

1. Análisis de Accuracy para el modelo.

Al correr nuestro modelo de **Clasificador de Árboles de Decisión** podemos observar que en la parte del entrenamiento obtenemos un **100 %** en la métrica de Accuracy, lo que nos indica que nuestro modelo estará en lo cierto, es decir, que clasificará de buena manera en el 100 % de las ocasiones. Ahora bien, en la parte de prueba podemos observar que el modelo recae en la métrica de Accuracy obteniendo tan solo un **69 %**.

Esto nos indica que el modelo en la parte de entrenamiento se está sobreajustando, para solucionar el problema de sobreajuste u overfit se tiene que limitar o ajustar el modelo para que aprenda a generalizar los datos en lugar de simplemente memorizarlos.

Por lo anterior, se procedió a ajustar ciertas características del modelo (en este caso la máxima profundidad del árbol y el criterio del mismo). Al hacer esto podemos observar que nuestro modelo obtiene un **77 %** de Accuracy en la parte de entrenamiento, lo que nos indica que el modelo ya no se está sobreajustando a los datos y está generalizándolos. En la parte de prueba el modelo aumenta casi un 10 % en la métrica de Accuracy, teniendo ahora un **77 %**.

Con esto podemos decir que el ajuste de las características específicas del modelo puede ayudarnos para evitar el sobreajuste y mejorar el desempeño del mismo en la parte de pruebas.

2. Análisis de la Confussion Matrix para el modelo.

La matrix de confusión nos es útil para que así podamos observar cómo el modelo está clasificando los datos y nos demos cuenta del desempeño de este de una manera más visual. Para la primera ejecución del modelo podemos observar que la matriz de confusión quedó de la siguiente manera:

		Clase real		Total
		Negativo	Positivo	
Clase Predicha	Negativo	358	0	358
	Positivo	0	179	179
Total		358	179	

Tabla 1: Matriz de Confusión en etapa de Entrenamiento.

		Clase real		Total
		Negativo	Positivo	
Clase Predicha	Negativo	111	31	142
	Positivo	40	49	89
Total		156	74	

Tabla 2: Matriz de Confusión en etapa de Pruebas.

Mientras que para el modelo en el cual hicimos ajustes, la matriz de confusión es de la siguiente manera:

		Clase real		Total
		Negativo	Positivo	
Clase Predicha	Negativo	313	45	358
	Positivo	78	101	179
Total		391	146	

Tabla 3: Matriz de Confusión en etapa de Entrenamiento.

		Clase real		Total
		Negativo	Positivo	
Clase Predicha	Negativo	127	15	142
	Positivo	37	52	89
Total		154	67	

Tabla 4: Matriz de Confusión en etapa de Pruebas.

Con esto podemos observar de manera detenida la mejora que tuvo el modelo al modificar sus hiperparámetros. Además de que se puede notar que el modelo clasificó de mejor manera comparado con la primera ejecución sin tocar nada.