

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report, confusion_ma

# Load Telco Churn dataset
url = "https://raw.githubusercontent.com/IBM/telco-customer-churn-on-icp4d/master"
data = pd.read_csv(url)

# Clean data
data.drop(columns=['customerID'], inplace=True)
data['TotalCharges'] = pd.to_numeric(data['TotalCharges'], errors='coerce')
data.dropna(inplace=True)

# Encode categorical features
data_encoded = pd.get_dummies(data, drop_first=True)

# Split features and target
X = data_encoded.drop('Churn_Yes', axis=1)
y = data_encoded['Churn_Yes']

# Train/test split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# Train Random Forest
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predict and evaluate
y_pred = model.predict(X_test)
print(f"Accuracy: {accuracy_score(y_test, y_pred):.4f}\n")
print(classification_report(y_test, y_pred))

# Confusion matrix visualization
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6,4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.title("Confusion Matrix")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.tight_layout()
plt.show()

# Feature importance plot
importances = pd.Series(model.feature_importances_, index=X.columns)
top10 = importances.sort_values(ascending=False).head(10)
plt.figure(figsize=(8,6))
sns.barplot(x=top10.values, y=top10.index)
plt.title("Top 10 Feature Importances")

```

```
plt.xlabel("Importance")  
plt.ylabel("Feature")  
plt.tight_layout()  
plt.show()
```

🔗 Accuracy: 0.7854

	precision	recall	f1-score	support
False	0.83	0.90	0.86	1033
True	0.63	0.48	0.54	374
accuracy			0.79	1407
macro avg	0.73	0.69	0.70	1407
weighted avg	0.77	0.79	0.78	1407

