```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix

# Load the dataset
df = pd.read_csv(r"C:\Users\ADMIN\Desktop\archive (5)\KaggleV2-May-
2016.csv")
# Clean and preprocess
df['ScheduledDay'] = pd.to_datetime(df['ScheduledDay'])
df['AppointmentDay'] = pd.to_datetime(df['AppointmentDay'])
df['AppointmentWeekday'] = df['AppointmentDay'].dt.weekday
df['ScheduledToAppointmentDays'] = (df['AppointmentDay'] -
df['ScheduledDay']).dt.days

# Fix invalid age values, etc.
df = df[df['Age'] >= 0]

# Convert target column
df['No-show'] = df['No-show'].map({'No': 0, 'Yes': 1})

features = ['Age', 'Scholarship', 'Hipertension', 'Diabetes',
            'Alcoholism', 'SMS_received', 'AppointmentWeekday',
'ScheduledToAppointmentDays']
X = df[features]
y = df['No-show']

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)


from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier(max_depth=5, random_state=42)
model.fit(X_train, y_train)

DecisionTreeClassifier(max_depth=5, random_state=42)

y_pred = model.predict(X_test)
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))

[[17705    10]
 [ 4382     9]]
              precision    recall  f1-score   support

           0       0.80      1.00      0.89     17715
           1       0.47      0.00      0.00      4391

    accuracy                           0.80     22106
   macro avg       0.64      0.50      0.45     22106
```

```
weighted avg       0.74      0.80      0.71     22106


df.to_csv(r"C:\Users\ADMIN\Desktop\cleaned_appointments.csv",
index=False)
```