Actividad 3.3 – IRIS con NaiveBayes y Cross Validation

Link al github: [github](https://github.com/Javiert54/bigData/tree/main/CEIABD_SNS/UT3%20-%20Algoritmos%20y%20herramientas%20para%20el%20aprendizaje%20supervisado/Actividad%203.3%20–%20IRIS%20con%20NaiveBayes%20y%20Cross%20Validation)

El objetivo de esta actividad es observar las diferencias en la Predicción en los modelos de   
NaiveBayes al utilizar o no la Validación cruzada en el entrenamiento de los modelos. Para ello   
anotaremos los resultados en la siguiente tabla comparativa:

|  |  |  |
| --- | --- | --- |
|  | Predicción | |
| Modelo | Sin CrossValidation | Con CrossValidation |
| GaussianNB | 0.64 | 0.75 |
| MultiNomialNB | 0.62 | 0.67 |
| BernouiliNB | 0.28 | 0.56 |
| ComplementNB | 0.62 | 0.67 |
| CategoricalNB | 0.98 |  |

La validación cruzada (cross validation) es una técnica utilizada en el aprendizaje automático para evaluar la capacidad de un modelo de generalizar a datos no vistos. Es decir, nos ayuda a entender cómo de bien funcionará nuestro modelo cuando se enfrente a nuevos datos que no ha visto durante el entrenamiento.

Usar la validación cruzada hace una diferencia importante porque:

- Evalúa el modelo varias veces con diferentes partes de los datos, dando una evaluación más precisa y confiable.

- Utiliza todos los datos tanto para entrenar como para probar, lo cual es útil si tenemos pocos datos.

- Ayuda a ver si el modelo está aprendiendo demasiado de los datos de entrenamiento y no generaliza bien a datos nuevos.