

# Test Design (Primera iteración)

En esta fase se define cómo se medirá la generalización del modelo y se especifica el particionado de datos. Se establecen las métricas primarias y secundarias en unidades operativas y se fijan umbrales de aceptación operativa que conectan el rendimiento técnico con decisiones veterinarias.

El resultado esperado es un plan de experimentos claro que evite fugas de información y permita comparar modelos de forma justa

## 2.1. División de datos

Para garantizar una evaluación robusta, entrenar bien al modelo y evitar overfitting, se implementó una estrategia de división de datos que incluye múltiples niveles de validación:

### 1. Split Initial Training-Testing:

La división inicial de 80-20% se fundamenta en principios estadísticos y prácticos de machine learning:

#### Training Set (80% ~3,200 registros):

- **Propósito principal:** Entrenamiento de modelos y ajuste fino de hiper parámetros
- **Composición:** Incluye toda la variabilidad del dataset para capturar patrones complejos
- **Ventaja:** Proporciona suficiente data para que modelos complejos como Random Forest aprendan relaciones no lineales sin overfitting excesivo

#### Test Set (20% ~800 registros):

- **Función crítica:** Simular condiciones de producción con datos completamente nuevos
- **Preservación de integridad:** Nunca se utiliza durante desarrollo, garantizando evaluación imparcial
- **Tamaño adecuado:** 800 registros proporcionan significancia estadística para métricas de evaluación

## 2.2. Cross-Validation con K-Folds:

La implementación de K-Fold con 3 splits representa un balance óptimo entre robustez y costo computacional.

Python

```
# K-Fold para evaluación durante desarrollo
kf = KFold(n_splits=3, shuffle=True, random_state=42)
```

El modelo se entrena K veces, cada vez utilizando K-1 folds para entrenamiento y el fold restante para validación. Finalmente, se promedian los resultados de las “K” iteraciones para obtener una estimación robusta del rendimiento del modelo.

- Test set reservado exclusivamente para evaluación final
- No utilizado en ningún proceso de entrenamiento o ajuste