

Análisis y reporte sobre el desempeño del modelo.



Curso:

**Inteligencia artificial avanzada
para la ciencia de datos I**

Presenta:

Flavio Ruvalcaba Leija

Fecha:

09-06-2023

Modelo utilizado:

Implementación de regresión logística con el uso de un framework.

Link:

<https://github.com/FlavioRr/InteligenciaArtificialGpo101/tree/main/MachineLearning1>

Puntos a considerar:

- **Separación y evaluación del modelo con un conjunto de prueba y un conjunto de validación (Train/Test/Validation).**

Con el fin de entrenar el modelo, decidí separar la base de datos `gpa_study_hours` en un 70%, un 15% y un 15%. Mi criterio para esta toma de decisión es el fin de reportar evaluaciones de mi modelo, explico:

Train:

El modelo de regresión logística se entrena en el conjunto de entrenamiento (**`df_train`**). Esto implica ajustar los parámetros del modelo utilizando estos datos, de manera que el modelo aprenda a hacer predicciones basadas en las características proporcionadas (en este caso, **`gpa` y `study_hours` (`x_train`)**) para predecir la variable objetivo (**`graduated` (`y_train`)**).

Test:

Una vez que el modelo ha sido entrenado, se evalúa en el conjunto de prueba (**`df_test`**). Esto proporciona una medida de qué tan bien el modelo se desempeña en datos que no ha visto durante el entrenamiento. La

métrica de exactitud (Accuracy) se calcula para medir su rendimiento en este conjunto.

Validation:

Después de evaluar el modelo en el conjunto de test, se utiliza el conjunto de datos de validation (**df_validation**) para obtener una evaluación adicional. Esto puede ayudar a verificar que el modelo no haya sobreajustado los datos de entrenamiento y que sea capaz de generalizar bien a datos nuevos y no vistos. La métrica de exactitud también se calcula en el conjunto de validación.

- **Diagnóstico y explicación el grado de bias o sesgo: bajo medio alto**

El sesgo o bias se refiere a un tipo de error sistemático que ocurre cuando un modelo no es capaz de capturar con precisión la relación entre las características de entrada (variables independientes) y la variable de salida (variable dependiente) en un conjunto de datos.

Sesgo en el Conjunto de Prueba (test):

- True Positives (TP): 29
- True Negatives (TN): 5
- False Positives (FP): 1
- False Negatives (FN): 0

Sesgo en el Conjunto de Validación (validation):

- True Positives (TP): 27
- True Negatives (TN): 5
- False Positives (FP): 2

- False Negatives (FN): 2

El modelo que se ha presentado tiene un diagnóstico de sesgo medio ya que, con los datos anteriores, podemos interpretar que el conjunto de test, el sesgo es bajo ya que el modelo tiene un buen rendimiento en la detección de casos positivos y la minimización de falsos positivos, sin embargo, en el conjunto de validación, el sesgo es medio-alto debido a la presencia de falsos negativos y falsos positivos, lo que indica que el modelo no está realizando predicciones tan precisas en este conjunto en comparación con el conjunto de prueba.

Sabemos esto ya que en el conjunto de prueba, la ausencia de False Negatives (FN = 0) y un número bajo de False Positives (FP = 1) indican que el modelo tiene un buen rendimiento en la detección de casos positivos y en la minimización de errores de clasificación de casos negativos, lo que sugiere un sesgo bajo y por el otro lado, en el conjunto de validación, la presencia de False Negatives (FN = 2) y False Positives (FP = 2) indica que el modelo está cometiendo errores en la detección de casos positivos y en la clasificación de casos negativos, lo que sugiere un sesgo medio-alto.

- **Diagnóstico y explicación el grado de varianza: bajo medio alto**

La varianza se refiere a cuán sensible es el rendimiento del modelo a las pequeñas variaciones en los datos de entrenamiento.

Utilizamos la herramienta cross validation que nos ayuda a identificar la variabilidad en un conjunto de datos.

Con la obtención de una precisión promedio en cross-validation (4-fold): 0.96 +/- 0.04 podemos identificar que tenemos una variabilidad de +/- 0.04.

Una varianza más baja (por ejemplo, "+/- 0.01") indicaría que la precisión fue bastante consistente y cercana al promedio. Por otra parte, Una varianza más alta (" +/- 0.10") indicaría que la precisión varió significativamente.

Por lo tanto, podemos definir que la varianza del modelo es media.

- **Diagnóstico y explicación el nivel de ajuste del modelo: underfitt
fitt overfitt**

El nivel de ajuste del modelo se refiere a cuán bien un modelo de aprendizaje automático representa o se adapta a los datos en el conjunto de entrenamiento.

Utilizamos `r2_score` de la librería de `sklearn`, lo cual nos ayudara a obtener el coeficiente de determinación r^2 .

Con r^2 ahora podemos calcular la calidad de un modelo de regresión. En pocas palabras r cuadrada es una métrica para evaluar cuanta variabilidad de la variable dependiente es explicada por nuestro modelo. Si esta cerca de uno significa un buen ajuste.

De la misma manera utilizamos `cross-validation` para detectar el sobreajuste al evaluar el rendimiento del modelo en diferentes conjuntos de datos de prueba lo que proporciona una evaluación más robusta y confiable del modelo.

Finalmente, con los resultados obtenidos:

1. True Positives (TP): 29:

- Indica el número de casos positivos que el modelo ha clasificado correctamente como positivos en el conjunto de test. En este caso, el modelo acertó 29 casos positivos.

2. True Negatives (TN): 5:

- Representa el número de casos negativos que el modelo ha clasificado correctamente como negativos en el conjunto de test. Aquí, el modelo acertó 5 casos negativos.

3. False Positives (FP): 1:

- Muestra el número de casos que el modelo ha clasificado incorrectamente como positivos cuando en realidad eran negativos en el conjunto de test. En este caso, el modelo cometió 1 error de este tipo.

4. False Negatives (FN): 0:

- Indica el número de casos que el modelo ha clasificado incorrectamente como negativos cuando en realidad eran positivos en el conjunto de test. En este caso, el modelo no cometió errores de este tipo.

5. Recall: 1.0 (100%):

- Recall es la proporción de casos positivos que el modelo ha identificado correctamente con respecto al total de casos positivos. Un valor de 1.0 significa que el modelo ha capturado todos los casos positivos en el conjunto de test.

6. Precision: 96.67%:

- La precisión mide la proporción de casos clasificados como positivos por el modelo que son verdaderos positivos. Un valor del 96.67% indica que la mayoría de las predicciones positivas realizadas por el modelo son correctas.

7. Accuracy en test: 97.14%:

- La exactitud (accuracy) es la proporción total de predicciones correctas realizadas por el modelo en el conjunto de test. Un valor del 97.14% muestra un alto nivel de aciertos en general en el conjunto de prueba.

8. R^2 ajustado en test: 0.80:

- R^2 tiene un valor de 0.80 lo que sugiere un buen ajuste del modelo en el conjunto de test, ya que indica que una parte significativa de la variabilidad en las predicciones se debe al modelo.

9. Accuracy en validation: 88.89%:

- La exactitud en el conjunto de validation es del 88.89%, lo que indica un buen rendimiento en datos de validación independientes.

10. R^2 ajustado en validation: 0.29:

- R^2 ajustado en el conjunto de validación es relativamente bajo (0.29), lo que sugiere que el modelo puede no ajustarse tan bien a los datos de validation como lo hace en el conjunto de test.

11. **Precisión promedio en cross-validation (4-fold): 0.96 +/- 0.04:**

- La precisión promedio en validación cruzada es del 96% en promedio, con una pequeña variación. Esto indica un buen rendimiento del modelo en diferentes divisiones del conjunto de entrenamiento durante cross validation.

En general, los resultados sugieren que el modelo tiene un buen ajuste a los datos de prueba y no muestra signos de underfitting u overfitting. Sin embargo, hay una diferencia en el rendimiento entre los conjuntos de prueba y validación, lo que indica cierto grado de variabilidad o sesgo en los datos de validación como establecimos anteriormente.

Basándote en lo encontrado en tu análisis utiliza técnicas de regularización o ajuste de parámetros para mejorar el desempeño de tu modelo y documenta en tu reporte cómo mejoró este.

La técnica de regularización utilizada fue por parte de la librería sklearn "regularización L2". La regularización L2 agrega una penalización a la función de coste del modelo de regresión logística en función de la suma de los cuadrados de los coeficientes del modelo.

Al agregar la regularización L2 en scikit-learn, configuramos el parámetro **penalty** en "l2" al crear una instancia del modelo **LogisticRegression**.

Antes:

```
Accuracy en test: 97.14%
R2 ajustado en test: 0.80
Accuracy en validation: 88.89%
R2 ajustado en validation: 0.29
Precisión promedio en cross-validation (4-fold): 0.96 +/- 0.04
```

Después:

```
Accuracy en test: 97.14%  
R2 ajustado en test: 0.80  
Accuracy en validation: 94.44%  
R2 ajustado en validation: 0.65  
Precisión promedio en cross-validation (4-fold): 0.95 +/- 0.06
```

Como podemos observar, con el uso de esta técnica de regularización, el modelo aumento su accuracy de 88.8 % a 94.44% y r^2 , de igual manera, recibió un aumento de 0.29 a 0.65 ambos en el conjunto de validation, en cambio al resultado de cross validation, recibimos una mayor desviación y un mayor promedio.