

# Proyecto de Machine Learning para adopción de mascotas

Ana Isabel Patiño Osorio  
*Estudiante*  
Universidad de Antioquia  
Heliconia, Antioquia  
anai.patino@udea.edu.co

Cristian Daniel Muñoz Botero  
*Estudiante*  
Universidad de Antioquia  
La Ceja, Antioquia  
cristian.munoz3@udea.edu.co

**Abstract**—Este trabajo explora la aplicación de técnicas de Machine Learning para predecir la probabilidad de adopción de mascotas en refugios, utilizando un conjunto de datos que incluye características clave de cada animal. El objetivo de esta herramienta es asistir a las organizaciones protectoras de animales en la toma de decisiones informadas, enfocándose en promover la adopción de aquellos animales con menores probabilidades de ser adoptados.

**Index Terms**—Machine Learning, Adopción de mascotas, Clasificación binaria

## I. INTRODUCCIÓN

Cada año, miles de mascotas son acogidas por refugios, muchos de los cuales tienen recursos limitados para cuidar de ellas por tiempos prolongados. Identificar qué mascotas tienen más o menos probabilidades de ser adoptadas permite a estas instituciones enfocar mejor sus esfuerzos. Este trabajo propone el uso de un modelo de Machine Learning que analice los atributos de cada animal para estimar su probabilidad de adopción.

## II. DESCRIPCIÓN DEL PROBLEMA

La adopción de mascotas desempeña un papel fundamental en nuestra sociedad, ya que brinda una nueva oportunidad de vida a animales que han sido abandonados o rescatados de situaciones que comprometían su bienestar. Sin embargo, muchos refugios enfrentan serios desafíos, como la sobre población, lo que dificulta una atención adecuada y eleva los costos operativos. Además, la permanencia prolongada de los animales en estos centros puede afectar negativamente su calidad de vida.

Debido a lo anterior, el desarrollo de una solución basada en técnicas de Machine Learning resulta altamente valioso. Esta permitiría predecir la probabilidad de adopción de una mascota según sus características. Por ejemplo, verificar si una mascota vacunada es más propensa a ser adoptada que una mascota que no cuente con sus vacunas al día. Esto facilitaría la implementación de estrategias más efectivas de promoción permitiendo priorizar campañas dirigidas a aquellos animales con menor probabilidad de ser adoptados. De esta manera, se optimizan los recursos disponibles y se incrementan las posibilidades de que más animales encuentren un hogar.

La base de datos para predecir el estado de adopción

de mascotas contiene información detallada sobre animales disponibles para adopción en refugios. A continuación presentamos una breve descripción:

Numero de muestras: 2007

Numero de variables: 13

Variables:

- PetID: Identificador único de cada animal.
- PetType: Tipo de mascota (Perro, gato, etc).
- Breed: Raza específica de cada mascota.
- AgeMonths: Edad de la mascota en meses.
- Color: Color de la mascota.
- Size: Categoría del tamaño de la mascota (Pequeño, Mediano, Grande).
- WeightKg: Peso de la mascota en kilogramos.
- Vaccinated: Estado de vacunación de la mascota (0 – No vacunado, 1 – Vacunado).
- HealthCondition: Estado de salud de la mascota (0 - Saludable, 1 - Condición médica).
- TimeInShelterDays: Tiempo que la mascota ha estado en el refugio en días.
- AdoptionFee: Tarifa de adopción que se cobra por la mascota en dólares.
- PreviousOwner: Si la mascota tuvo un dueño anterior (0 - No, 1 - Sí).
- AdoptionLikelihood: Probabilidad de que la mascota sea adoptada (0 - Poco probable, 1 - Probable).

No hay existencia de datos faltantes

Tipo de codificación usado para las diferentes variables del problema:

Variable	Tipo	Codificación
PetType	Categorico nominal	One hot encoding
Breed	Categorico nominal	Label encoding
AgeMonths	Numérico	Normalización
Color	Categorico nominal	One hot encoding

Variable	Tipo	Codificación
Size	Categorico ordinal	Ordinal encoding
WeightKg	Numérico	Normalización
Vaccinated	Numérico binario	—
HealthCondition	Numérico binario	—
TimeInShelter-Days	Numérico	Normalización
AdoptionFee	Numérico	Normalización
PreviousOwner	Numérico binario	—
AdoptionLike-lihood	Numérico binario	—

Se optó por utilizar el paradigma de aprendizaje supervisado, dado que se cuenta con un conjunto de datos etiquetado. En particular, se empleó una metodología de clasificación binaria, ya que la variable objetivo, «Probabilidad de adopción», toma únicamente dos valores posibles: Poco probable (0) y Probable (1).

### III. ESTADO DEL ARTE

#### A. Sistema de recomendación para adopción de animales PetMatch [1]

Este trabajo realizado por Nathalie Estrella utiliza un modelo híbrido de recomendación basado en contenido y filtrado colaborativo empleando técnicas como Random Forest para la clasificación y una red neuronal convolucional (CNN) para el análisis de las características visuales de las mascotas.

La validación del sistema se realizó mediante pruebas de precisión (accuracy) y recall, comparando las recomendaciones generadas con las elecciones reales de los adoptantes.. Una métrica adicional utilizada fue la similitud coseno, la cual permite medir el grado de similitud entre un adoptante y una mascota comparando sus vectores de características. La similitud coseno mide el ángulo entre dos vectores en un espacio multidimensional, y se define matemáticamente como:

$$\text{Sim}(A,B) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

Un valor cercano a 1 indica alta similitud, mientras que valores cercanos a 0 indican poca o ninguna similitud.

En cuanto a los resultados, el primer modelo implementado fue K-Nearest Neighbors (KNN), que obtuvo una precisión inicial del 33.31%. Tras ajustes se logró mejorar esta precisión hasta un 43.05%, alcanzando además una similitud promedio de 0.78, utilizando valores de k entre 5 y 10. Posteriormente, se probó el modelo Random Forest, que incrementó la precisión a 44.91% y alcanzó una similitud promedio de 0.99. Finalmente, el modelo seleccionado como solución definitiva

fue XGBoost. En su configuración inicial obtuvo una precisión de 46.82%, pero tras realizar ajustes, se logró alcanzar una precisión del 94.46%, lo que representa una mejora significativa frente a los modelos anteriores. Además, el sistema logró un recall del 92.37%, lo que indica una alta capacidad para identificar recomendaciones relevantes.

#### B. Increasing adoption rates at animal shelters: a two-phase approach to predict length of stay and optimal shelter allocation [2]

En este trabajo, desarrollado por Janae Bradley y Suchithra Rajendran, se usó un paradigma de aprendizaje supervisado, tratando de predecir la duración de la estancia de los animales en el refugio según varias características: el tipo de animal, edad, género, raza, tamaño y ubicación del refugio.

Se usaron como técnicas de aprendizaje la regresión logística, GBoost, Random Forest y redes neuronales artificiales. Para el entrenamiento y validación de los mismos no se menciona dentro del texto una metodología de validación explícita. Simplemente se menciona que se dividen los datos en datos de entrenamiento y test y las métricas que se usaron para evaluar dichos modelos fueron: Recall, precisión y  $F_\beta$  score, con  $\beta = 1$ .

Generalmente, se pudo ver que para predecir tiempos cortos de estancia, aunque se tuvo un rendimiento bajo en todos los modelos, el que más destacó fue el Random Forest; para estancias medias todos los modelos tuvieron un bajo rendimiento; para estancias largas el Gradient Boosting mostró un excelente desempeño lo mismo que para estancias muy largas, en los que se alcanzaron métricas entre 70% y 80% tanto para Gradient Boosting como para Random Forest.

En promedio, el algoritmo de Gradient Boosting obtuvo un desempeño cercano al 60% en todas las métricas, mientras que la regresión logística fue el peor. Por último, se hizo un análisis de características en el cual se pudo observar que la edad avanzada, el tamaño de la mascota y el color fueron factores clave en la predicción de la duración de la estancia. Las demás variables no tuvieron un impacto significativo.

### REFERENCES

- [1] N. B. E. Mejía, «Sistema de recomendación para adopción de animales “PetMatch”», *Tesis de licenciatura*, 2025.
- [2] S. R. Janae Bradley, «Increasing adoption rates at animal shelters: a two-phase approach to predict length of stay and optimal shelter allocation», *BMC Veterinary Research*, 2021.