|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Referência | |
|  |  | Evento | No Event |
| Predito | Evento | A (TP) | B (FP) |
| No Event | C (FN) | D (TN) |

# Métricas

Sensitivity | Recall = TP/(TP+FN)  
Quantos eventos reais (positivos) foram corretamente identificados.

Specificity = TN/(FP+TN)  
Quantos não eventos (negativos) foram corretamente identificados.

Prevalence = (TP+FN)/(TP+FP+FN+TN)  
Proporção de eventos reais na população (eventos positivos).

PPV (Pos Pred Value) | Precision [1] = (sensitivity \* prevalence)/((sensitivity\*prevalence) + ((1-specificity)\*(1-prevalence))) = TP/(TP+FP)  
Proporção de eventos previstos como positivos que eram realmente positivos.

NPV (Neg Pred Value) | Precison [0] = (specificity \* (1-prevalence))/(((1-sensitivity)\*prevalence) + ((specificity)\*(1-prevalence))) = TN/(FN+TN)  
Proporção de predições negativas que eram realmente negativas.

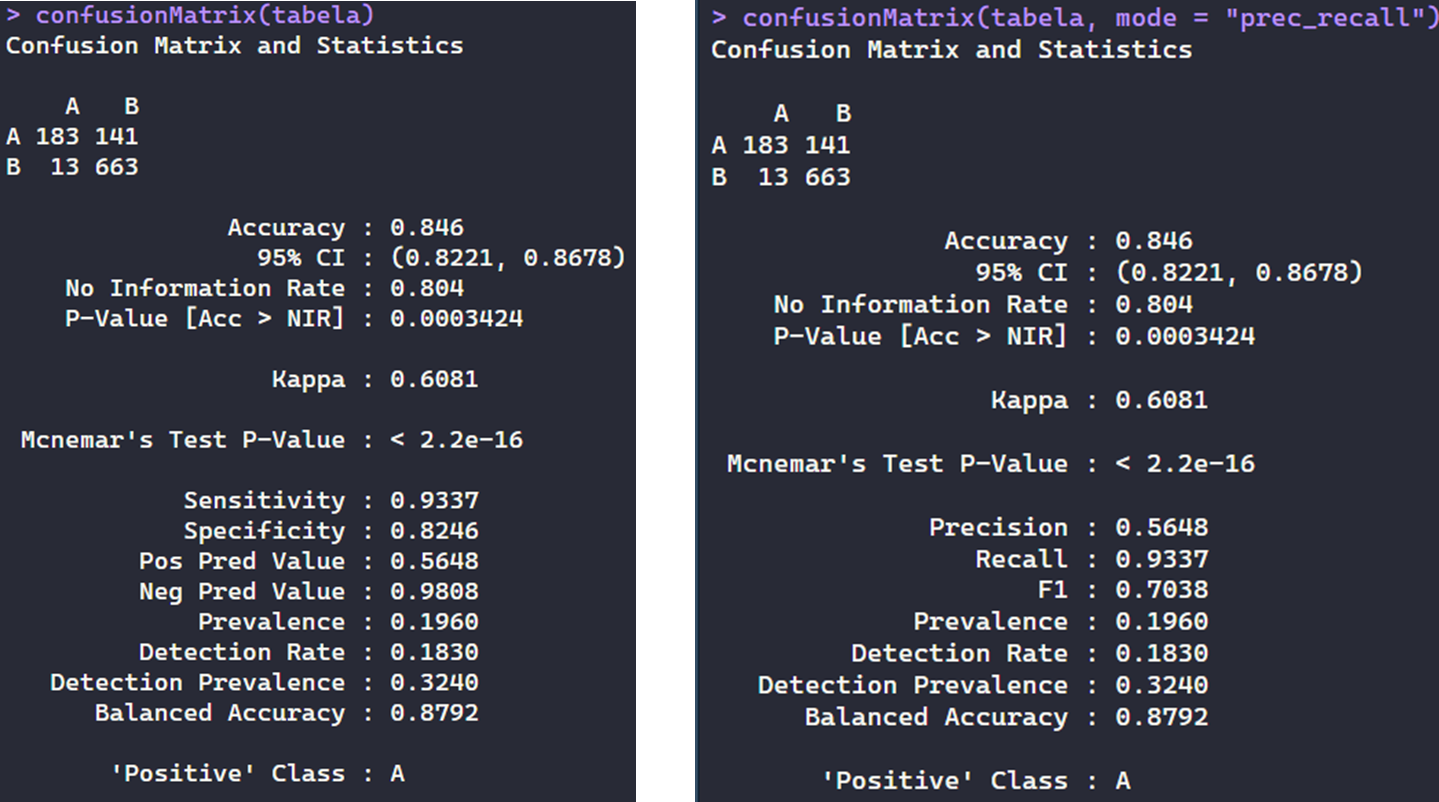
Detection Rate = TP/(TP+FP+FN+TN)  
Proporção de eventos reais detectados em relação ao total da população.

Detection Prevalence = (TP+FP)/(TP+FP+FN+TN)  
Proporção de predições positivas feitas pelo modelo (independentemente de serem corretas ou não).

Balanced Accuracy = (sensitivity+specificity)/2  
Média das taxas de acerto para eventos e não eventos. Usada quando há desequilíbrio entre classes.

F1 Score /Fβ Score = (1+β^2)\*precision\*recall/((β^2 \* precision)+recall)  
Média harmônica entre precisão e recall

# Saida no R:



É expresso duas métricas novas

**Kappa**: Mede a concordância entre as predições do modelo e os rótulos reais, corrigindo pela concordância que poderia ocorrer ao acaso. Útil especialmente quando há desbalanceamento entre classes.

**McNemar's Test p-value**: Testa a simetria dos erros do modelo, especificamente se há uma diferença significativa entre os falsos positivos e falsos negativos.

# Saida no Python:

Uma imagem contendo Interface gráfica do usuário

Descrição gerada automaticamente

Mais sobre F-β Score

* **β>1**: Atribui mais peso à **recall** do que à **precisão**. β=2, por exemplo, significa que a recall é duas vezes mais importante do que a precisão.
* **β<1**: Dá mais importância à **precisão** do que à **recall**. β=0.5, por exemplo, indica que a precisão é mais importante.
* **β=1**: Resulta no F1 Score, onde precisão e recall têm igual peso.

Em termos de limites:

* **β→0**: Nesse limite, o Fβ​ Score tende a focar exclusivamente na precisão, ignorando praticamente a recall.
* **β→∞**: Neste caso, o Fβ​ Score dá foco exclusivo no recall, ignorando a precisão.

Na prática, os valores comuns de β são 0.5, 1 e 2, mas não há um limite máximo ou mínimo teórico, desde que seja maior ou igual a zero.

Para ajustar o valor de β no F-β Score sem desequilibrar excessivamente a métrica entre precisão e recall, alguns valores de referência podem ser considerados:

1. **β = 0.5**: Coloca uma leve ênfase na precisão em relação à recall. Esse valor pode ser útil quando os falsos positivos são mais prejudiciais e você quer garantir uma maior precisão, mas sem descartar completamente a recall.

2. **β = 1**: Este é o F1 Score, que trata precisão e recall como igualmente importantes. É uma escolha equilibrada e um valor de referência comum em muitos cenários onde não há uma clara preferência entre as duas métricas.

3. **β = 2**: Coloca uma leve ênfase no recall em relação à precisão. Esse valor é interessante quando os falsos negativos são mais prejudiciais e é preferível capturar a maioria dos exemplos positivos, mesmo que ocasionalmente haja falsos positivos.

Considerações Gerais

- **Valores de β entre 0.5 e 2**: São considerados moderados e ajudam a não inclinar a balança de forma drástica para um lado ou outro.

- **β < 0.5 ou β > 2**: Pode começar a causar desequilíbrios acentuados. Valores muito baixos (próximos de zero) focam quase exclusivamente na precisão, enquanto valores muito altos (acima de 5, por exemplo) dão ênfase quase exclusiva à recall.

Esses valores de referência 0.5, 1 e 2 costumam ser suficientes para a maioria dos casos, mantendo o equilíbrio entre as métricas e ajustando a importância de cada uma sem exageros.