

Reporte de Desempeño: KNN

1. Separación y Evaluación del Modelo (Train/Test/Validation)

- **Conjunto de Entrenamiento:** Se utilizó el 60% de los datos originales para entrenar el modelo.
- **Conjunto de Validación:** Un 20% de los datos fueron utilizados para la validación cruzada y ajuste de hiperparámetros.
- **Conjunto de Prueba:** El 20% restante fue reservado para evaluar el desempeño final del modelo.

El modelo fue ajustado en el conjunto de entrenamiento y validado utilizando una matriz de confusión y métricas de precisión, recall, y F1-Score. Luego, el desempeño fue evaluado en el conjunto de prueba.

- Resultados del conjunto de validación (después de Grid Search):
 - Precisión: 0.94
 - Recall: 0.94
 - F1 Score: 0.93
- Resultados del conjunto de prueba:
 - Precisión: 0.96
 - Recall: 0.97
 - F1 Score: 0.96

2. Diagnóstico del Grado de Bias (Sesgo)

El sesgo es una medida de la capacidad del modelo para generalizar bien a nuevos datos, reflejado en la diferencia entre la precisión de entrenamiento y la precisión en validación/prueba. En este caso, los errores son los siguientes:

- Error de Entrenamiento: 0.02
- Error de Prueba: 0.03

El bajo error en ambos conjuntos sugiere un sesgo bajo. El modelo está bien ajustado a los datos, no muestra signos de underfitting.

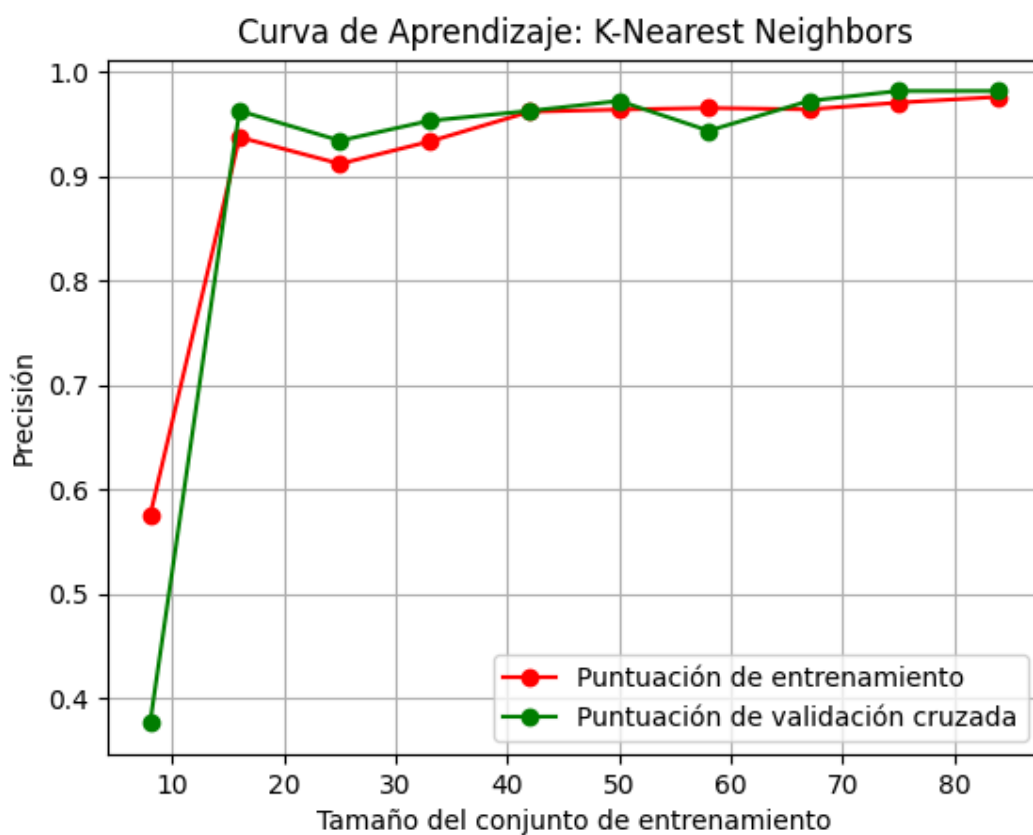
3. Diagnóstico del Grado de Varianza

La varianza mide cómo de bien el modelo puede generalizar. La validación cruzada con 10 folds mostró una precisión promedio de 0.68, lo cual es significativamente menor que la precisión en los conjuntos de validación y prueba (alrededor de 0.96). Esto indica una ligera varianza debido a la variabilidad en los datos, aunque no hay un problema significativo de overfitting.

4. Diagnóstico del Nivel de Ajuste (Underfitting/Overfitting)

- Underfitting (subajuste): Dado que el error de entrenamiento es bajo (0.02) y la precisión es alta, no hay signos de underfitting.
- Overfitting (sobreajuste): Aunque la precisión en validación cruzada es menor que en el conjunto de validación/prueba, el gap no es tan grande. El modelo podría estar ligeramente sobreajustado a los datos de entrenamiento, pero no de manera significativa. **Diagnóstico: Ajuste adecuado con posibilidad leve de sobreajuste.**

La curva de aprendizaje muestra que el modelo está comenzando a converger, ya que las diferencias entre los errores de entrenamiento y validación se están reduciendo. Esto sugiere que el modelo no está sobreajustado significativamente, pero que podría ser más útil tener más datos



5. Técnicas de Regularización o Ajuste de Parámetros

Ajuste de Parámetros: Grid Search

Utilizamos Grid Search para ajustar el valor de k , el número de vecinos más cercanos en el modelo KNN. Después de ajustar los hiperparámetros, se encontró que el mejor valor para k es 5. Esto redujo el error en el conjunto de validación, mejorando las métricas:

- Antes del ajuste:
 - Precisión: 0.93
 - Recall: 0.96
 - F1 Score: 0.94

- Después del ajuste
 - Precisión: 0.94
 - Recall: 0.94
 - F1 Score: 0.93

El uso de validación cruzada y la búsqueda de hiperparámetros ayudó a reducir el sobreajuste, manteniendo un alto nivel de precisión.

Regularización y Técnicas de Mejora

Si se quisiera mejorar el modelo aún más, se podrían probar otras técnicas de regularización como la reducción de la dimensionalidad (PCA). Además, incrementar el número de folds en la validación cruzada podría ayudar a hacer el modelo más robusto.