



```
In [4]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, ConfusionMatrixDisplay
import matplotlib.pyplot as plt

# --- Step 1: Load dataset from URL ---
url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indian

# Column names as per dataset description
columns = ["Pregnancies", "Glucose", "BloodPressure", "SkinThickness",
           "Insulin", "BMI", "DiabetesPedigreeFunction", "Age", "Outcome"]

df = pd.read_csv(url, header=None, names=columns)

print("Dataset preview:")
print(df.head())

# --- Step 2: Split into features and target ---
X = df.drop("Outcome", axis=1)
y = df["Outcome"]

# --- Step 3: Train-test split ---
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, rand

# --- Step 4: Train Random Forest classifier ---
rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)

# --- Step 5: Predictions & Evaluation ---
y_pred = rf.predict(X_test)
print("\nAccuracy:", accuracy_score(y_test, y_pred))
```

Dataset preview:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
0	6	148	72	35	0	33.6	
1	1	85	66	29	0	26.6	
2	8	183	64	0	0	23.3	
3	1	89	66	23	94	28.1	
4	0	137	40	35	168	43.1	

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

Accuracy: 0.734375

```
In [3]: # --- Step 6: Feature Importance ---
importances = rf.feature_importances_
features = X.columns
indices = importances.argsort()[::-1]
```

```
plt.figure(figsize=(10,6))
plt.title("Feature Importance")
plt.bar(range(X.shape[1]), importances[indices], align="center")
plt.xticks(range(X.shape[1]), [features[i] for i in indices], rotation=45)
plt.ylabel("Importance")
plt.tight_layout()
plt.show()
```

