

Módulo 2: Análisis y reporte sobre el desempeño del modelo.

Alan Rodrigo Vega Reza; A01750658

El modelo que yo elegí para el análisis de desempeño es el utilizado en la retroalimentación 2.

Ahora bien, para este análisis se hizo la división debida del dataset en las 3 partes solicitadas, después se agregaron al código las secciones necesarias para medir el bias y la varianza del modelo, en la primera ejecución se utilizó un RandomForestClassifier con los siguientes parámetros:

```
model = RandomForestClassifier(n_estimators=100, max_depth=5, random_state=1)
model.fit(X_train, y_train)
```

El resultado de esta ejecución dio un bias de 0.18, una varianza de -0.06 y un estado de Underfitting:

```
Train Accuracy: 0.82
Validation Accuracy: 0.76
Test Accuracy: 0.94
Bias (Error en entrenamiento): 0.18
Variance (Diferencia entre validación y entrenamiento): -0.06
Model Fit Status: Underfitting
```

Se ajustaron los parámetros a lo siguiente:

```
# Entrenar el modelo
model = RandomForestClassifier(n_estimators=150, max_depth=8, random_state=1)
model.fit(X_train, y_train)
```

Dando los siguientes resultados en ejecución:

```
Train Accuracy: 0.83
Validation Accuracy: 0.75
Test Accuracy: 0.93
Bias (Error en entrenamiento): 0.17
Variance (Diferencia entre validación y entrenamiento): -0.07
Model Fit Status: Underfitting
Precision: 0.89
Recall: 0.92
F1-Score: 0.90
```

Dado que esto seguía juzgando el modelo como overfiteado probé a seguir modificando parámetros, aumentando los estimators y el depth pero continuó mostrando valores similares de bias y de varianza, utilizando los siguientes parámetros:

```
Classifier(n_estimators=10000, max_depth=100, random_state=1)  
_train)
```

Dio estos resultados:

```
Train Accuracy: 0.83  
Validation Accuracy: 0.77  
Test Accuracy: 0.93  
Bias (Error en entrenamiento): 0.17  
Variance (Diferencia entre validación y entrenamiento): -0.06  
Model Fit Status: Underfitting  
Precision: 0.92  
Recall: 0.89  
F1-Score: 0.91
```