

## KNN

O modelo KNN com  $k=5$  aplicado ao dataset de diabetes apresentou desempenho moderado, com Acurácia de 68,83%, Precisão de 69,02%, Recall de 68,83% e F1-Score de 68,92%. O AUC-ROC de 0,7162 indica que o classificador possui uma capacidade razoável de distinguir entre pacientes com e sem diabetes. Esses valores mostram que o modelo consegue identificar corretamente a maioria dos casos, mas ainda deixa escapar uma quantidade considerável de diagnósticos.

```
# AUC-ROC (como é binário, podemos calcular diretamente)
auc_roc = roc_auc_score(y_test, y_pred_proba)

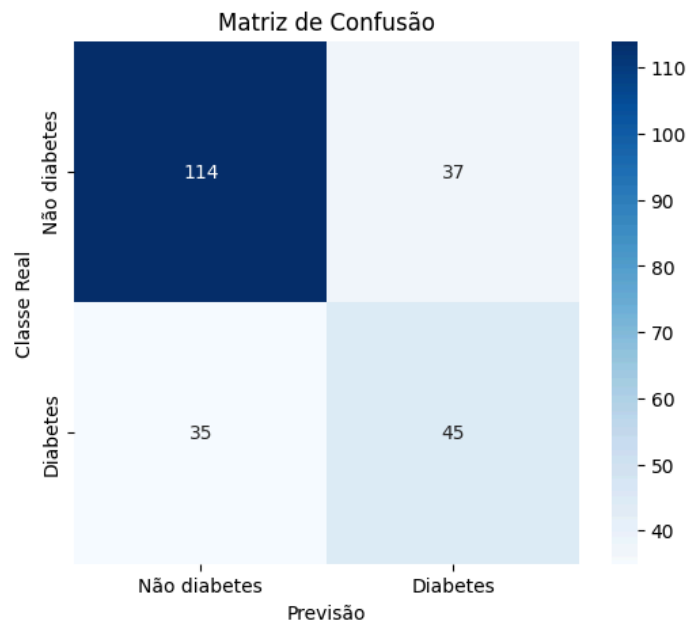
# Acurácia
accuracy = accuracy_score(y_test, y_pred)
# Precision
precision = precision_score(y_test, y_pred, average="weighted")
# Recall
recall = recall_score(y_test, y_pred, average="weighted")
# F1-Score
f1 = f1_score(y_test, y_pred, average="weighted")

# Exibir resultados
print(f"AUC-ROC Score: {auc_roc:.4f}")
print(f"Acurácia: {accuracy:.4f}")
print(f"Precisão: {precision:.4f}")
print(f"Recall: {recall:.4f}")
print(f"F1-Score: {f1:.4f}")
```

[6] ✓ 0.0s

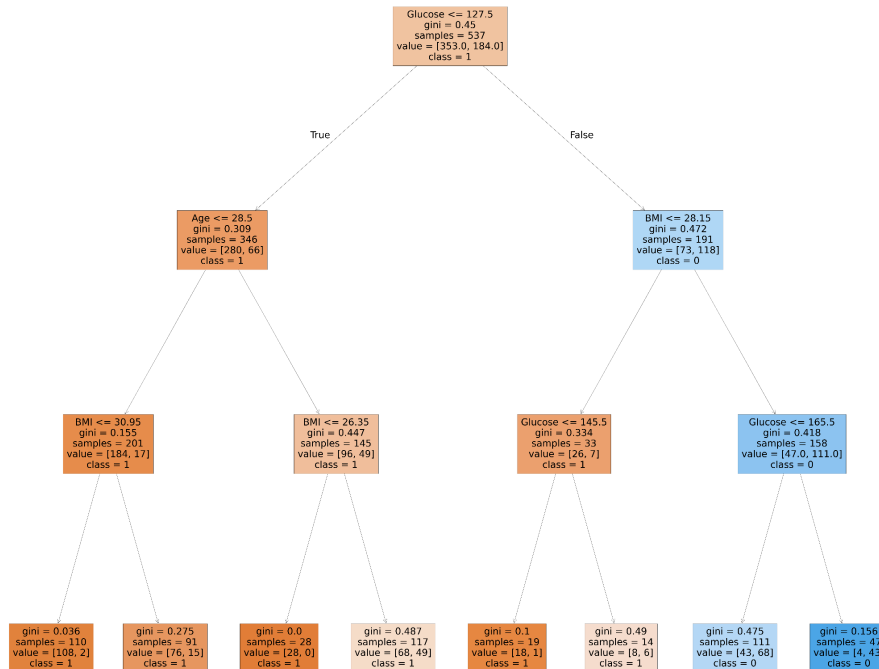
```
... AUC-ROC Score: 0.7162
Acurácia: 0.6883
Precisão: 0.6902
Recall: 0.6883
F1-Score: 0.6892
```

A matriz de confusão evidencia que 114 pacientes foram corretamente classificados como “Não diabéticos” e 45 como “Diabéticos”, enquanto ocorreram 37 falsos positivos e 35 falsos negativos. Esses últimos são especialmente críticos em um contexto médico, já que representam pacientes com diabetes que não seriam diagnosticados.



# Árvore de decisão

O classificador de árvore de decisão apresentou uma acurácia de 74,03% no conjunto de teste, indicando um desempenho razoável na predição de diabetes. Esse resultado mostra que o modelo consegue identificar padrões importantes.



Quanto à importância das variáveis, a glicose (Glucose) teve maior relevância (0,6100), seguida pelo índice de massa corporal (BMI, 0,2660) e idade (Age, 0,1240). As demais variáveis, como número de gestações (Pregnancies, 0,0000), pressão arterial (BloodPressure, 0,0000), espessura da pele (SkinThickness, 0,0000), insulina (Insulin, 0,0000) e função pedigree de diabetes (DiabetesPedigreeFunction, 0,0000), não contribuíram para o modelo. Esses valores indicam que glicose e BMI são os principais preditores para diabetes neste conjunto de dados

```
# Fazer previsões
y_pred = decision_T.predict(X_test)

# Avaliar o modelo
print("Acurácia:", metrics.accuracy_score(y_test, y_pred))
[97] ✓ 0.0s

... Acurácia: 0.7402597402597403

# Importância das variáveis
importances = decision_T.feature_importances_
for i, feature in enumerate(X.columns):
    print(f"feature: {importances[i]:.4f}")
[98] ✓ 0.0s

... Pregnancies: 0.0000
Glucose: 0.6100
BloodPressure: 0.0000
SkinThickness: 0.0000
Insulin: 0.0000
BMI: 0.2660
DiabetesPedigreeFunction: 0.0000
Age: 0.1240
```