

# Técnicas Avanzadas de la Clasificación de Patrones

*Casos extremos y patológicos de las medidas de desempeño*

Semestre B25

---

Dr. Cornelio Yáñez Márquez

Gerardo Acevedo Sánchez

Jorge Alberto Pacheco Senard

Alejandro Monroy Azpeitia

Carlos Jair Ramos González

César Augusto Pilon Alcalá

13 de Septiembre de 2025

## Índice

<b>1</b>	<b>Parte 1 - Dataset</b>	<b>1</b>
<b>2</b>	<b>Parte 2 - Diversos Modelos</b>	<b>1</b>
<b>3</b>	<b>Parte 3 - Matrices de confusión</b>	<b>2</b>
3.1	1NN	2
3.2	3NN	3
3.3	5NN	4
3.4	SVM	5
3.5	Naive Bayes	6
3.6	C5.0	7
3.7	RF	8
3.8	XGBoost	9
<b>4</b>	<b>Parte 4 - Medidas de desempeño</b>	<b>10</b>

## Índice de cuadros

1	Métricas de modelos con clase positiva 0	10
2	Métricas de modelos con clase positiva 1	10

## Índice de figuras

1	Matriz de confusión del modelo 1NN, con la clase 0 como 'Positive'.	2
2	Matriz de confusión del modelo 1NN, con la clase 1 como 'Positive'.	2
3	Matriz de confusión del modelo 3NN, con la clase 0 como 'Positive'.	3
4	Matriz de confusión del modelo 3NN, con la clase 1 como 'Positive'.	3
5	Matriz de confusión del modelo 5NN, con la clase 0 como 'Positive'.	4
6	Matriz de confusión del modelo 5NN, con la clase 1 como 'Positive'.	4
7	Matriz de confusión del modelo SVM, con la clase 0 como 'Positive'.	5
8	Matriz de confusión del modelo SVM, con la clase 1 como 'Positive'.	5
9	Matriz de confusión del modelo Naive Bayes, con la clase 0 como 'Positive'.	6
10	Matriz de confusión del modelo Naive Bayes, con la clase 1 como 'Positive'.	6
11	Matriz de confusión del modelo C5, con la clase 0 como 'Positive'.	7
12	Matriz de confusión del modelo C5, con la clase 1 como 'Positive'.	7
13	Matriz de confusión del modelo RF, con la clase 0 como 'Positive'.	8
14	Matriz de confusión del modelo RF, con la clase 1 como 'Positive'.	8
15	Matriz de confusión del modelo XGBOOST, con la clase 0 como 'Positive'.	9
16	Matriz de confusión del modelo XGBOST, con la clase 1 como 'Positive'.	9

# Propósito de la tarea

Analizar los casos extremos y patológicos de las medidas de desempeño derivadas de una matriz de confusión.

## 1. Parte 1 - Dataset

Utiliza la partición fija del dataset asignado por equipo. Los dataset elegidos son biclase, de atributos numéricos y depurados.

Para poder trabajar comodamente el dataset se hizo un mapeo de los valores de las etiquetas, pasando las clases con valor 1 a 0, y las clases con valor 2 a 1. De esta manera se continuó con el análisis encontrando que existe un desbalance en las clases resultantes. Mediante la aplicación de una función para calcular el valor de 'Imbalance Ratio' se obtuvo un valor de 2.38, lo que significa que aunque se haga el calculo de las medidas de desempeño como Accuracy o Error Rate, estos no van a ser representativos para el modelo.

Haciendo una observación en los valores de las características que tenía cada vector, se determinó conveniente hacer una estandarización de dichos valores ya que se hace una implementación de algoritmos basados en distancias y utilizar los vectores si una respectiva estandarización puede tener un efecto negativo para la tarea de clasificación.

Este es el único preprocesamiento realizado en el dataset proporcionado, ya que este no contenía valores nulos, repetidos o indistinguibles para ser descartados.

## 2. Parte 2 - Diversos Modelos

Con el set Train entrenar los clasificadores:

- 1NN
- 3NN
- 5NN
- SVM
- Naive Bayes
- C5.0
- RF
- XGBoost

Usa los patrones del set Test para probar la fase de entrenamiento de cada clasificador. Considera cada clase como 'positive'. Es decir, en la primera iteración del Train-Test usa la primera clase como 'positive'. En la segunda iteración usa la segunda clase como 'positive'.

### 3. Parte 3 - Matrices de confusión

Reporta gráficamente la 'Matriz de Confusión' para cada iteración generada en la parte 2. Es decir, dos matrices de confusión por dataset y clasificador.

#### 3.1. 1NN

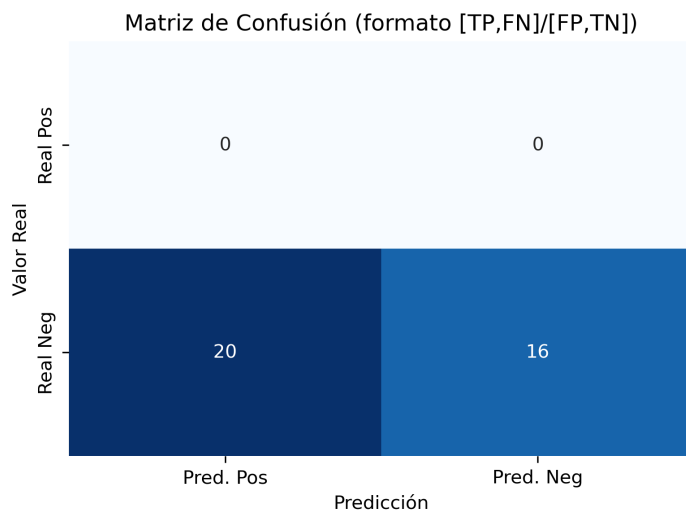


Figura 1: Matriz de confusión del modelo 1NN, con la clase 0 como 'Positive'.

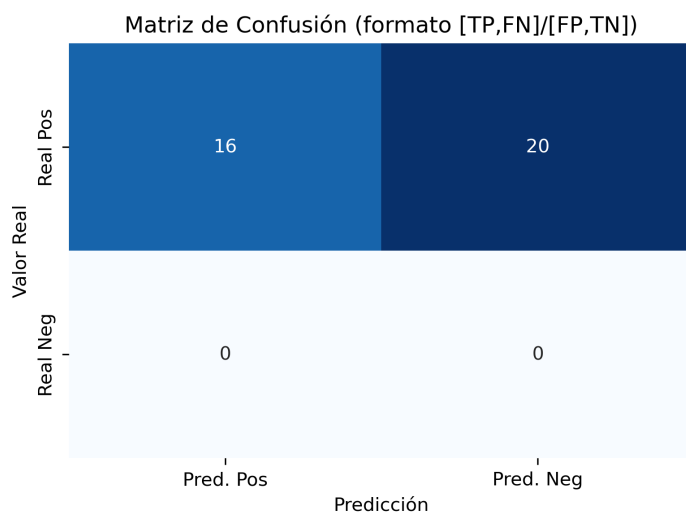


Figura 2: Matriz de confusión del modelo 1NN, con la clase 1 como 'Positive'.

### 3.2. 3NN

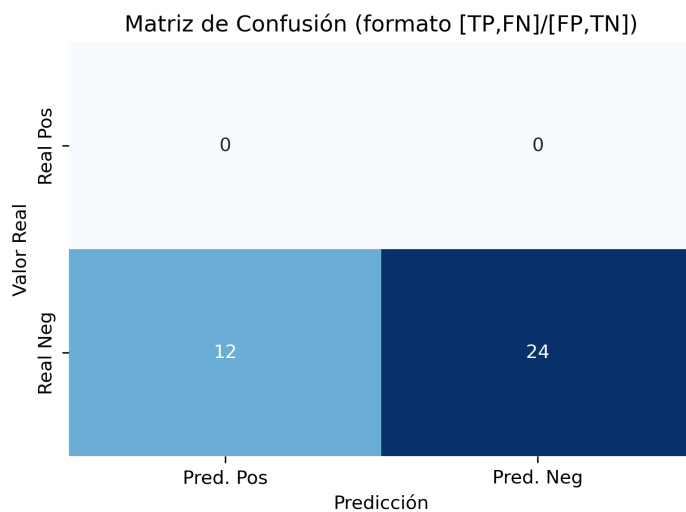


Figura 3: Matriz de confusión del modelo 3NN, con la clase 0 como 'Positive'.

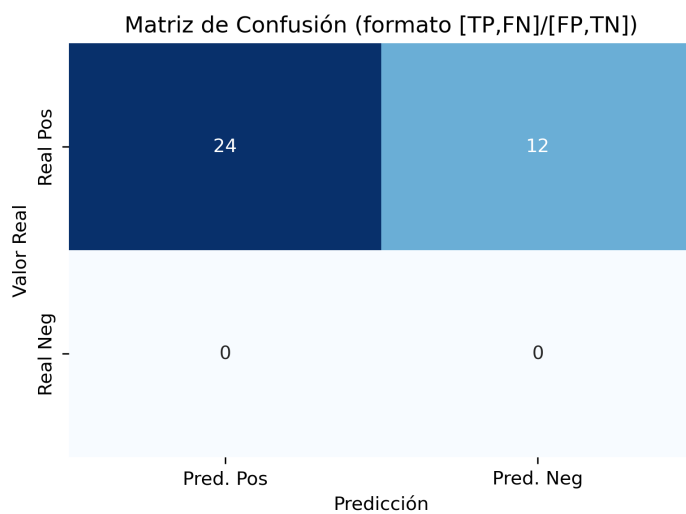


Figura 4: Matriz de confusión del modelo 3NN, con la clase 1 como 'Positive'.

### 3.3. 5NN

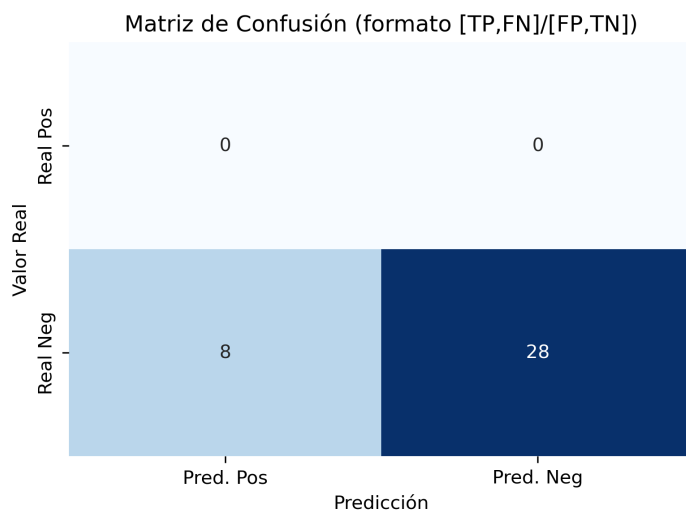


Figura 5: Matriz de confusión del modelo 5NN, con la clase 0 como 'Positive'.

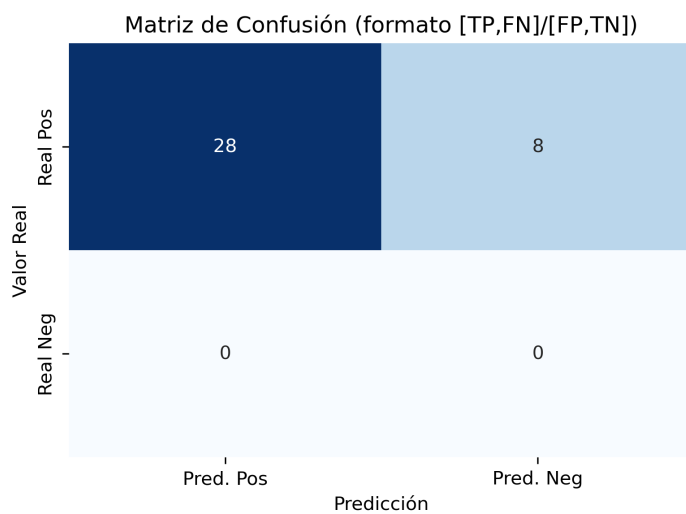


Figura 6: Matriz de confusión del modelo 5NN, con la clase 1 como 'Positive'.

### 3.4. SVM

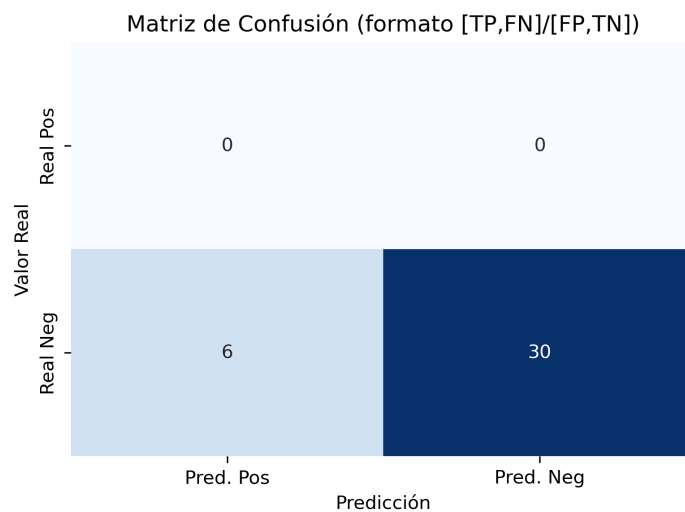


Figura 7: Matriz de confusión del modelo SVM, con la clase 0 como 'Positive'.

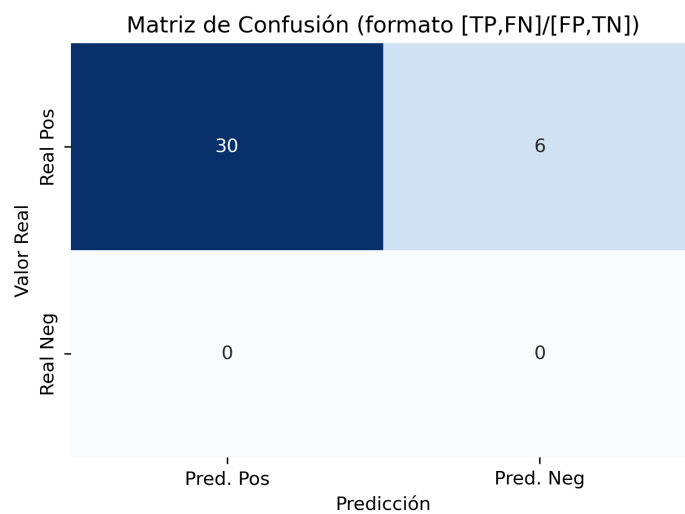


Figura 8: Matriz de confusión del modelo SVM, con la clase 1 como 'Positive'.

### 3.5. Naive Bayes

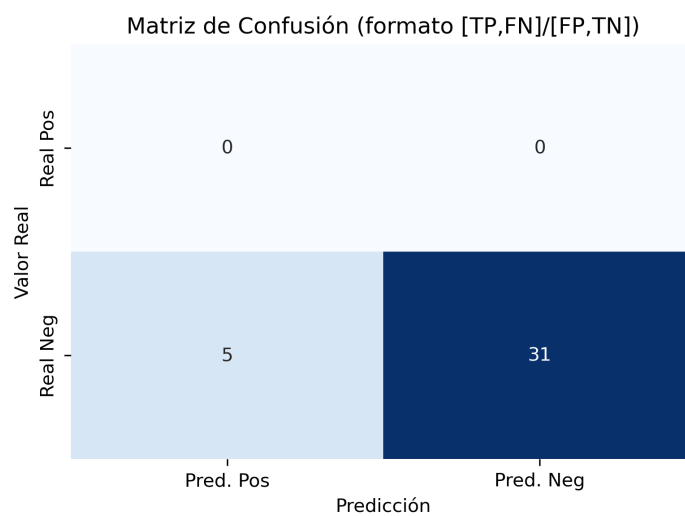


Figura 9: Matriz de confusión del modelo Naive Bayes, con la clase 0 como 'Positive'.

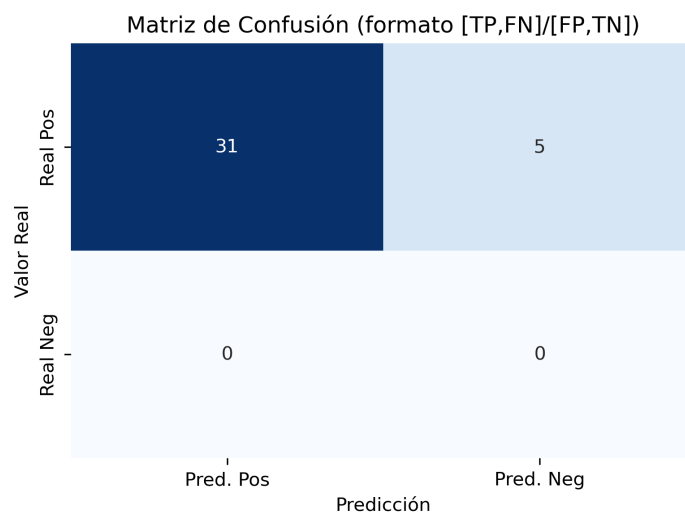


Figura 10: Matriz de confusión del modelo Naive Bayes, con la clase 1 como 'Positive'.



### 3.6. C5.0

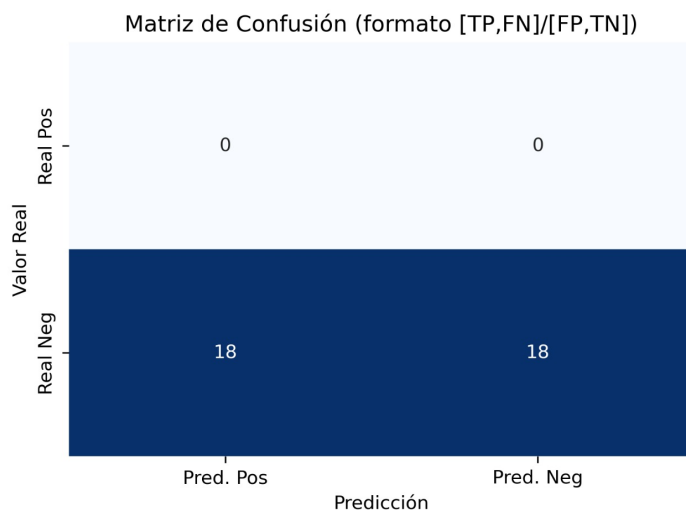


Figura 11: Matriz de confusión del modelo C5, con la clase 0 como 'Positive'.

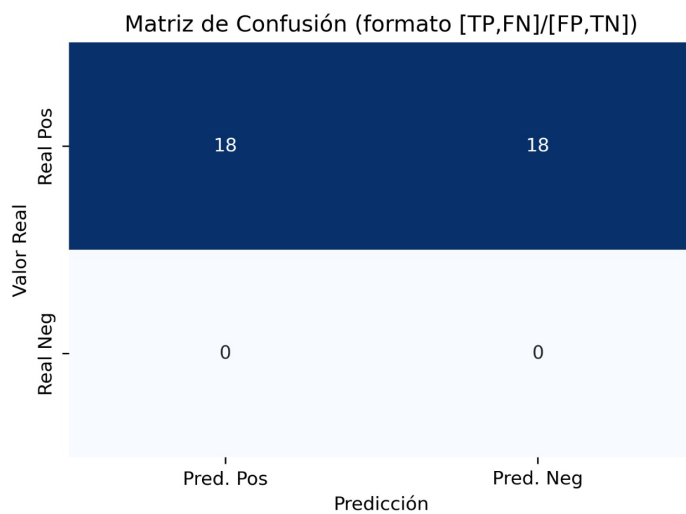


Figura 12: Matriz de confusión del modelo C5, con la clase 1 como 'Positive'.

### 3.7. RF

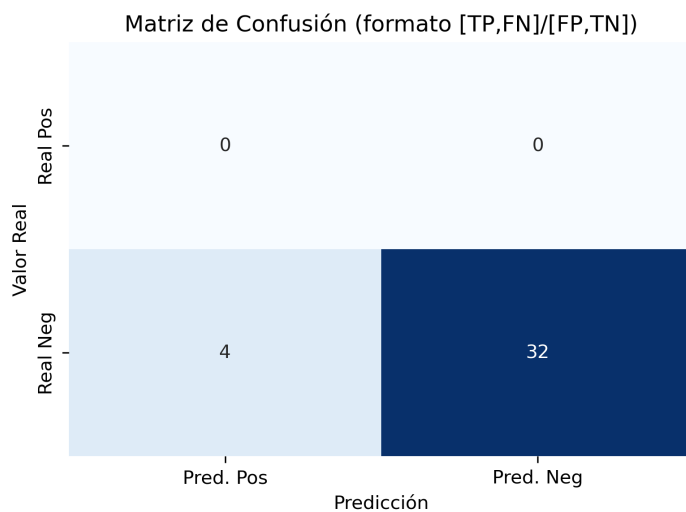


Figura 13: Matriz de confusión del modelo RF, con la clase 0 como 'Positive'.

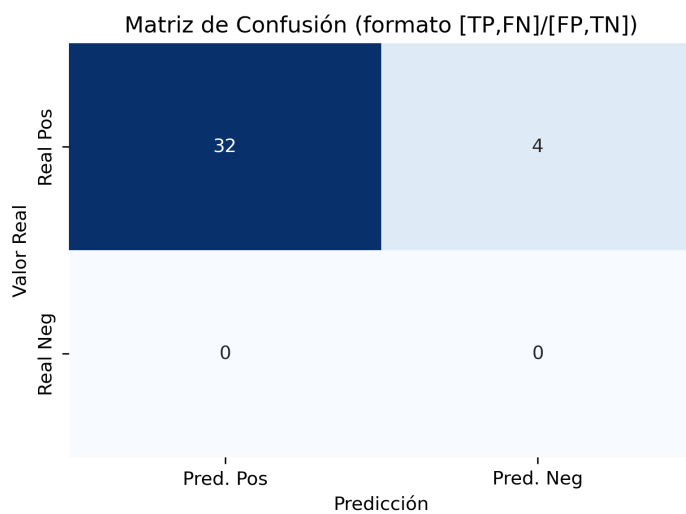


Figura 14: Matriz de confusión del modelo RF, con la clase 1 como 'Positive'.

### 3.8. XGBoost

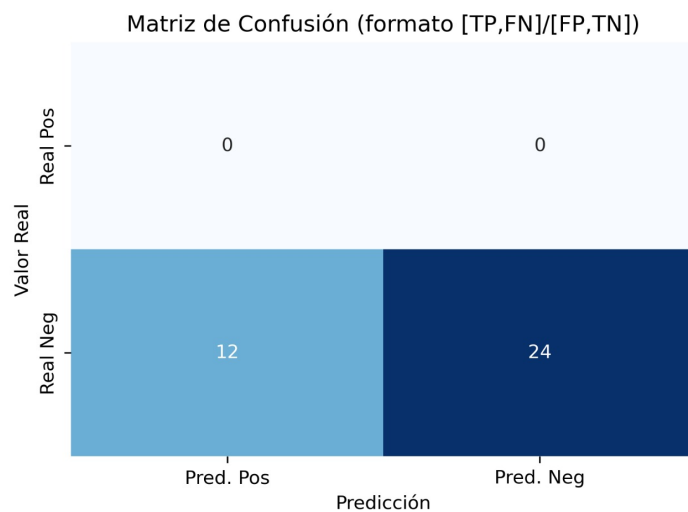


Figura 15: Matriz de confusión del modelo XGBOOST, con la clase 0 como 'Positive'.

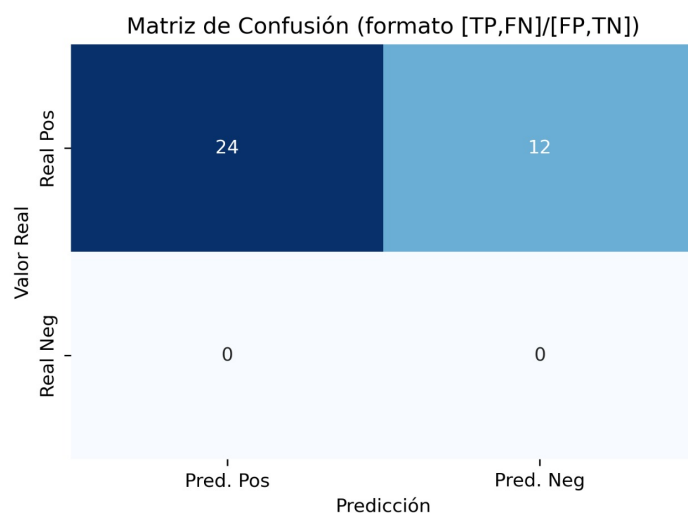


Figura 16: Matriz de confusión del modelo XGBOOST, con la clase 1 como 'Positive'.

## 4. Parte 4 - Medidas de desempeño

Reporta las medidas de desempeño por cada matriz de confusión/clasificador:

- Accuracy
- Error Rate
- Recall
- Specificity
- Balanced Accuracy
- Pprecision
- F1-Score
- MCC

Cuadro 1: Métricas de modelos con clase positiva 0

Modelo	Accuracy	Balanced Acc.	Error Rate	F1-Score	MCC	Precision	Recall	Specificity
1NN	0.4444	0.2222	0.5556	0.0	0.0	0.0	0.0	0.4444
3NN	0.6667	0.3333	0.3333	0.0	0.0	0.0	0.0	0.6667
5NN	0.7778	0.3889	0.2222	0.0	0.0	0.0	0.0	0.7778
SVM	0.8333	0.4167	0.1667	0.0	0.0	0.0	0.0	0.8333
Naive Bayes	0.8611	0.4306	0.1389	0.0	0.0	0.0	0.0	0.8611
RF	0.8889	0.4444	0.1111	0.0	0.0	0.0	0.0	0.8889
C5.0	0.5000	0.2500	0.5000	0.0	0.0	0.0	0.0	0.5000
XGBOOST	0.6667	0.3333	0.3333	0.0	0.0	0.0	0.0	0.6667

La tabla anterior nos muestra los resultados obtenidos mediante el uso de scikit-learn, pero es importante aclarar que cuando el valor de la medida de desempeño es "0.0", en realidad estamos hablando de un valor de una medida que ha sido indeterminada por los valores obtenidos tanto en las predicciones que fueron TP y FN.

Cuadro 2: Métricas de modelos con clase positiva 1

Modelo	Accuracy	Balanced Acc.	Error Rate	F1-Score	MCC	Precision	Recall	Specificity
1NN	0.4444	0.2222	0.5556	0.6154	0.0	1.0	0.4444	0.0
3NN	0.6667	0.3333	0.3333	0.8000	0.0	1.0	0.6667	0.0
5NN	0.7778	0.3889	0.2222	0.8750	0.0	1.0	0.7778	0.0
SVM	0.8333	0.4167	0.1667	0.9091	0.0	1.0	0.8333	0.0
Naive Bayes	0.8611	0.4306	0.1389	0.9254	0.0	1.0	0.8611	0.0
RF	0.8889	0.4444	0.1111	0.9412	0.0	1.0	0.8889	0.0
C5.0	0.5000	0.2500	0.5000	0.6667	0.0	1.0	0.5000	0.0
XGBOOST	0.6667	0.3333	0.3333	0.8000	0.0	1.0	0.6667	0.0

En la tabla anterior, los resultados son correspondientes a utilizar la clase positiva con el valor de 1. El valor de MCC tiene como resultado 0 ya que el cálculo queda indeterminado teniendo un valor para Specificity de 0, al tratarse de un caso extremo.