

Momento de Retroalimentación

Modulo 2

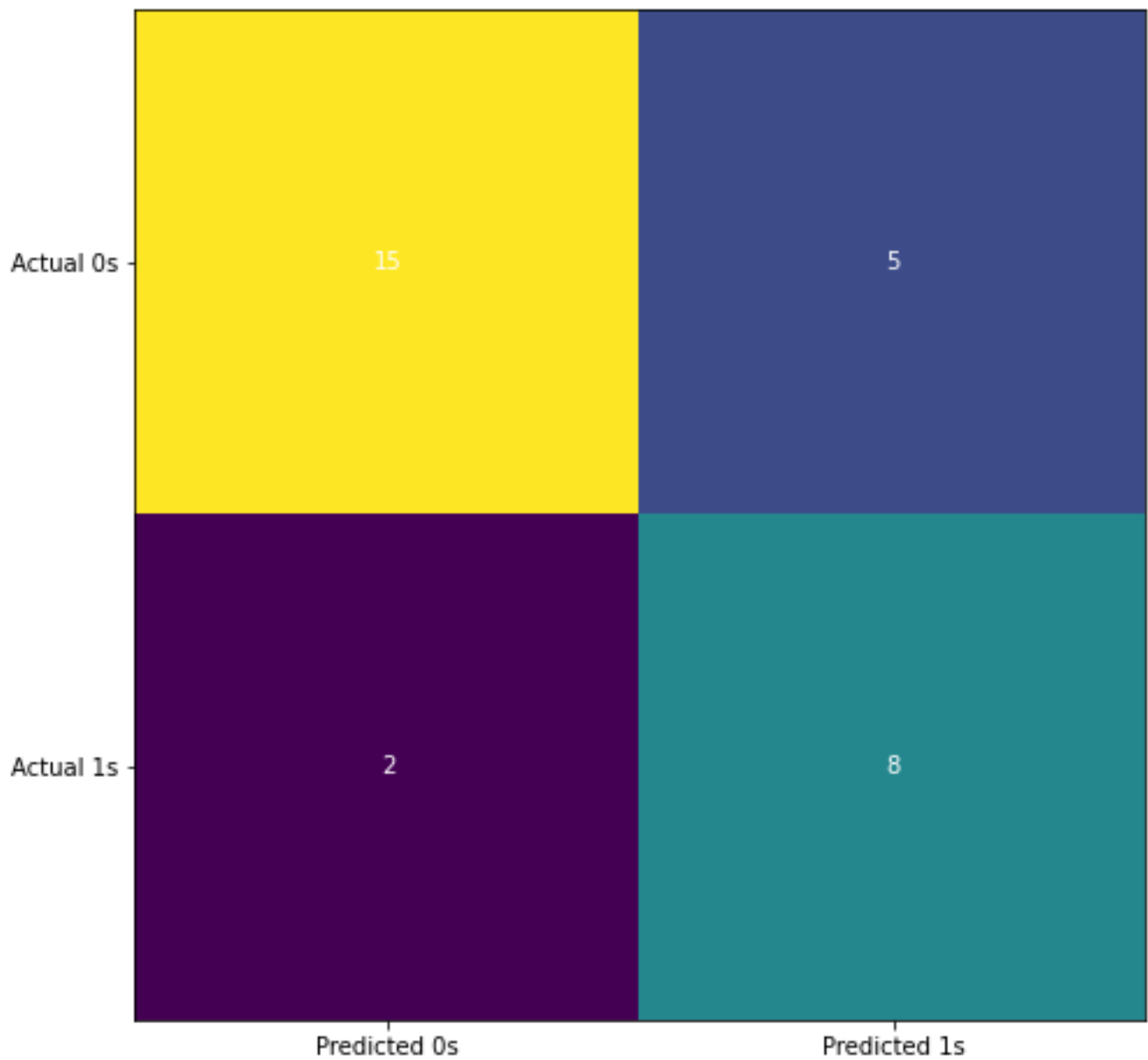
Análisis y reporte sobre el desempeño del modelo

Reporte

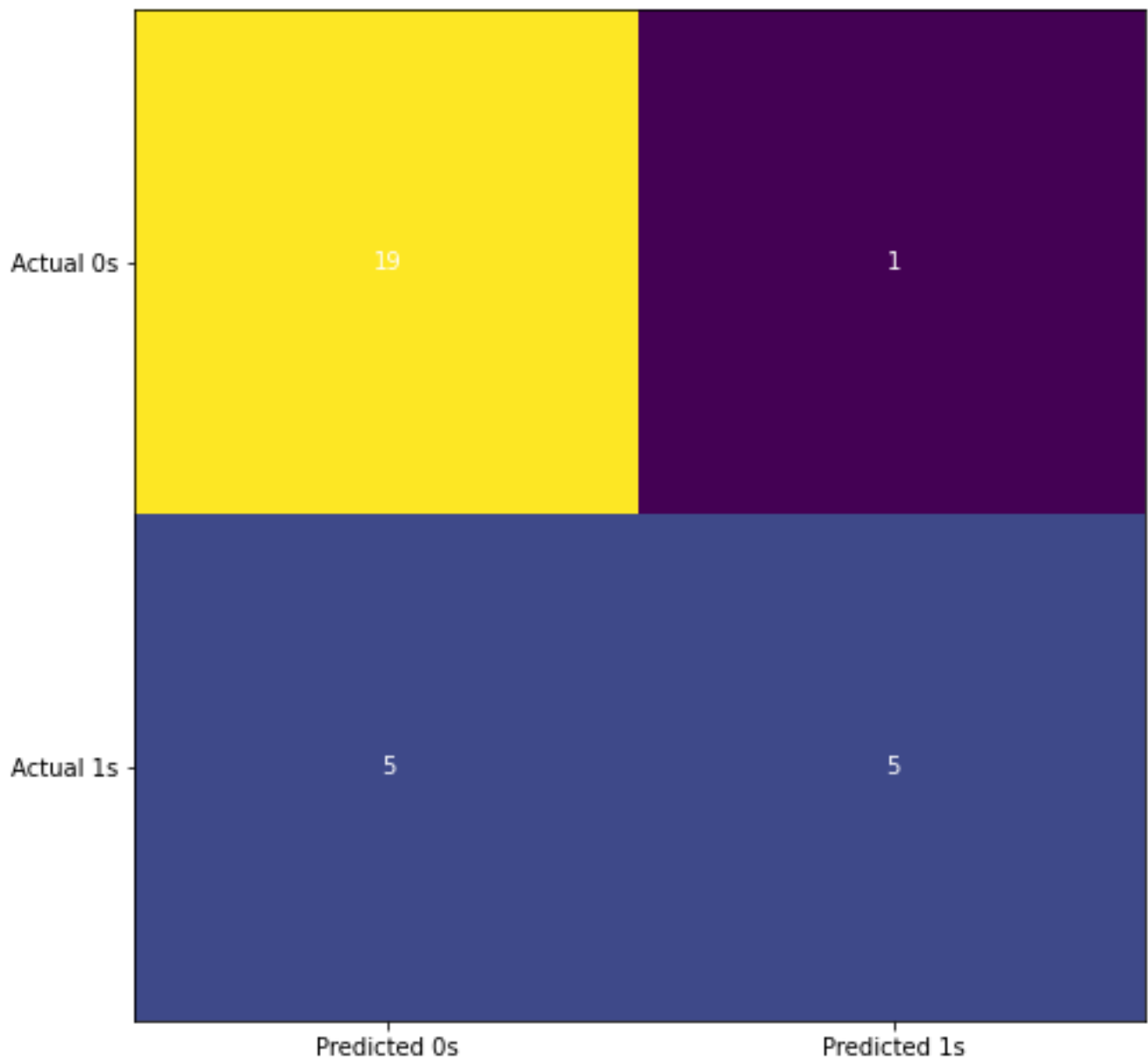
Para este análisis se hizo uso del segundo modelo implementado, modelo de regresión logística utilizando librerías. Junto a esto, se utilizó el *dataset* de *iris* que identifica tipos de flor dependiendo de las dimensiones de la misma. Para el primer modelo se buscaba identificar la flor *Iris-setosa* de las demás utilizando el ancho del sépalo. Además, se establecieron parámetros generales que serán de utilidad más adelante. Estos son un tamaño de datos de entrenamiento de 80%; estados de *random* para separación de datos, entrenamiento y estimación de *bias* y varianza de 42 (arbitrariamente escogido); ningún penalti en entrenamiento; tolerancia de 0.1; sin intercepción; utilizando el método de solución “sag”; y con un máximo de iteraciones de 100.

El *bias* detectado fue de 0.23 y una varianza de 0.0. Tomando en cuenta esto, la varianza es baja, pero el *bias* pareciera ser medio. Se observó también una evaluación en entrenamiento de 0.78 y una evaluación en validación de 0.77. Esto refleja la varianza baja ya que el modelo no está en un *overfitt* dado que las evaluaciones son similares. De la misma forma, muestra un *bias* presente, mas no alto, ya que el modelo tiene una evaluación relativamente baja aunque no mala. Por esta misma razón, el modelo está pasando de *underfitt* a *fitt*.

En el segundo modelo generado se buscó mejorar la evaluación general. El resultado de *bias* del modelo fue de 0.2 y con una varianza de 0.02. Esta variación nos muestra un valor de varianza bajo, pero que a incrementado un poco, a cambio de reducir el valor del *bias*. El *bias* sigue estando presente, pero en menor medida y podría empezar a considerarse como bajo dada la evaluación final del modelo. Esta evaluación resultó ser de 0.83 en entrenamiento y de 0.8 en validación. Se puede observar que el modelo mejoró de forma considerable y ahora podría considerarse como *fitt*. En cuanto a los cambios realizados en los parámetros, se cambió el penalti a “l2”; la tolerancia ahora es de 0.001; se considera una intercepción; y el método de resolución ahora es “saga”.



Se incluyen las matrices de confusión de los modelos como comparación. Modelo 1



Modelo 2

Los modelos presentados muestran una mejora y esta se puede observar y evaluar con el *bias*, la varianza y la evaluación en entrenamiento y validación. Esto demuestra el valor de estas métricas para ayudar a identificar las características del modelo y sus áreas de mejora. Esto, a su vez, es de utilidad para mejorar dicho modelo y obtener un modelo que será de mayor utilidad para el propósito final del análisis.