

Test Design (Tercera iteración)

2.1. División de datos

Parámetro	Valor	Justificación
Train	70%	Suficiente para aprendizaje sin comprometer validación
Test	30%	Representativo para evaluación final
Estratificación	Sí	Mantener proporción original de clases
Shuffle	Sí (seed = 42)	Mezclar datos para evitar sesgos temporales

Python

```
# Estrategia de división
X = df_balanced.drop(columns=['Estado de Producción', 'Número del animal'])
y = df_balanced['Estado de Producción']

X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.3,
    random_state=42,
    stratify=y # Mantener balance entre clases
)
```

2.2. Estrategia de evaluación

Métrica	Importancia por Modelo
Accuracy	Baseline para todos los modelos
Precisión por clase	Crítico para "En Monitoreo" (evitar secado prematuro)
Recall por clase	Crítico para "Previo a Secado"
F1-Score	Balance entre precisión y recall
Matriz de Confusión	Análisis de errores por etiqueta

2.3. Plan de experimentación

El plan de experimentación está sustentado en:

- Análisis exploratorio del dataset original
- Data Augmentation para evitar desbalances entre clases
- Modelado y selección de hiperparametros específicos por modelo
- Realización de predicciones

2.4. Validación durante el entrenamiento

Para la validación de que los entrenamientos están teniendo un buen desempeño es con el uso de una división del training set. Esto con el objetivo de que en cada de las iteraciones de los entrenamientos se pueda hacer la validación de que el modelo se entrena de manera adecuada sin llegar al overfitting.

2.5. Criterios de éxito por modelo

Modelo	Accuracy Mínima	Precisión "En Monitoreo"	Recall "Previo a Secado"	Criterios Adicionales
Regresión Logística	≥ 85%	≥ 80%	≥ 75%	Baseline estable y punto de partida para modelos más complejos
Random Forest	≥ 88%	≥ 85%	≥ 80%	Interpretabilidad clara Overfitting
XGBoost	≥ 90%	≥ 88%	≥ 85%	Interpretabilidad clara Overfitting
MLP	≥ 91%	≥ 90%	≥ 88%	Interpretabilidad clara y no Overfitting
TabNet	≥ 85%	≥ 80%	≥ 85%	Interpretabilidad clara Overfitting al ser una arquitectura de datos tabulares