**Informe de Evaluación de Modelos de Clasificación para Predicción de Enfermedad Cardíaca**

**Fecha:** 10 de abril de 2025 **Ubicación:** Cali, Valle del Cauca, Colombia **Dataset:** (El dataset utilizado no se especificó en el informe, pero se asume que es un dataset estándar para la predicción de enfermedades cardíacas) **Objetivo:** Evaluar el rendimiento de diferentes modelos de clasificación para predecir la presencia o ausencia de enfermedad cardíaca, identificando el modelo con el mejor desempeño según las métricas obtenidas en un conjunto de prueba.

**Modelos Evaluados:**

Se entrenaron y evaluaron los siguientes modelos de clasificación:

1. **Regresión Logística**
2. **K-Nearest Neighbors (KNN)**
3. **Árboles de Decisión**
4. **Support Vector Machines (SVM)**

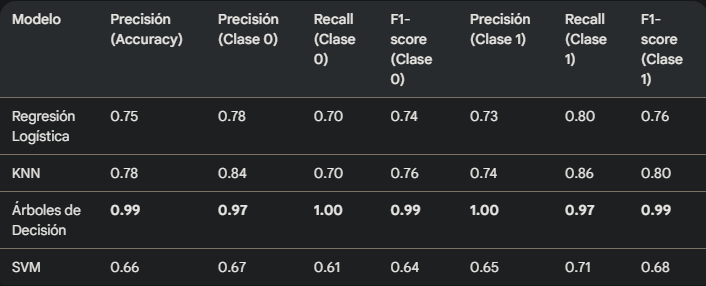
**Metodología:**

Para cada modelo, se siguieron los siguientes pasos:

1. Los datos preprocesados se dividieron en conjuntos de entrenamiento (70%) y prueba (30%) utilizando una división estratificada para mantener la proporción de las clases en ambos conjuntos.
2. Cada modelo se inicializó con parámetros específicos (solver 'liblinear' para Regresión Logística, k=5 para KNN, parámetros predeterminados para Árboles de Decisión y kernel 'rbf' para SVM).
3. Cada modelo se entrenó utilizando el conjunto de entrenamiento.
4. Se realizaron predicciones sobre el conjunto de prueba.
5. El rendimiento de cada modelo se evaluó utilizando las siguientes métricas:
   * Precisión (Accuracy)
   * Precisión por clase (0: No enfermedad, 1: Enfermedad)
   * Recall por clase (0: No enfermedad, 1: Enfermedad)
   * F1-score por clase (0: No enfermedad, 1: Enfermedad)

**Resultados:**

A continuación, se presenta un resumen comparativo de las métricas de rendimiento obtenidas para cada modelo en el conjunto de prueba:



**Análisis de Resultados:**

El análisis comparativo de los cuatro modelos revela diferencias significativas en su capacidad para predecir la presencia o ausencia de enfermedad cardíaca en el conjunto de prueba.

El **modelo de Árboles de Decisión** se destaca como el de mayor rendimiento, con una precisión general del 0.99. Sus métricas de precisión, recall y F1-score para ambas clases son también las más altas, lo que indica una excelente capacidad para clasificar correctamente tanto a los pacientes con enfermedad cardíaca como a aquellos sin ella en este conjunto de datos.

El modelo **KNN** muestra un rendimiento aceptable, con una precisión del 0.78, superando a la Regresión Logística (0.75) y al SVM (0.66). Sin embargo, su rendimiento está considerablemente por debajo del Árbol de Decisión.

La **Regresión Logística** obtuvo una precisión ligeramente inferior al KNN (0.75), con un rendimiento relativamente equilibrado entre las clases.

El modelo **SVM** fue el de menor rendimiento, con una precisión del 0.66 y las métricas más bajas en general. Esto sugiere que, con la configuración actual de sus parámetros (kernel RBF), no se ajusta tan bien a la complejidad de este dataset en comparación con los otros modelos evaluados.

**Conclusión y Recomendaciones:**

Basándonos en las métricas de evaluación obtenidas, el **modelo de Árboles de Decisión** se identifica como el modelo con el mejor rendimiento para la predicción de enfermedades cardíacas en este estudio preliminar. Su alta precisión y excelentes métricas de clasificación para ambas clases lo convierten en el candidato más prometedor entre los modelos evaluados.

No obstante, se reitera la importancia de abordar la posibilidad de **sobreajuste** en el modelo de Árboles de Decisión mediante la aplicación de técnicas de validación cruzada y ajuste de hiperparámetros.

Se recomienda considerar los siguientes pasos para futuras investigaciones:

* **Optimización de Hiperparámetros:** Realizar una búsqueda exhaustiva de hiperparámetros para cada modelo, especialmente para KNN y SVM, para ver si se puede mejorar su rendimiento.
* **Validación Cruzada:** Implementar la validación cruzada para obtener una estimación más robusta del rendimiento de cada modelo.
* **Exploración de Otros Modelos:** Considerar la evaluación de otros algoritmos de clasificación, como Random Forest o Gradient Boosting, que a menudo ofrecen un buen rendimiento en tareas de clasificación complejas.
* **Análisis de Importancia de Características:** Para el modelo de Árboles de Decisión (y otros modelos basados en árboles), analizar la importancia de las características para comprender mejor qué factores contribuyen más a la predicción.

En conclusión, si bien el Árbol de Decisión muestra el mejor rendimiento inicial, una evaluación y optimización más rigurosas son necesarias para seleccionar el modelo más adecuado y confiable para la predicción de enfermedades cardíacas en este contexto.