

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]
```

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.

