



Inteligencia artificial avanzada para la ciencia de datos

Momento de Retroalimentación: Módulo 2 Análisis y Reporte sobre el desempeño del modelo. (Portafolio Análisis)

Alex Federico Núñez Escobar

A01751559

11 de Septiembre 2023

Instituto Tecnológico y de Estudios Superiores de Monterrey.
Campus Estado de México.

“Yo, como integrante de la comunidad estudiantil del Tecnológico de Monterrey, soy consciente de que la trampa y el engaño afectan mi dignidad como persona, mi aprendizaje y mi formación, por ello me comprometo a actuar honestamente, respetar y dar crédito al valor y esfuerzo con el que se elaboran las ideas propias, las de los compañeros y de los autores, así como asumir mi responsabilidad en la construcción de un ambiente de aprendizaje justo y confiable.”

Análisis del modelo

Separación y evaluación del modelo

1. Conjunto de Entrenamiento (Training Set): Este conjunto se utiliza para entrenar el modelo.
2. Conjunto de Validación (Validation Set): Este conjunto se utiliza para evaluar su rendimiento para determinar los mejores hiper parámetros una vez que el modelo está entrenado.
3. Conjunto de Prueba (Test Set): El conjunto de prueba se usa para evaluar el rendimiento del modelo. Esta evaluación proporciona una estimación imparcial del rendimiento del modelo en datos completamente nuevos.

```
# División del dataset en conjuntos de entrenamiento y temporal (test + validación)
X_train_temp, X_temp, y_train_temp, y_temp = train_test_split(X, y, test_size=0.3, random_state=42)

# División del conjunto temporal en conjuntos de prueba y validación
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, random_state=42)
```

Diagnósticos

- Bias / Sesgo

El bias o sesgo es la capacidad de un modelo para representar con precisión el mapeo entre las entradas y las salidas en los datos. [1]

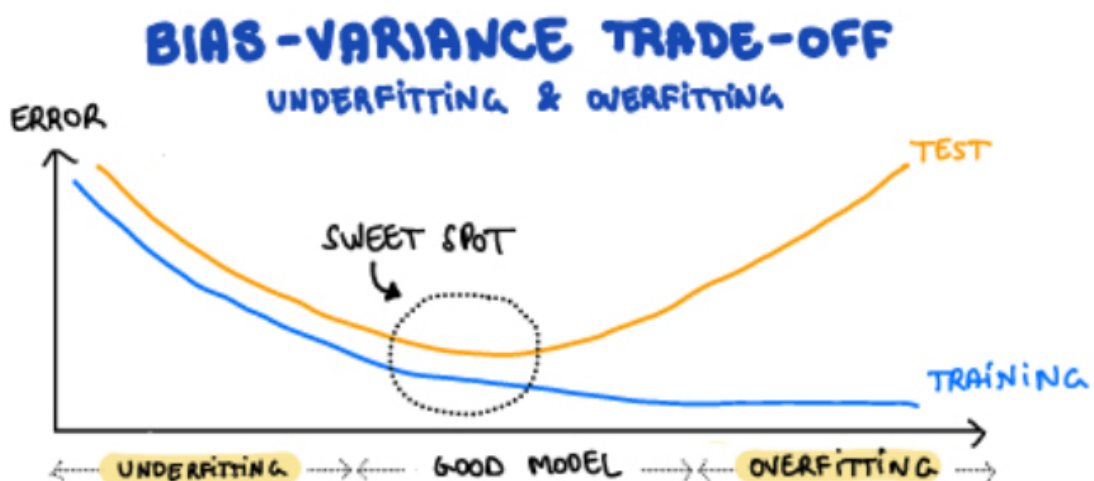


Imagen 1. Bias-Varianza(Kepler's crew, 2022)

En base al dataset que elegí ocupe el siguiente Bias [2]

1. Nivel Bajo de Bias (Sesgo)
 - Rango: 0.0 - 0.1

- En árboles de decisión, un bajo sesgo se ajusta adecuadamente a los datos de entrenamiento y es capaz de capturar con precisión las relaciones subyacentes entre las características y la variable objetivo.
- Generalmente, esto implica que el árbol tiene una profundidad.

2. Nivel Medio de Bias (Sesgo)

- Rango: 0.1 - 0.3
- En árboles de decisión puede no estar capturando completamente la complejidad de los datos.
- Puede deberse a que el árbol no es lo suficientemente profundo o a que se han ignorado algunas características importantes.

3. Nivel Alto de Bias (Sesgo)

- Rango: >0.3
- Se debe a una profundidad demasiado superficial, divisiones inapropiadas o la falta de características relevantes.

El grado de Bias del modelo es:

Grado de Bias (Sesgo): 0.0508 - Bajo

- Varianza

La varianza es una medida de cuánto cambian las predicciones del modelo para diferentes conjuntos de entrenamiento.

1. Alta Varianza:

- El modelo ha aprendido "demasiado" de los datos de entrenamiento, incluido el ruido y las fluctuaciones aleatorias.
- Se desempeña muy bien en los datos de entrenamiento pero tiene un rendimiento pobre en validación o prueba.

2. Baja Varianza:

- El modelo no es sensible a las fluctuaciones en el conjunto de entrenamiento.

El grado de varianza del modelo es:

Grado de Varianza: 0.0169 - Bajo

- Nivel de ajuste

El nivel de ajuste del modelo es:

El modelo tiene un buen equilibrio entre bias y varianza (fitt)

Mejorar el modelo

- Aumentando el nivel de profundidad del árbol disminuye el nivel de vías y de la varianza pero corremos el riesgo de llegar a un nivel de overfit por lo que el modelo es aceptable con el nivel de profundidad predicha. Llega a un accuracy de 1.0.

Referencias:

- [1] Gonzalez, L. (2022). Sesgo y varianza en machine learning. 🤖 Aprende IA. <https://aprendeia.com/bias-y-varianza-en-machine-learning/>
- [2] Romero, I. (2021, 4 marzo). La dicotomía sesgo-varianza en modelos de machine learning - Keeper | Cloud Data Driven Partner. Keeper | Cloud Data Driven Partner. [https://keeper.io/es/2021/03/la-dicotomia-sesgo-varianza-en-modelos-de-machine-learning/#:~:text=El%20sesgo%20\(o%20bias\)%20es,y%20la%20variable%20a%20predecir.](https://keeper.io/es/2021/03/la-dicotomia-sesgo-varianza-en-modelos-de-machine-learning/#:~:text=El%20sesgo%20(o%20bias)%20es,y%20la%20variable%20a%20predecir.)
- [3] Huilgol, P. (2023). Bias and Variance in Machine Learning – A fantastic guide for beginners! *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2020/08/bias-and-variance-tradeoff-machine-learning/>