

Reporte de Regresión Logística - Breast Cancer

Información del Dataset

Dataset utilizado: Breast Cancer Wisconsin (Diagnostic)

Este dataset contiene información de tumores malignos y benignos, con variables que describen características de las células obtenidas mediante imágenes.

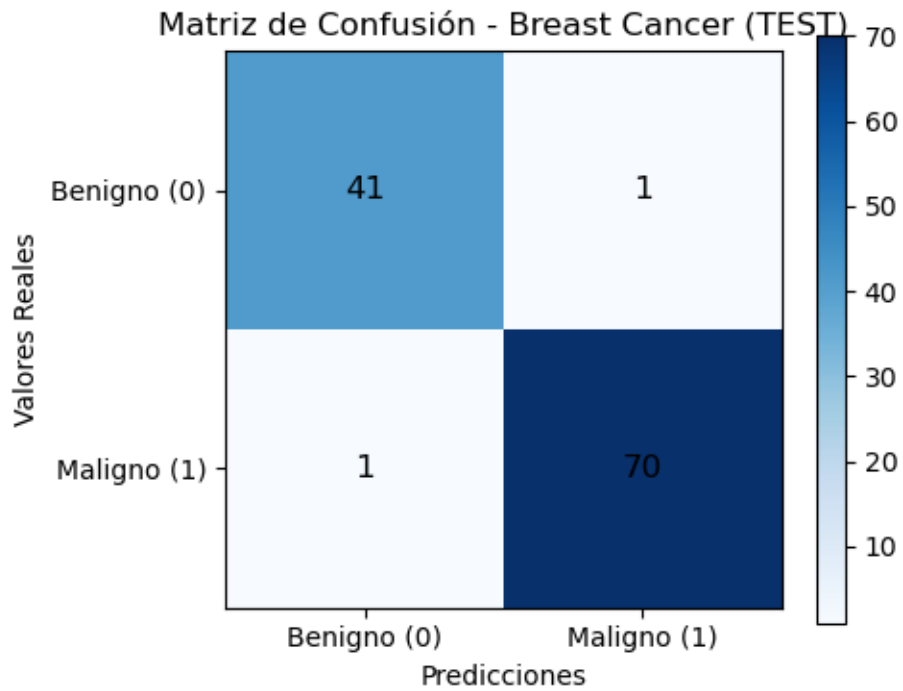
Detalles del dataset:

- Total de muestras: 569
- Número de características: 30
- Tamaño de entrenamiento: 455
- Tamaño de prueba: 114

Evaluación del Modelo

A continuación se muestran los resultados obtenidos en la fase de prueba, incluyendo la matriz de confusión y métricas clave para evaluar el desempeño del modelo.

Matriz de Confusión con seed 42



Métricas con Seed 42

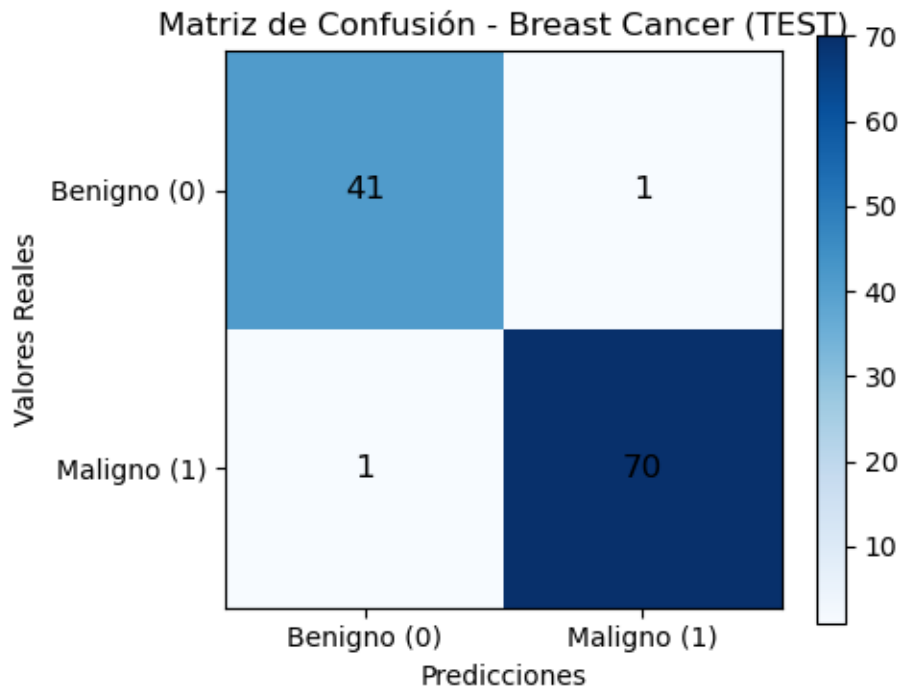
Accuracy: 0.982

Precision: 0.986

Recall: 0.996

F1-score: 0.986

Matriz de Confusión con seed 1



Métricas con Seed 1

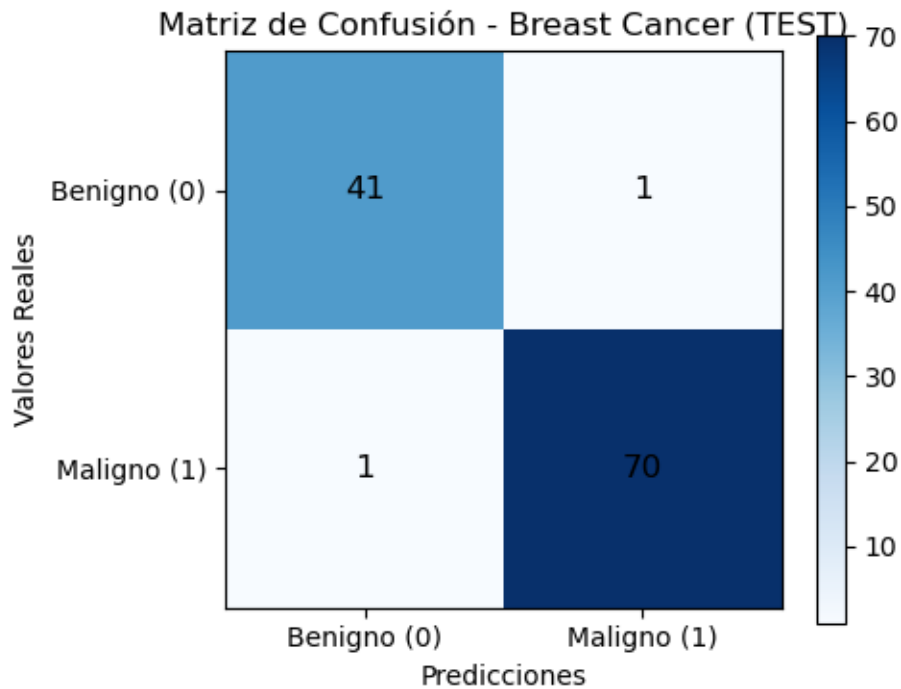
Accuracy: 0.982

Precision: 0.986

Recall: 0.996

F1-score: 0.986

Matriz de Confusión con seed 10



Métricas con Seed 10

Accuracy: 0.982

Precision: 0.986

Recall: 0.996

F1-score: 0.986

Análisis y Conclusión

Interpretación:

El alto valor de recall indica que el modelo identifica correctamente la mayoría de los casos malignos, lo cual es esencial en problemas médicos donde se debe minimizar el riesgo de no detectar un tumor cancerígeno. La precisión elevada confirma que la mayoría de las predicciones de tumores malignos son correctas, evitando falsas alarmas innecesarias. El F1-score balanceado demuestra que existe un buen equilibrio entre precisión y recall.

Conclusión:

El modelo de regresión logística mostró un excelente desempeño en la clasificación de tumores, logrando predecir de manera precisa y confiable. Esto lo convierte en una herramienta útil como apoyo para la toma de decisiones en diagnóstico médico. Sin embargo, se recomienda en futuras versiones:

- Aplicar regularización L2 para reducir sobreajuste y mejorar la generalización.
- Ajustar el umbral de clasificación, priorizando el recall si se busca detectar todos los casos malignos, o la precision si se desea minimizar falsos positivos.
- Probar técnicas más avanzadas como redes neuronales o árboles de decisión para comparar el desempeño.

Estos resultados demuestran que la regresión logística es una técnica simple pero poderosa para problemas de clasificación médica, siendo un primer paso sólido hacia sistemas de diagnóstico asistido