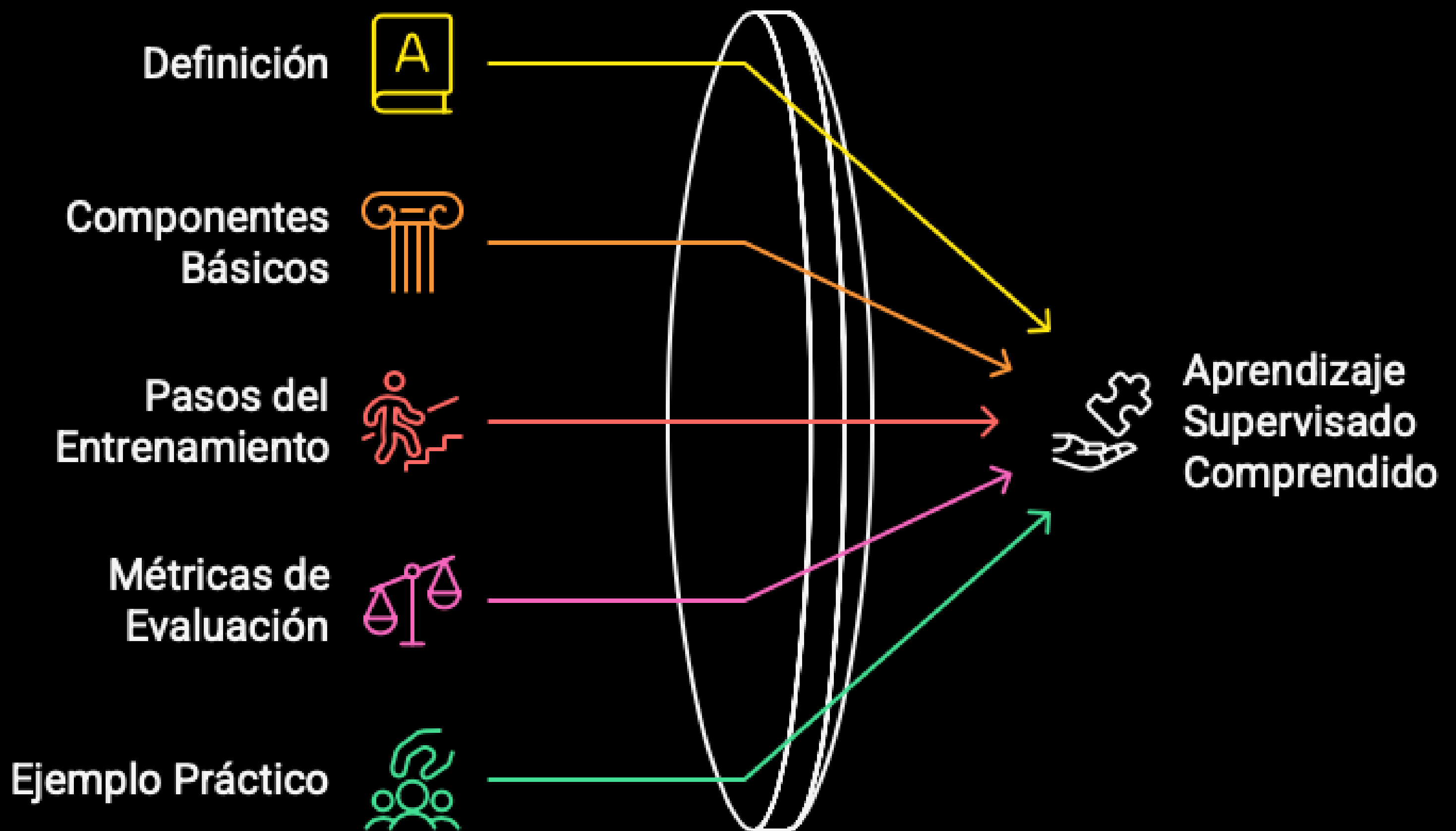
**Una guía completa para
construir Modelos
Supervisados**



Edgar Rios Linares

Ruta

Camino al Dominio del Aprendizaje Supervisado



Definición

Le "enseñamos" a una máquina a partir de datos etiquetados para que aprenda a predecir resultados en el futuro.



Definición

El aprendizaje supervisado es una técnica en la que un modelo es entrenado con un conjunto de datos etiquetados.

Cada ejemplo de entrenamiento está acompañado de una salida deseada.

El objetivo es que el modelo aprenda a mapear entradas(features) a salidas correctas(labels).

Componentes

Conjunto de datos etiquetados

Características (features)

Etiquetas (labels)

Función objetivo



Componentes

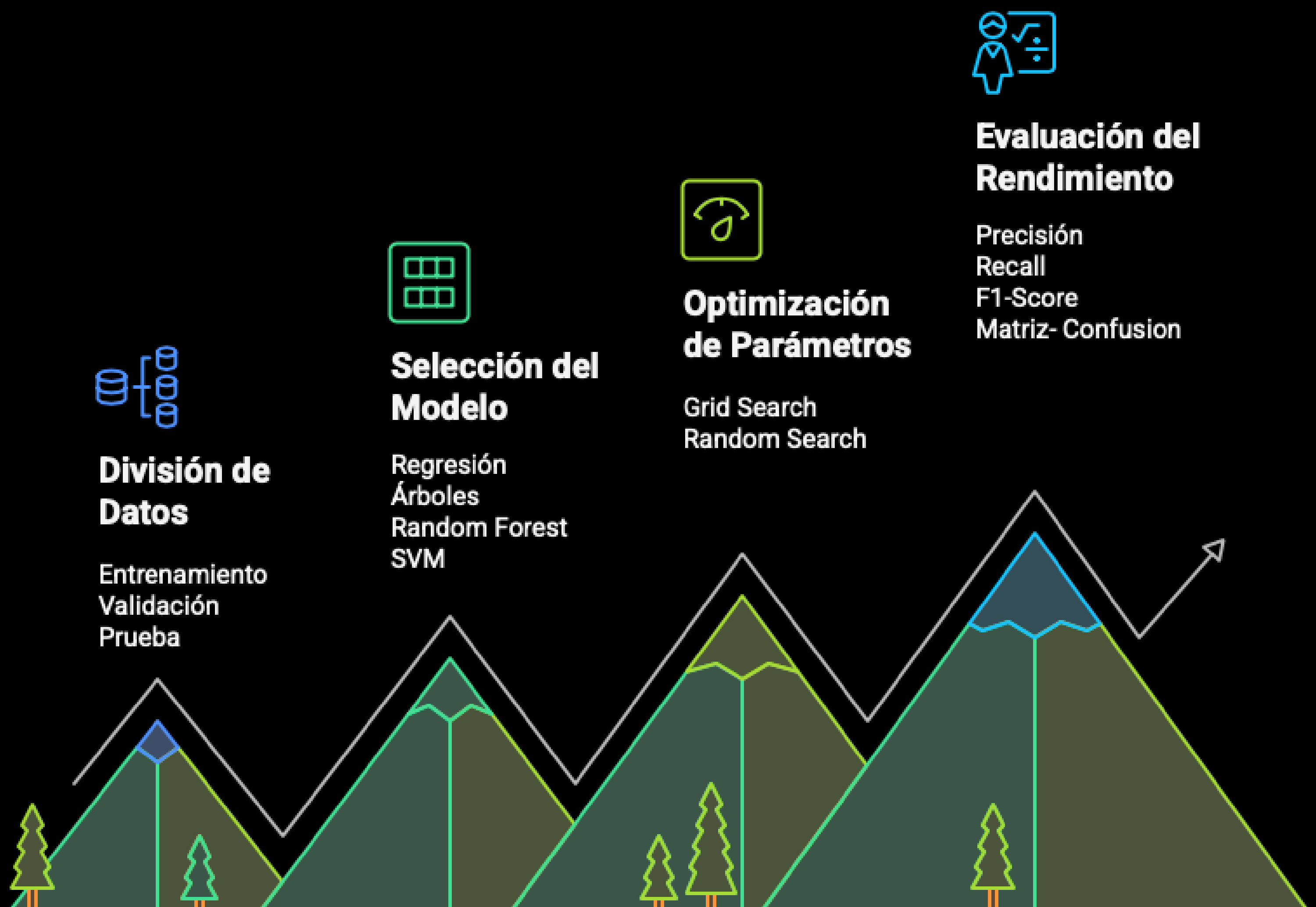
Conjunto de datos etiquetados: Datos de entrada con sus respectivas salidas esperadas.

Características (features): Atributos relevantes que se utilizarán para hacer predicciones (píxeles de una imagen)

Etiquetas (labels): Valores objetivo que el modelo debe aprender a predecir (perro o gato)

Función Objetivo: Métrica que el modelo optimiza durante el entrenamiento para mejorar sus predicciones.

Pasos del Entrenamiento



Pasos del Entrenamiento

1. División de Datos

Entrenamiento: Entrenar el modelo. (70-80%)

Validación: Ajustar hiperparámetros. (10-15%)

Prueba: Evaluar el desempeño final. (10-15%)

2. Selección del Modelo

Regresión, árboles, SVM, redes, etc.

3. Optimización de Parámetros

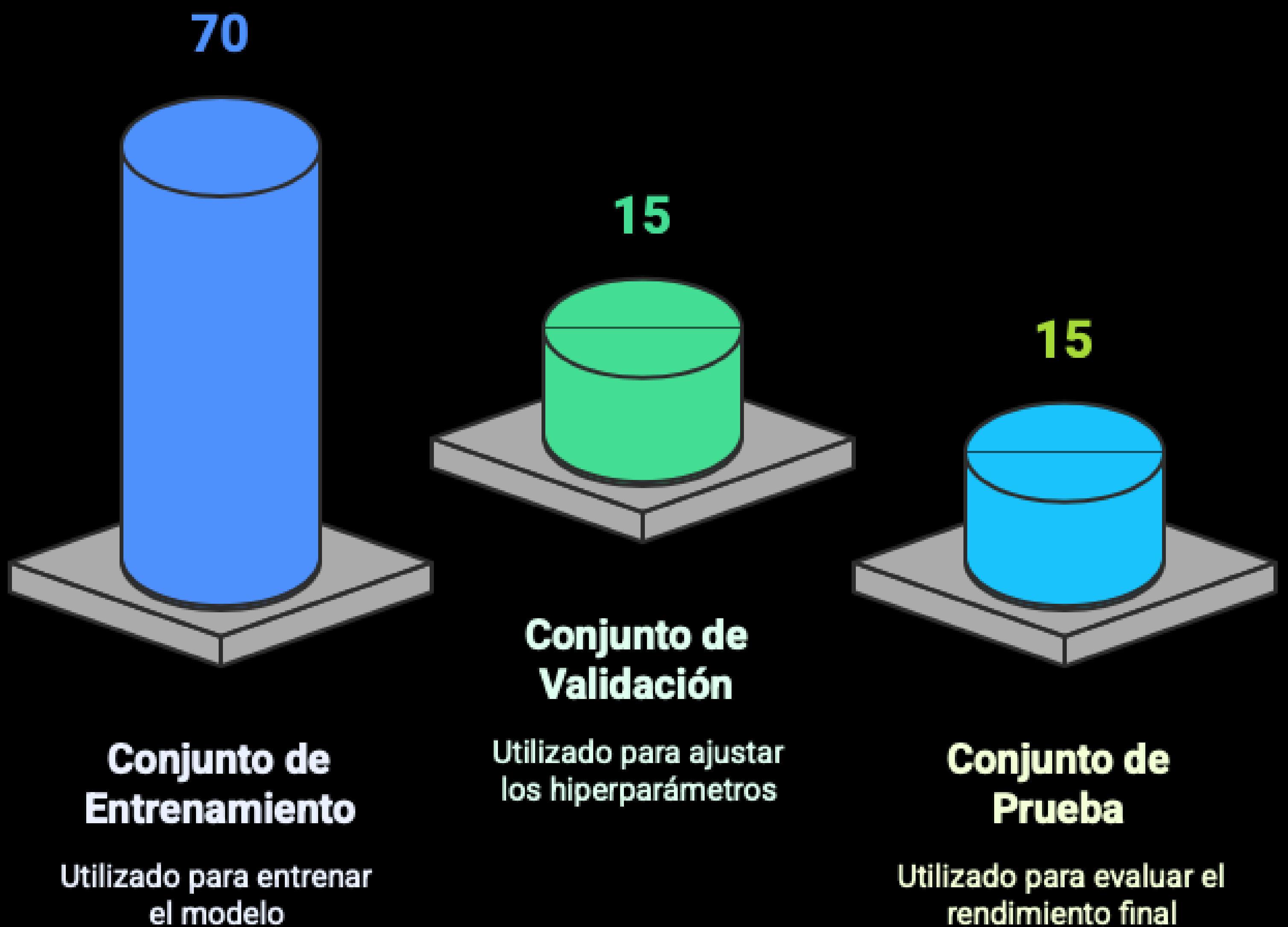
Validación cruzada, Grid Search, etc.

4. Evaluación del Rendimiento

Rendimiento en validación y prueba.

Pasos del Entrenamiento

1. División de Datos



Pasos del Entrenamiento

1. División de Datos

Features (X)				Label (y)
X1	X2	...	Xn	y
999	787	678	554	CAT
123	332	444	555	DOG
123	333	405	554	DOG
999	787	678	554	CAT
123	333	405	554	DOG
999	787	678	554	CAT
999	787	678	554	CAT
123	787	405	554	DOG
999	787	678	554	CAT
999	787	678	554	CAT
123	333	543	233	DOG

Training 70%

Test 15%

Validation 15%

Pasos del Entrenamiento

2. Selección del Modelo



Pasos del Entrenamiento

3. Optimización de Parámetros

Features (X)				Label (y)
X1	X2	...	Xn	y
999	787	678	554	CAT
123	332	444	555	DOG
123	333	405	554	DOG
999	787	678	554	CAT
123	333	405	554	DOG
999	787	678	554	CAT
999	787	678	554	CAT
123	787	405	554	DOG
999	787	678	554	CAT
999	787	678	554	CAT
123	333	543	233	DOG

K1

K2

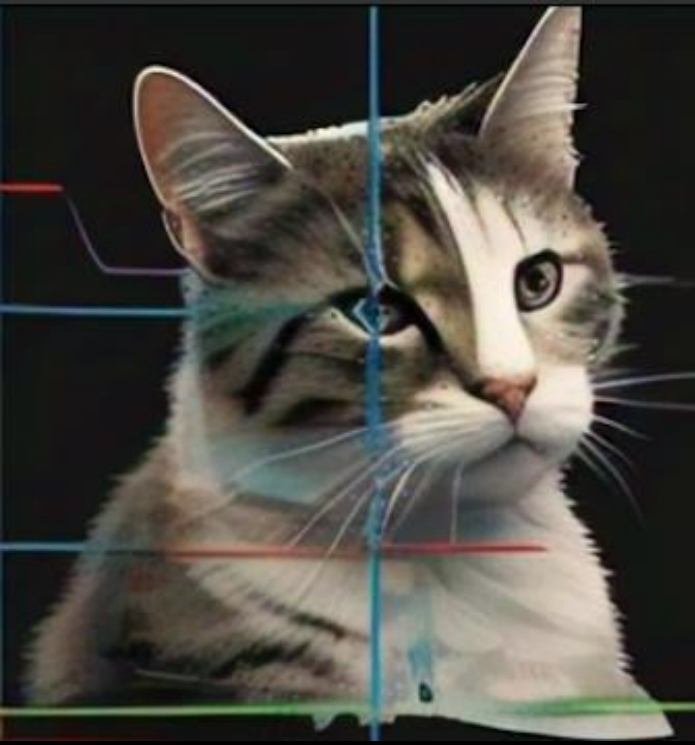
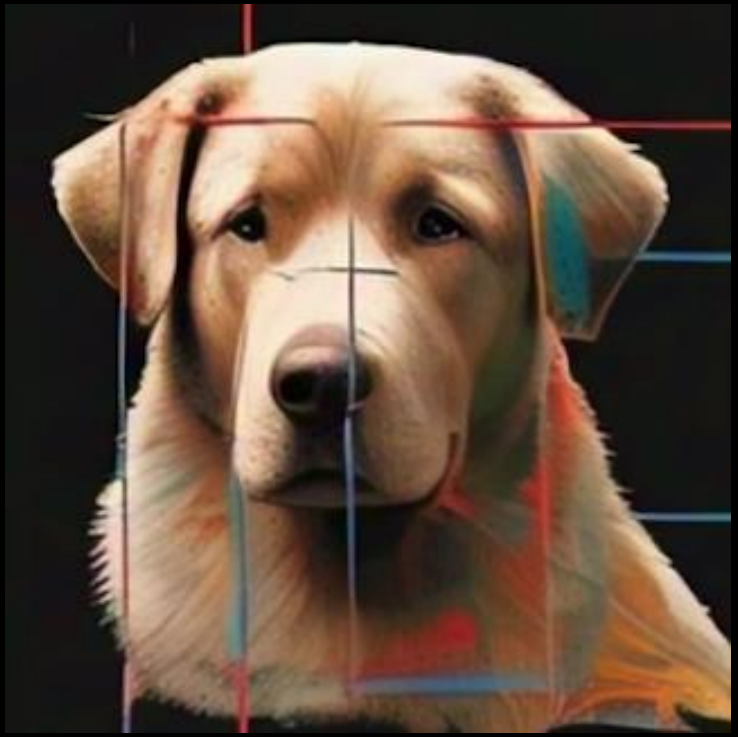
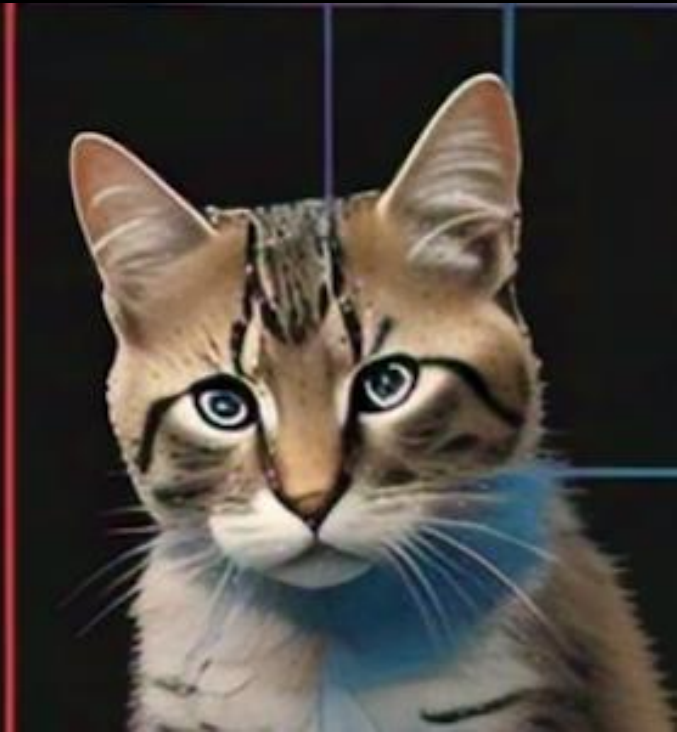
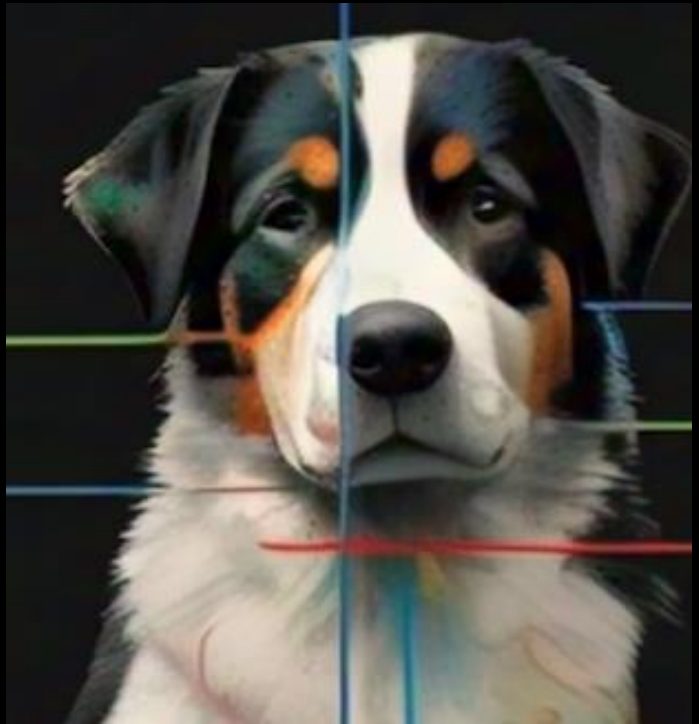
Training

Training

Test

Pasos del Entrenamiento

4. Evaluación del Rendimiento

		ACTUAL	
		CAT	NO CAT
PREDICTION	CAT	 TRUE POSITIVE	 FALSE POSITIVE ERROR
	NO CAT	 FALSE NEGATIVE ERROR	 TRUE NEGATIVE

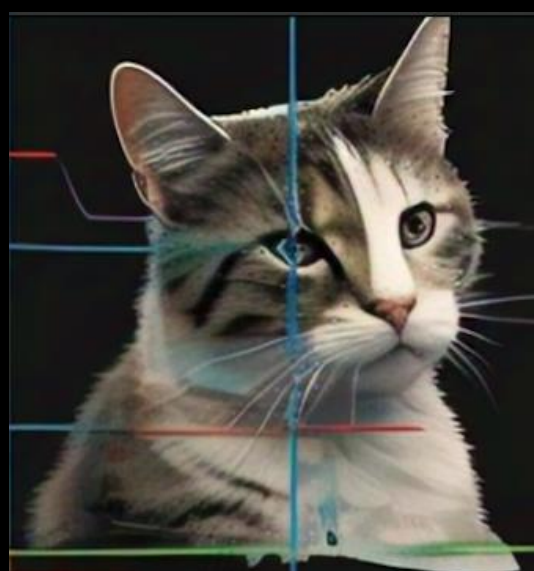
Métricas de Evaluación



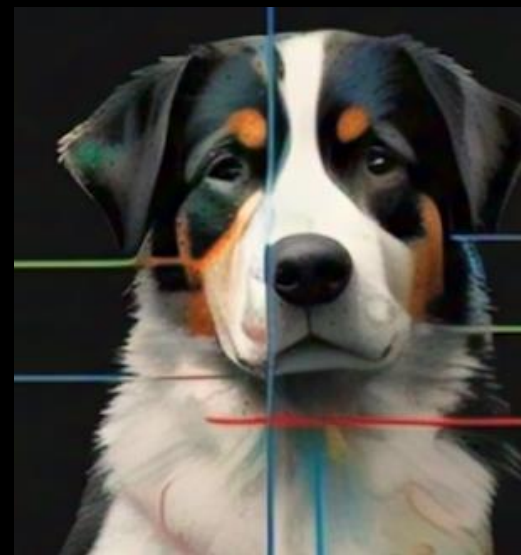
Métricas de Evaluación

Precisión (Accuracy) Proporción de predicciones correctas sobre el total de predicciones.

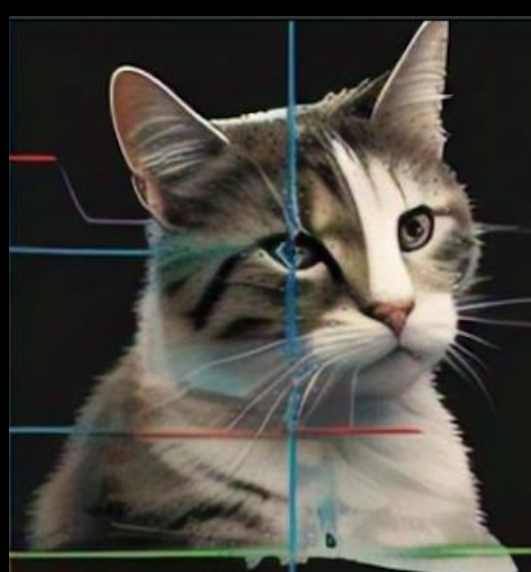
Fórmula: (Verdaderos positivos + Verdaderos negativos) / Total.



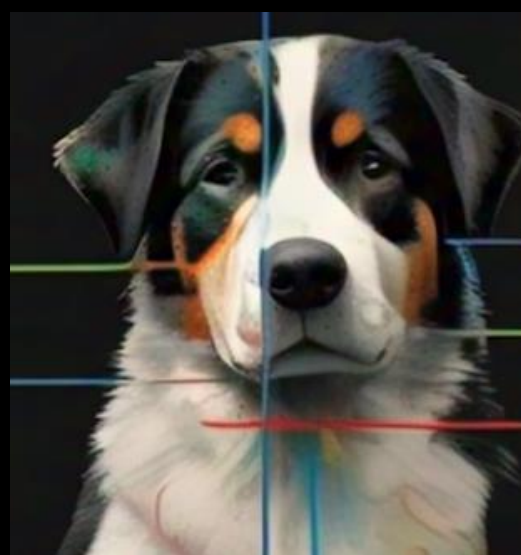
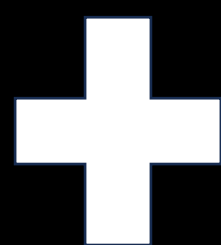
TRUE POSITIVE



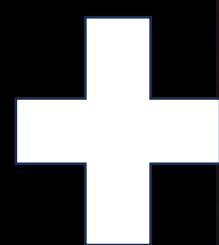
TRUE NEGATIVE



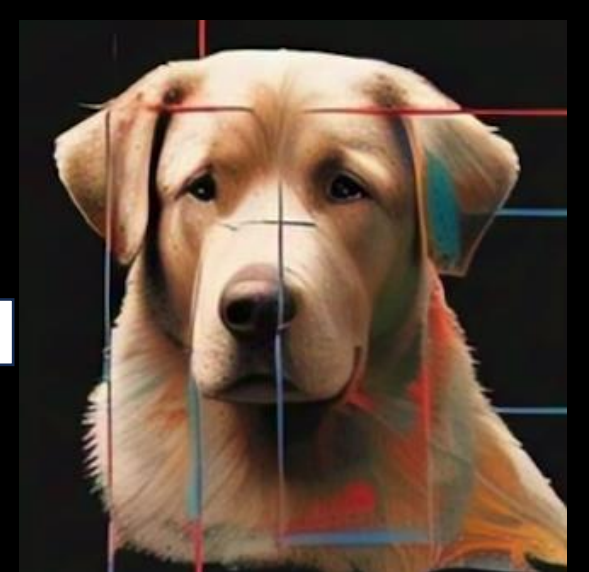
TRUE POSITIVE



TRUE NEGATIVE



FALSE NEGATIVE

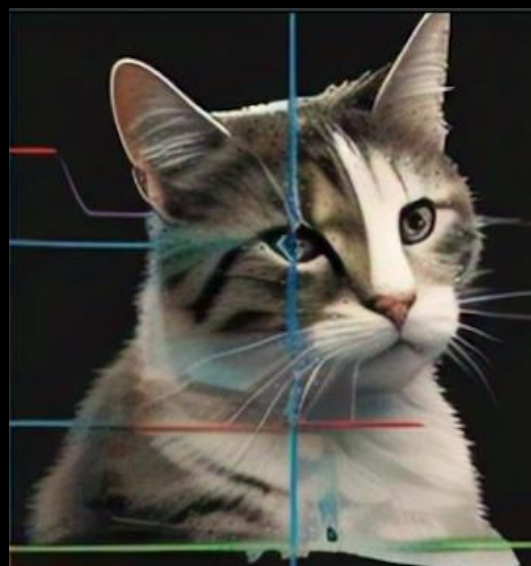


FALSE POSITIVE

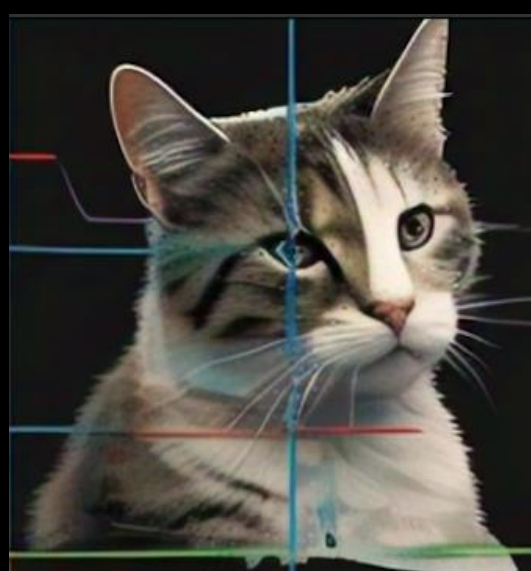
Métricas de Evaluación

Recall (Sensibilidad) Proporción de verdaderos positivos sobre el total de positivos reales.

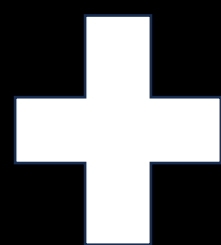
Fórmula: Verdaderos positivos / (Verdaderos positivos + Falsos negativos).



TRUE POSITIVE



TRUE POSITIVE



FALSE NEGATIVE

Métricas de Evaluación

F1-Score, Media de precisión y recall. Útil cuando hay desbalance en las clases.

Fórmula: $2 * (\text{Precisión} * \text{Recall}) / (\text{Precisión} + \text{Recall})$.

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Métricas de Evaluación

MSE (Error Cuadrático Medio): Para problemas de regresión, mide la diferencia entre valores reales y predichos

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Donde:

y_i son los valores reales.

\hat{y}_i son las predicciones del modelo.

n es el número total de observaciones.

Ejemplo Práctico

Predicción de Empresas Solventes

Paso 1: Recopilación de Datos

Se obtiene un conjunto de datos con características como ingresos, egresos, créditos y solvencia.

Paso 2: Preprocesamiento

Limpieza de datos y tratamiento de valores nulos.

Normalización de variables numéricas.

Paso 3: Entrenamiento del Modelo

División de datos en entrenamiento, validación y prueba.

Modelo LDA- relación de características y solvencia.

Paso 4: Evaluación

Se calculan, Precision, Recall, F1-Score y Matriz de

Confusión

Se ajustan hiperparámetros y se reentrena.

Ejemplo Práctico

PASO 1

	income	debt	credit_history	solvent
0	57450.712295	26483.242833	392.115420	0
1	47926.035482	33365.916483	357.512312	0
2	59715.328072	10210.026983	650.036637	1
3	72845.447846	23940.784657	688.561650	1
4	46487.699379	15445.502016	317.372380	1

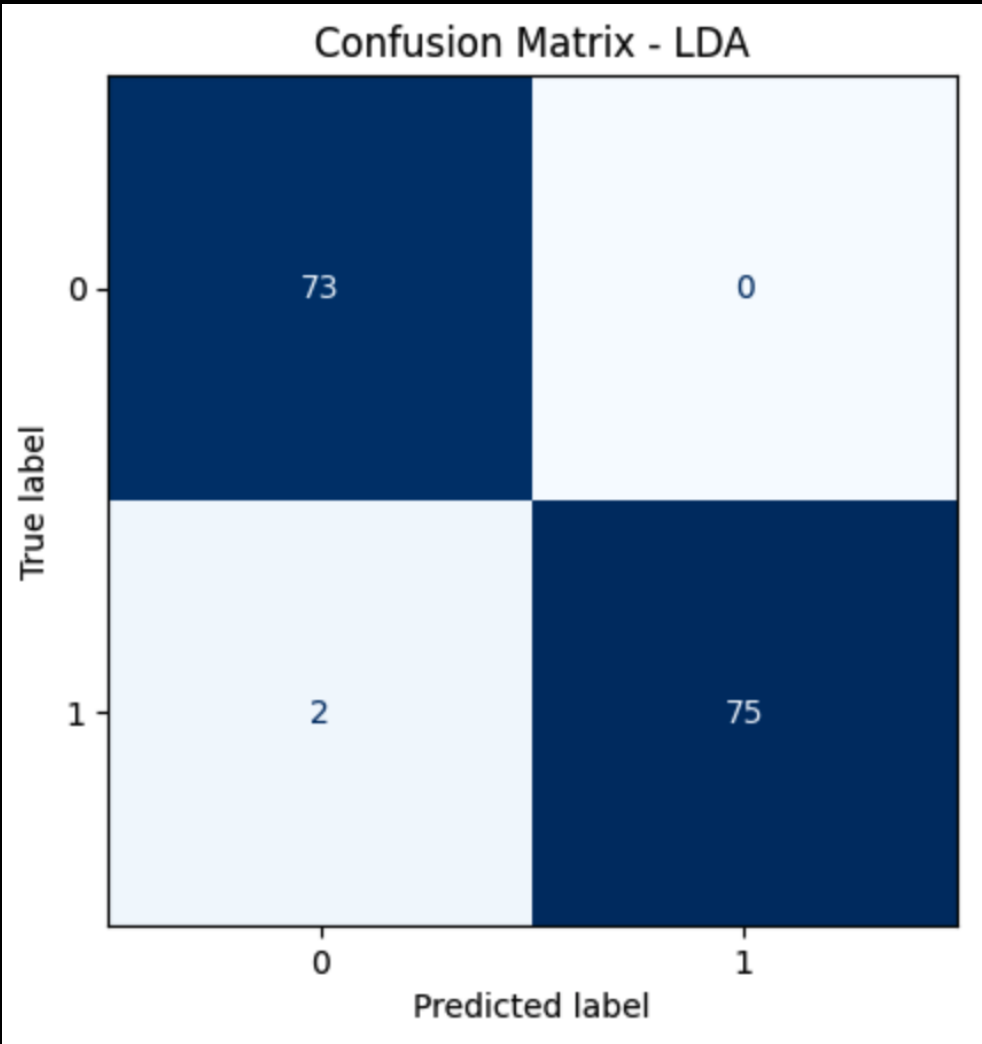
PASO 2

	income	debt	credit_history
0	-0.245825	-0.531158	1.500187
1	-0.042386	0.079393	-0.803370
2	-0.741307	-0.848214	-0.136724
3	-1.040191	0.480125	0.156964
4	-0.720783	0.168569	-1.356440

PASO 3

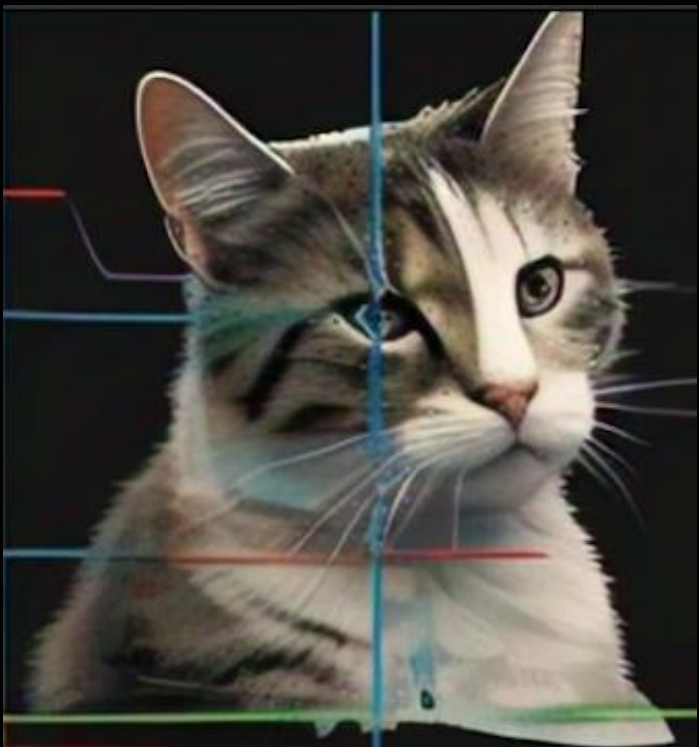
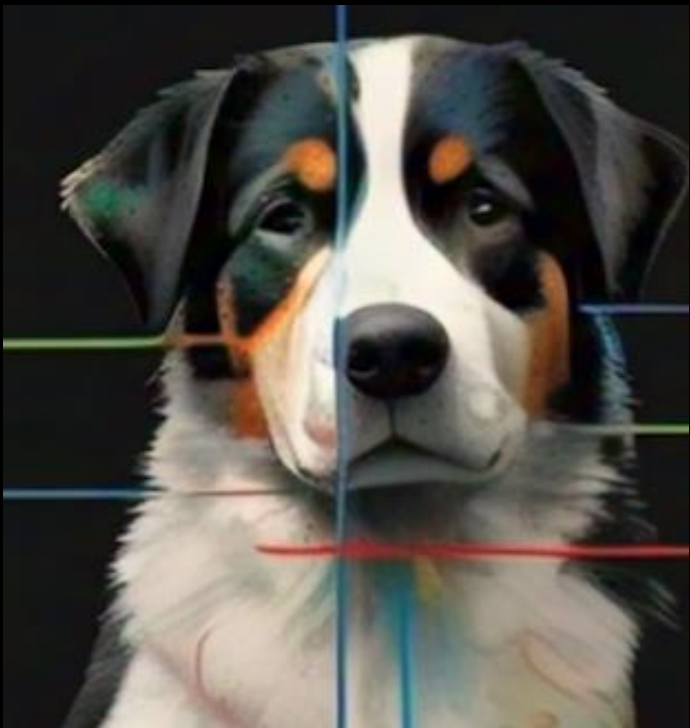
```
X_lda
LinearDiscriminantAnalysis
LinearDiscriminantAnalysis()
```

PASO 4



OVERFITTING

El sobreajuste ocurre cuando un modelo aprende demasiado bien los detalles y el ruido del conjunto de entrenamiento, pero no generaliza bien a nuevos datos.
Memoriza los datos.

		ACTUAL	
		CAT	NO CAT
PREDICTION	CAT	 TRUE POSITIVE	0
	NO CAT	0	 TRUE NEGATIVE

Educator in AI

**Artificial
Intelligence**

Data Engineering



Machine Learning

Data Science

📌 **Linkedin** —> <https://www.linkedin.com/in/erlinares/>

👋 **Follow us on X**: <https://x.com/erlinares>^[SEP]

💻 **GitHub**: https://github.com/erlinares/365_AI_Journey/

💬 **Discord**: <https://discord.gg/5fFM2zh8>



Edgar Rios Linares