**INTEGRANTES:** Castillo Elian, Muñoz Jhon, Niebles Yurliana, Quiñonez Jose

Tabla

El contenido generado por IA puede ser incorrecto.

1. Indique la cantidad de datos muestrales, verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Interprete estos valores.

R=hay 100 datos muestrales, 50 Verdaderos Positivos, 17 Falsos positivos, 23 Falsos Negativos y 10 Verdaderos Negativos. El clasificador tiene buen desempeño en la detección de casos positivos, pero todavía presenta un número considerable de falsos positivos y falsos negativos que limitan su confiabilidad.

1. Calcule las métricas para evaluar la calidad de los resultados. ¿Qué significa el resultado obtenido en cada métrica?

|  |  |
| --- | --- |
| Metrica | Resultado |
| Exactitud (ACC) |  |
| Sensibilidad (SEN) |  |
| Especificidad (SPE) |  |
| Valor predictivo positivo (TPV) |  |
| Valor predictivo negativo (NPV) |  |
| Tasa de descubrimiento falso (FDR) |  |
| Tasa de falsos negativos |  |
| Tasa de falsos positivos |  |
| Indice de elevación (Lift) |  |
| F1 score |  |

**1. Exactitud (ACC = 0.6)**

Proporción de predicciones correctas (VP + VN) sobre el total de muestras.

* **Interpretación**: El modelo acierta el 60% de las veces, esta puede ser engañosa si la clase está desbalanceada (muchos más positivos que negativos o viceversa).

**2. Sensibilidad / Recall (SEN = 0.68)**

Mide la capacidad de detectar los positivos reales:

* **Interpretación**: El modelo detecta correctamente el 68% de los casos positivos, es útil cuando es importante no dejar escapar positivos (ejemplo: diagnóstico de una enfermedad).

**3. Especificidad (SPE = 0.37)**

Proporción de negativos correctamente identificados:

* **Interpretación**: Solo el 37% de los negativos reales fueron reconocidos como negativos.

**4. Valor Predictivo Positivo (TPV = 0.793)**

Probabilidad de que un positivo predicho sea realmente positivo:

* **Interpretación**: El 79.3% de las veces que el modelo predice “positivo”, acierta, es útil cuando interesa la confianza en los positivos detectados.

**5. Valor Predictivo Negativo (NPV = 0.303)**

Probabilidad de que un negativo predicho sea realmente negativo:

* **Interpretación**: Solo el 30.3% de los negativos predichos son correctos, es decir el modelo falla mucho al confirmar que algo es negativo.

**6. Tasa de Descubrimiento Falso (FDR = 0.253)**

Proporción de positivos predichos que en realidad son falsos positivos:

* **Interpretación**: El 25.3% de las predicciones positivas fueron equivocadas.

**7. Tasa de Falsos Negativos (FNR = 0.315)**

Proporción de positivos reales que fueron clasificados como negativos:

* **Interpretación**: El modelo deja escapar el 31.5% de los positivos reales.

**8. Tasa de Falsos Positivos (FPR = 0.629)**

Proporción de negativos reales que fueron clasificados como positivos:

* **Interpretación**: El modelo confunde un 62.9% de los negativos con positivos.

**9. Índice de Elevación (Lift = 1.022)**

Compara la efectividad del modelo con un clasificador aleatorio:

* **Interpretación**: Con un valor cercano a 1, el modelo apenas es un poco mejor que adivinar al azar.

**10. F1 Score (0.714)**

Media armónica entre precisión (TPV) y recall (SEN):

* **Interpretación**: Balancea la calidad entre “detectar positivos” y “que esos positivos predichos sean correctos”, con **0.714** el modelo tiene un rendimiento aceptable en positivos, pero podría mejorar bastante.

1. Teniendo en cuenta las métricas obtenidas, describa la calidad de los resultados indicando si son confiables. ¿Qué se debe mejorar?

El modelo no es totalmente confiable, ya que aunque reconoce la mayoría de los positivos, confunde demasiados negativos.

* **Sensibilidad (SEN = 0.68):** El modelo detecta el 68% de los casos positivos reales, lo cual es importante en contextos médicos, ya que no se deben pasar por alto tumores malignos.
* **Valor Predictivo Positivo (TPV = 0.793):** Cuando predice un caso positivo, acierta en el 79.3% de las veces, lo que da confianza en los diagnósticos positivos.
* **Especificidad (SPE = 0.37):** Solo el 37% de los negativos fueron identificados correctamente, evidenciando un bajo desempeño en la detección de casos sanos.
* **Tasa de Falsos Positivos (FPR = 0.629):** El 62.9% de los negativos fueron clasificados como positivos, generando un número excesivo de falsos diagnósticos.
* **Tasa de Falsos Negativos (FNR = 0.315):** El modelo deja escapar un 31.5% de los positivos reales, lo que representa un riesgo clínico considerable.
* **Índice F1 (0.714):** Muestra un rendimiento aceptable en el balance entre precisión y sensibilidad, aunque con margen de mejora.
* **Índice de Elevación (Lift = 1.022):** Con un valor cercano a 1, el modelo apenas supera a una clasificación aleatoria.

**¿Qué se debe mejorar?**

Se recomienda mejorar reduciendo falsos positivos y negativos mediante el ajuste del umbral de decisión, técnicas de balanceo de datos y el uso de algoritmos más robustos.