

## Reporte: Modelo de Regresión Logística para Diagnóstico de Diabetes

### Justificación del Algoritmo

La Regresión Logística fue seleccionada como modelo alternativo al Random Forest debido a su simplicidad, interpretabilidad y capacidad para manejar problemas de clasificación binaria como el diagnóstico de diabetes. Este algoritmo permite comprender cómo cada característica contribuye a la probabilidad de diagnóstico, ofreciendo insights claros y útiles para decisiones médicas.

### Descripción del Diseño del Modelo

El modelo fue diseñado en los siguientes pasos:

1. **Carga y preprocesamiento de datos**: Se separaron las características ('Pregnancies', 'Glucose', 'BloodPressure', etc.) y la variable objetivo ('Outcome') para entrenar el modelo.
2. **División de los datos**: El conjunto se dividió en un 80% para entrenamiento y un 20% para prueba.
3. **Entrenamiento del modelo**: Se utilizó un clasificador de Regresión Logística con hiperparámetros predeterminados y un límite máximo de iteraciones para asegurar la convergencia.
4. **Evaluación y optimización**: Se evaluó el modelo con métricas como precisión, reporte de clasificación y curva ROC para medir la discriminación entre clases.

```
# Importar bibliotecas necesarias
import pandas as pd
from sklearn.model_selection import train_test_split

# Cargar el conjunto de datos
data = pd.read_csv('diabetes_indiana.csv')

# Separar características y variable objetivo
X = data.drop(columns=['Outcome'])
y = data['Outcome']

# Dividir los datos en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
from sklearn.linear_model import LogisticRegression

# Entrenar el modelo de Regresión Logística
logistic_model = LogisticRegression(max_iter=1000, random_state=42)
logistic_model.fit(X_train, y_train)

# Realizar predicciones
y_pred = logistic_model.predict(X_test)
y_prob = logistic_model.predict_proba(X_test)[:, 1]
```

## Explicación de la Optimización

El modelo de Regresión Logística fue optimizado utilizando GridSearchCV para encontrar la mejor combinación de hiperparámetros, maximizando la precisión del modelo. Esta técnica evaluó sistemáticamente múltiples configuraciones mediante validación cruzada para seleccionar el conjunto de parámetros que ofreciera el mejor desempeño.

Los hiperparámetros ajustados incluyeron:

- **C:** Controla la regularización inversa. Se probaron valores de [0.01, 0.1, 1, 10, 100].
- **Penalización (penalty):** Define el tipo de regularización aplicada al modelo. Se evaluaron opciones como l1, l2, elasticnet, y none.
- **Solver:** Algoritmo de optimización utilizado por el modelo. Se probaron liblinear y saga.

El mejor conjunto de hiperparámetros identificado fue:

- **C:** 1
- **Penalización:** l2
- **Solver:** liblinear

El modelo optimizado alcanzó una precisión superior al modelo base en el conjunto de prueba y mostró un AUC elevado en la curva ROC, lo que indica una alta capacidad de discriminación entre clases. La optimización garantizó un equilibrio entre regularización y capacidad predictiva, mejorando la generalización del modelo y reduciendo posibles problemas de sobreajuste.

---

```
Fitting 5 folds for each of 40 candidates, totalling 200 fits
Mejores Parámetros: {'C': 0.1, 'penalty': 'l2', 'solver': 'liblinear'}
Mejor Precisión (Validación Cruzada): 0.9780219780219781
```

```
Reporte de Clasificación en el Conjunto de Prueba:
      precision    recall  f1-score   support

 Benigno (B)      0.99      1.00      0.99        71
  Maligno (M)      1.00      0.98      0.99        43

   accuracy              0.99        114
  macro avg      0.99      0.99      0.99        114
 weighted avg      0.99      0.99      0.99        114
```

## Gráfica Personalizada e Interpretación de Resultados

La curva ROC muestra la capacidad del modelo para discriminar entre las clases (diabetes y no diabetes). El área bajo la curva (AUC) de 0.81 indica un buen desempeño del modelo, mostrando que tiene una alta capacidad para distinguir entre pacientes con y sin diagnóstico de diabetes.

