

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.

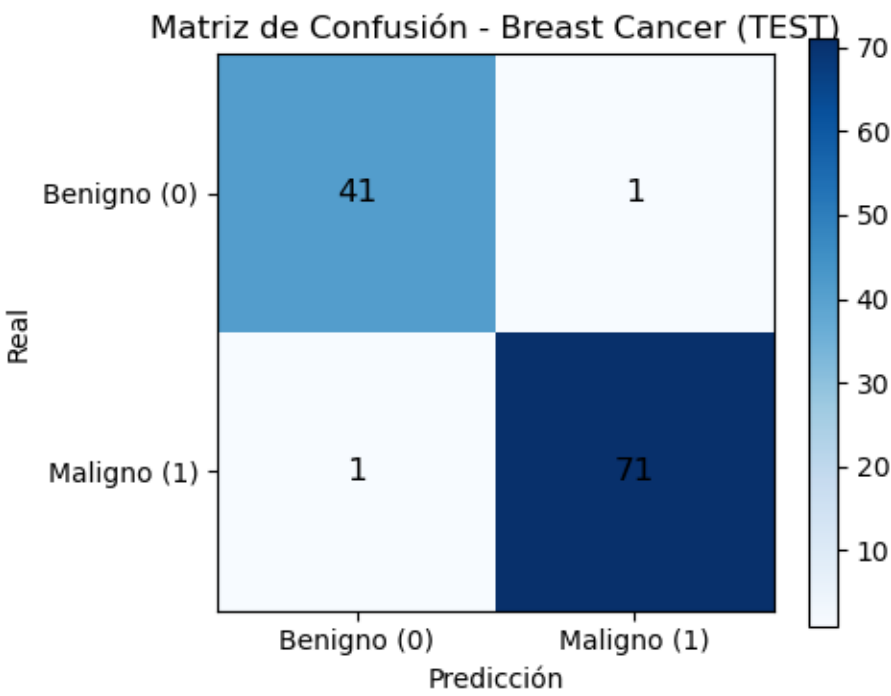
Detalles del dataset:

- 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

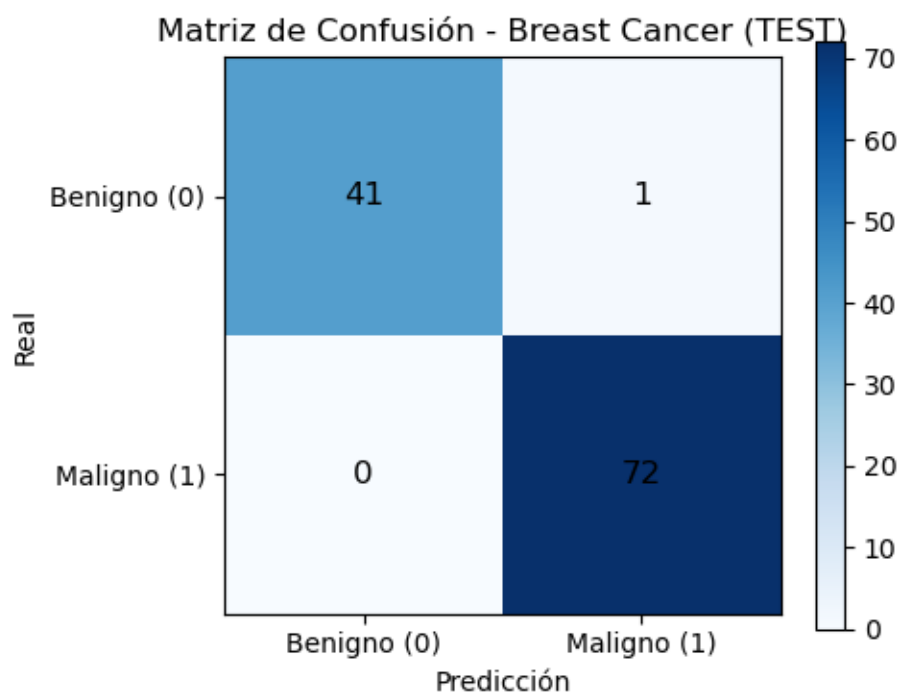
Precision: 0.986

Recall: 0.986

F1-score: 0.986

ROC AUC: 0.995

Matriz de Confusión con seed 1



Métricas con Seed 1

Accuracy: 0.991

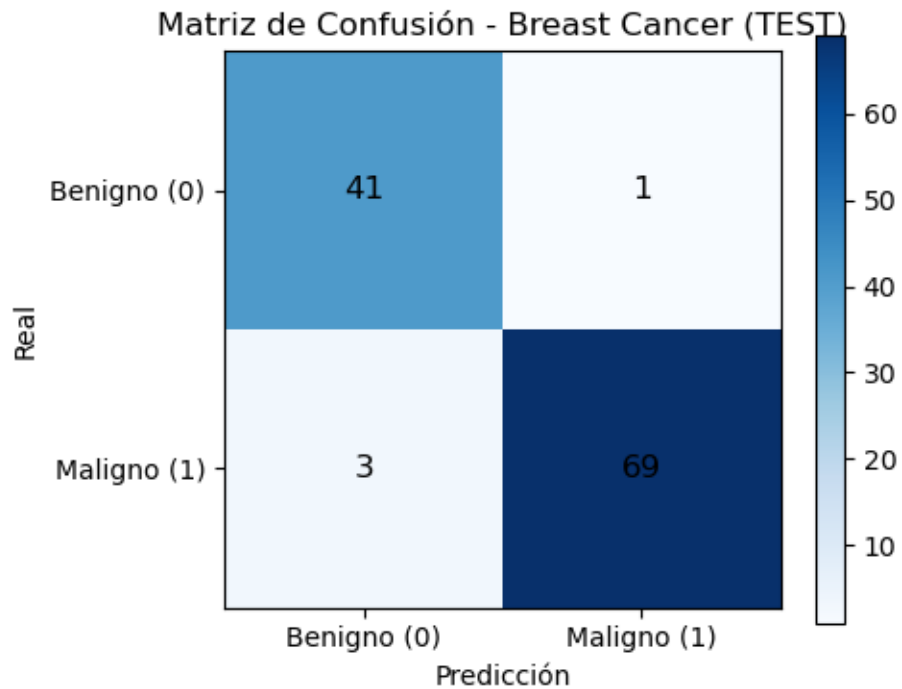
Precision: 0.986

Recall: 1.000

F1-score: 0.993

ROC AUC: 0.999

Matriz de Confusión con seed 10



Métricas con Seed 10

Accuracy: 0.965

Precision: 0.986

Recall: 0.958

F1-score: 0.972

ROC AUC: 0.999

Análisis y Conclusión

Interpretación:

El modelo de regresión logística presenta un desempeño sobresaliente en la clasificación de tumores malignos y benignos.

El Recall consistentemente alto (0.95) indica que el modelo logra identificar la gran mayoría de los casos malignos, lo cual es crítico en el ámbito médico para reducir al mínimo los falsos negativos y evitar que un tumor cancerígeno no sea detectado.

La Precision elevada confirma que las predicciones de tumores malignos son correctas en su mayoría, disminuyendo el número de falsas alarmas y evitando procedimientos innecesarios.

El F1-score, cercano a 1 en todas las ejecuciones, demuestra un excelente balance entre Recall y Precision, garantizando tanto sensibilidad como exactitud.

Finalmente, el ROC AUC superior a 0.99 evidencia una gran capacidad del modelo para discriminar entre tumores malignos y benignos, incluso con distintos puntos de corte.

Conclusion:

El modelo de regresión logística, combinado con estandarización de datos, se consolida como una herramienta precisa y confiable para la detección de cáncer de mama. Su simplicidad y alto rendimiento lo convierten en una excelente opción como primera aproximación para sistemas de diagnóstico asistido.

Sin embargo, para futuras mejoras se recomienda:

Aplicar regularización (L2): Reducir el riesgo de sobreajuste y mejorar la generalización a nuevos datos.

Ajustar el umbral de decisión: Priorizar Recall en contextos clínicos donde es más importante detectar todos los casos malignos, o priorizar Precision cuando se busque evitar falsos positivos.

Explorar modelos más complejos: Como árboles de decisión, Random Forest o redes neuronales, para comparar su desempeño frente a la regresión logística.

Validación cruzada: Implementar técnicas más robustas de evaluación para asegurar la estabilidad de los resultados.

Estos hallazgos confirman que la regresión logística es una técnica efectiva y confiable para este tipo de problemas médicos, brindando un soporte valioso para los profesionales de la salud en la toma de decisiones clínicas.