

The background is a vibrant jungle scene. On the left, a black toucan with a large orange beak is perched on a branch. On the right, a red parrot is perched on a branch. The background is filled with various green leaves and palm trees. In the center, there is a large, light green circular area. Above this circle, there is a bright sun and two small birds flying in the sky.

# zoo

Machine Learning aplicado a la  
Clasificación de animales

Nerea Gómez



# OBJETIVO

Construir y optimizar un modelo que prediga correctamente el tipo de animal usando sus características.

# ÍNDICE

01

Tratamiento de datos,  
distribución y agrupación de  
las clases de animales

02

Features, target y matriz de  
correlaciones

03

Modelos de Machine Learning  
y optimización

04

Ejemplo del modelo final

05

Conclusiones y aprendizajes

# TRATAMIENTO DE DATOS

01

## **DATASET LIMPIO Y COMPACTO**

Pocas filas y columnas, sin valores faltantes

02

## **CARACTERÍSTICAS BINARIAS**

Todas representadas con 0/1 ( excepto número de patas)

03

## **COLUMNAS CATEGÓRICAS**

Tipo y nombre del animal, tratadas aparte.

04

## **AGRUPACIÓN CLASES MINORITARIAS**

Debido a un gran desbalance

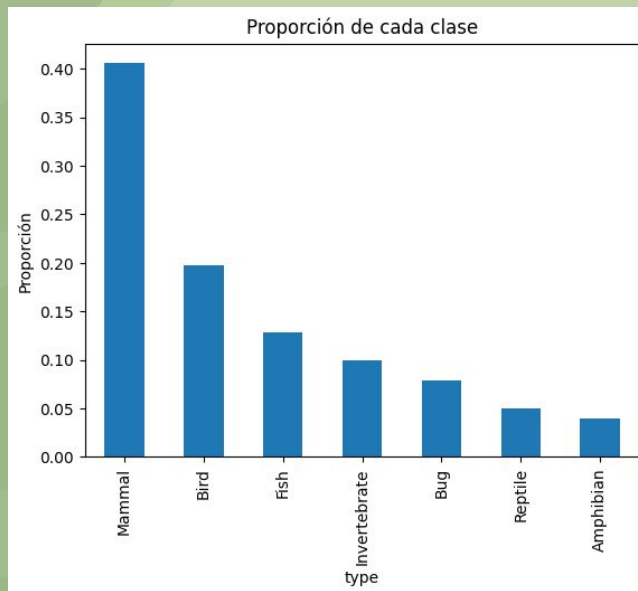
05

## **PREPARACIÓN PARA MACHINE LEARNING**

Dataset listo para entrenar modelos



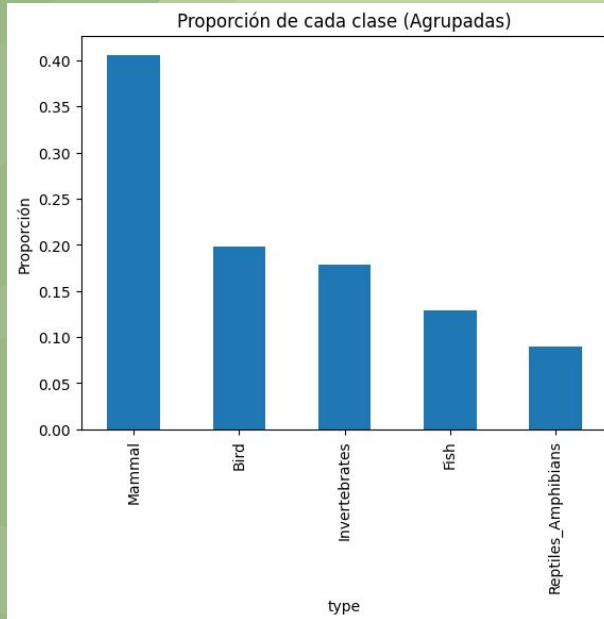
# DISTRIBUCIÓN Y PROPORCIÓN



El gráfico evidencia un **fuerte desbalance** entre las clases, destacando la gran mayoría de **mamíferos** frente al resto.



# ACCIÓN DE AGRUPAR LOS GRUPOS MINORITARIOS



**Grupos minoritarios** agrupados para reducir el desbalance en la distribución.

- **Invertebrados**, ahora tiene los grupos de invertebrados y bichos
- **Reptiles\_anfibios**, contienen ambas clases



# FEATURES Y TARGET

**FEATURES (X)**

**Características de los animales:** *pelo, plumas, huevos, leche, volador, acuático, depredador, dientes, columna vertebral, respira, venenoso, aletas, cola, doméstico, tamaño de gato y patas.*

**TARGET (Y)**

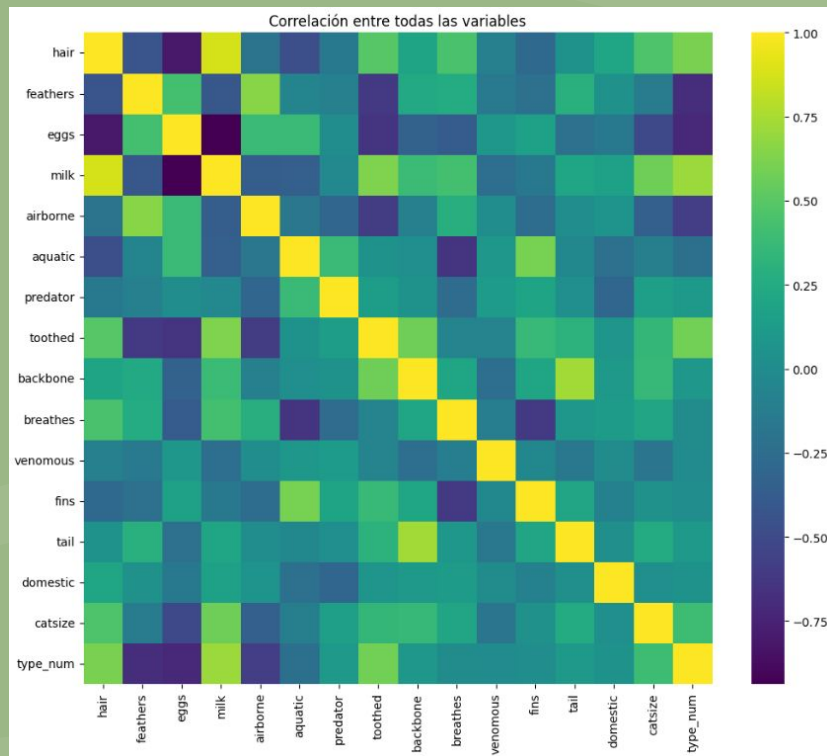
Variable a predecir, **Clase de Animal:**  
*Mamíferos, Ave, Pez, Invertebrados y Reptiles\_Anfibios*



# MATRIZ DE CORRELACIONES

type_num	1.000000
milk	0.711456
hair	0.606299
toothed	0.585993
catsize	0.397822
predator	0.111054
tail	0.102857
backbone	0.092515
domestic	0.048922
fins	0.001322
breathes	-0.003116
venomous	-0.014861
aquatic	-0.229000
airborne	-0.589256
feathers	-0.687023
eggs	-0.726156

Las características **más fuertemente correlacionadas** con el tipo de animal son leche, pelo, dientes y tamaño, mientras que huevos, plumas y vuelo son **correlaciones negativas** con el tipo.





# MODELOS DE MACHINE LEARNING

<u>MODELOS</u>	<u>ACCURACY</u>
KNN	0.33
ÁRBOL DE DECISIÓN	0.57
RANDOM FOREST	0.57
BAGGING	0.57
ADABOOST	0.61
GRADIENT BOOSTING	0.66
REGRESIÓN LOGÍSTICA	0.71



# OPTIMIZACIÓN DE LOS MEJORES MODELOS

Modelos seleccionados para optimización:

- Gradient Boosting
- Regresión Logística

Métodos de optimización aplicados:

- GridSearchCV
- RandomSearchCV

<u>MODELO</u>	<u>ACCURACY INICIAL</u>	<u>ACCURACY OPTIMIZADO</u>	<u>MEJOR MODELO</u>
Gradient Boosting	0.66	0.76	✓
Regresión Logística	0.71	0.56	✗



# EJEMPLO DE PREDICCIÓN CON EL MODELO FINAL

```
nuevo_prueba = pd.DataFrame([{\n    'hair':0,\n    'feathers':1,\n    'eggs':1,\n    'milk':0,\n    'airborne':1,\n    'aquatic':0,\n    'predator':0,\n    'toothed':0,\n    'backbone':1,\n    'breathes':1,\n    'venomous':0,\n    'fins':0,\n    'legs':2,\n    'tail':1,\n    'domestic':0,\n    'catsize':0\n}])\n\n# Respetar el orden exacto de X\nnuevo_prueba = nuevo_prueba[X.columns]\n\n# Escalado\nnuevo_prueba_scaled = scaler.transform(nuevo_prueba)\n\n# Predicción\npred_prueba = best_model.predict(nuevo_prueba_scaled)\npred_prueba[0]\n\n✓ 0.0s\n\n'Bird'
```

Predicción del tipo de animal a partir de sus características.

En este caso se trata de un Ave

# CONCLUSIONES

- **Limpieza y preparación del dataset:** esencial para un buen rendimiento del modelo.
- **Desbalance de clases:** la agrupación de clases minoritarias ayudó a mejorar la predicción.
- **Selección de features:** las características más influyentes fueron leche, pelo, huevos y plumas.
- **Optimización de modelos:** Gradient Boosting superó a Regresión Logística tras GridSearch/RandomSearch.
- **Rendimiento final:** el modelo alcanzó un accuracy de ~76%.
- **Aprendizajes y mejoras futuras:** más datos con más características y más animales, oversampling de clases minoritarias y visualización interactiva con Streamlit.



**¡¡MUCHAS GRACIAS!!**