import pandas as pd

import pickle

from sklearn.pipeline import Pipeline

from sklearn.compose import ColumnTransformer

from sklearn.preprocessing import OneHotEncoder, LabelEncoder

from sklearn.model\_selection import train\_test\_split

from sklearn.ensemble import RandomForestClassifier

# Load data

df = pd.read\_csv("Tourism\_Master.csv")

df.dropna(inplace=True)

# Features and target

X = df[["Continent", "Region", "VisitSeason", "AttractionType"]]

y = df["VisitModeLabel"]

# Encode target

le = LabelEncoder()

y\_encoded = le.fit\_transform(y)

# Save label encoder for streamlit

with open("label\_encoder.pkl", "wb") as f:

    pickle.dump(le, f)

# Preprocess categorical features

preprocessor = ColumnTransformer([

    ("cat", OneHotEncoder(handle\_unknown='ignore'), X.columns.tolist())

])

# Build pipeline

pipeline = Pipeline([

    ("preprocessor", preprocessor),

    ("clf", RandomForestClassifier(n\_estimators=100, random\_state=42))

])

# Train-test split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y\_encoded, test\_size=0.2, random\_state=42)

pipeline.fit(X\_train, y\_train)

# Save model

with open("classification\_model.pkl", "wb") as f:

    pickle.dump(pipeline, f)

print("✅ Classification model trained and saved.")