

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

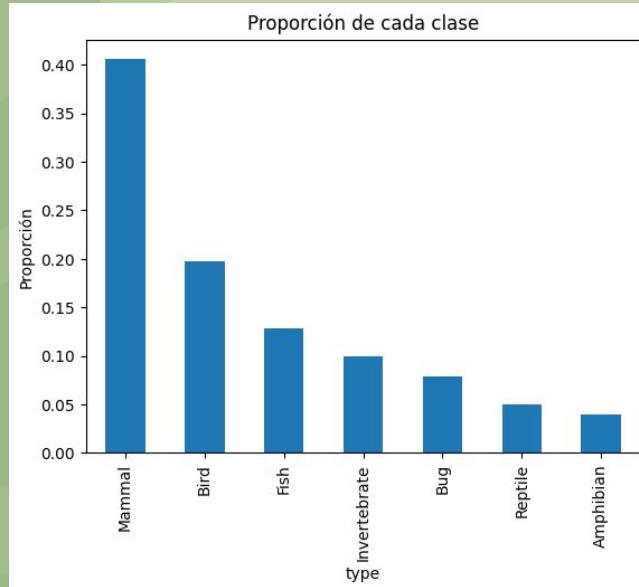
05

PREPARACIÓN PARA MACHINE LEARNING

Dataset listo para entrenar modelos

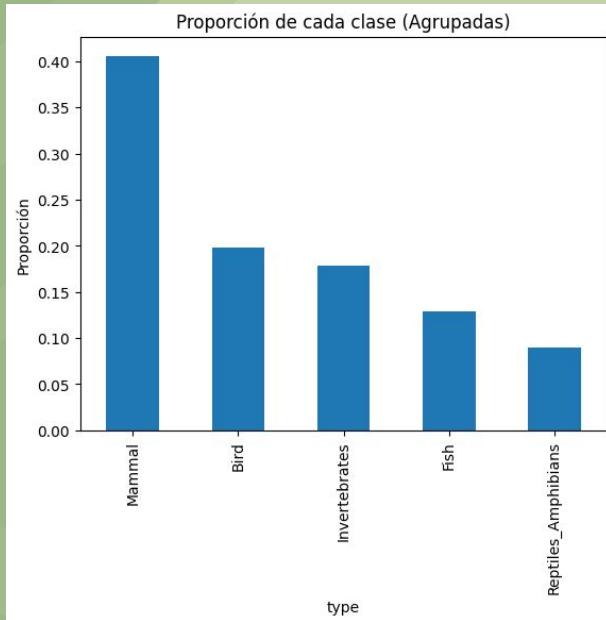


DISTRIBUCIÓN Y PROPORCIÓN



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)

TARGET (Y)

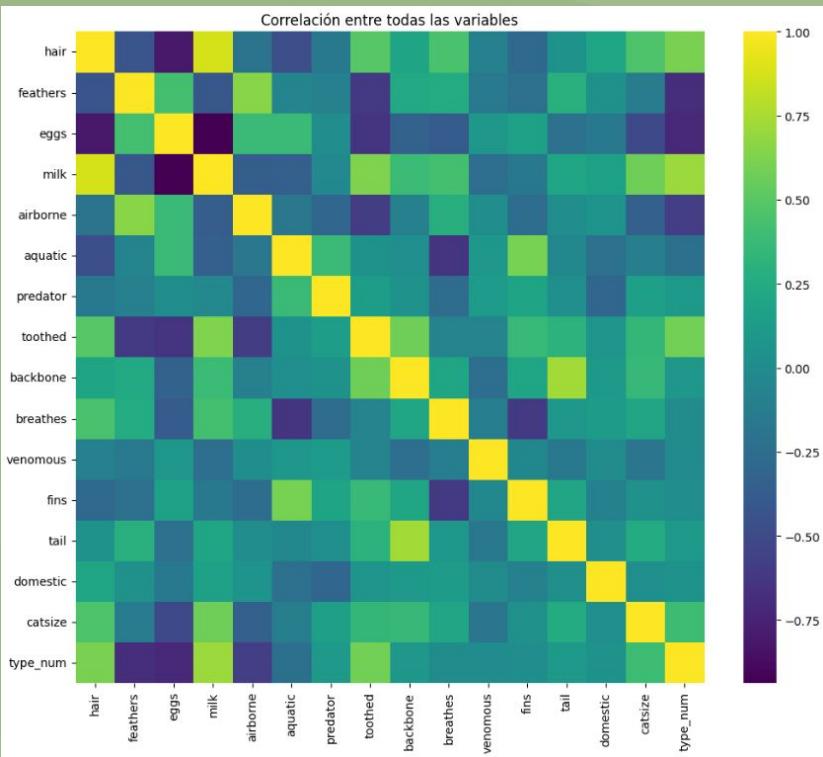
Características de los animales: pelo, plumas, huevos, leche, volador, acuático, depredador, dientes, columna vertebral, respira, venenoso, aletas, cola, doméstico, tamaño de gato y patas.

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	-0.687823
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([{
    'hair':0,
    'feathers':1,
    'eggs':1,
    'milk':0,
    'airborne':1,
    'aquatic':0,
    'predator':0,
    'toothed':0,
    'backbone':1,
    'breathes':1,
    'venomous':0,
    'fins':0,
    'legs':2,
    'tail':1,
    'domestic':0,
    'catsize':0
}])

# Respetar el orden exacto de X
nuevo_prueba = nuevo_prueba[X.columns]

# Escalado
nuevo_prueba_scaled = scaler.transform(nuevo_prueba)

# Predicción
pred_prueba = best_model.predict(nuevo_prueba_scaled)
pred_prueba[0]

✓ 0.0s
'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.

A cartoon illustration of a zebra's head and upper body on the left, looking towards the right. The background features a large, bright yellow sun in the upper right, casting a warm glow over a lush green jungle with various tropical plants and trees.

¡¡MUCHAS GRACIAS!!