

```
In [6]: # Importación de librerías necesarias
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, roc_auc_score, confusion

import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load_breast_cancer

# Cargar un conjunto de datos de ejemplo
data = load_breast_cancer()
X = data.data
y = data.target

# Dividir los datos en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Métricas de Evaluación de Modelos

Las métricas de evaluación como precisión, recall, y F1-score son esenciales para entender diferentes aspectos del rendimiento del modelo. Aquí demostramos cómo calcular estas métricas.

```
In [7]: model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Calcular métricas
accuracy = accuracy_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
roc_auc = roc_auc_score(y_test, y_pred)

print(f"Accuracy: {accuracy:.2f}")
print(f"Recall: {recall:.2f}")
print(f"Precision: {precision:.2f}")
print(f"F1 Score: {f1:.2f}")
print(f"ROC AUC Score: {roc_auc:.2f}")
```

```
Accuracy: 0.96
Recall: 0.99
Precision: 0.96
F1 Score: 0.97
ROC AUC Score: 0.96
```

Estas métricas indican un alto desempeño del modelo en el conjunto de prueba, reflejando su capacidad para predecir correctamente tanto las clases positivas como las negativas.

Explicación de las Métricas de Evaluación y sus Resultados

En el contexto de la validación y evaluación de modelos de machine learning, es fundamental entender qué mide cada métrica y cómo interpretar los resultados obtenidos. A continuación, se presentan las definiciones y el análisis de las métricas de evaluación: Accuracy, Recall, Precision, F1 Score y ROC AUC Score, junto con sus resultados específicos.

Accuracy (Precisión)

Definición: La precisión es la proporción de predicciones correctas realizadas por el modelo sobre el total de predicciones. Es una métrica global que indica qué tan bien está funcionando el modelo en general.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Donde:

- TP: Verdaderos Positivos
- TN: Verdaderos Negativos
- FP: Falsos Positivos
- FN: Falsos Negativos

Resultado: 0.96

- **Interpretación:** El modelo predice correctamente el 96% de los casos. Esto indica un buen rendimiento general, pero no distingue entre las diferentes clases de errores (falsos positivos y falsos negativos).

Recall (Sensibilidad o Tasa de Verdaderos Positivos)

Definición: El recall mide la capacidad del modelo para identificar todas las instancias positivas. Es especialmente importante en contextos donde no detectar una instancia positiva tiene un alto costo.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Resultado: 0.99

- **Interpretación:** El modelo identifica correctamente el 99% de las instancias positivas. Este alto valor de recall es crucial en situaciones como el diagnóstico médico, donde es importante no pasar por alto ningún caso positivo.

Precision (Precisión o Valor Predictivo Positivo)

Definición: La precisión mide la exactitud de las predicciones positivas del modelo. Indica la proporción de verdaderos positivos sobre todas las predicciones positivas.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Resultado: 0.96

- **Interpretación:** El 96% de las instancias que el modelo predice como positivas son realmente positivas. Una alta precisión es importante en casos donde los falsos positivos son costosos.

F1 Score

Definición: El F1 Score es la media armónica de la precisión y el recall. Proporciona una única métrica que equilibra la precisión y el recall, especialmente útil cuando se necesita un balance entre ambas.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Resultado: 0.97

- **Interpretación:** El F1 Score de 0.97 indica un equilibrio casi perfecto entre precisión y recall, mostrando que el modelo es muy bueno tanto en identificar instancias positivas como en evitar falsos positivos.

ROC AUC Score (Área bajo la curva ROC)

Definición: El ROC AUC Score mide la capacidad del modelo para distinguir entre clases. La curva ROC traza la tasa de verdaderos positivos frente a la tasa de falsos positivos en varios umbrales de clasificación.

$$\text{AUC} = \int_{\text{ROC}} d(\text{ROC})$$

Resultado: 0.96

- **Interpretación:** Un ROC AUC Score de 0.96 significa que el modelo tiene una excelente capacidad para distinguir entre clases positivas y negativas. Cuanto más cercano a 1, mejor es la discriminación.

Conclusión

Las métricas obtenidas reflejan un modelo que no solo es preciso en sus predicciones generales (accuracy), sino que también es extremadamente eficaz en identificar correctamente las instancias positivas (recall) y en asegurar que las predicciones positivas sean verdaderamente positivas (precision). El F1 Score y el ROC AUC Score complementan esta evaluación al mostrar un excelente balance y capacidad de discriminación, respectivamente. Estos resultados indican que el modelo está bien ajustado y es robusto para

In []:

