GITHUB URL: https://github.com/MsFT100/Road\_analysis\_assignment

# Written By Lennox Mathew Omoiti

import os

import pandas as pd

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

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

import joblib

# Step 1: Check the current working directory

print("Current working directory:", os.getcwd())

# Step 2: load and Understand the Dataset

file\_path = 'c:/Users/hp480/OneDrive/Desktop/Road\_analysis/data/road\_accident.csv'

# Check file content

with open(file\_path, 'r') as file:

    content = file.read()

    print("File content:")

    print(content)

try:

    df = pd.read\_csv(file\_path)  # Use absolute path if necessary

    print(df.head())

    print(df.describe())

    # Print actual column names

    print("Column names:", df.columns.tolist())

    # Strip any leading/trailing spaces from column names

    df.columns = df.columns.str.strip()

    # Identify the dependent and independent variables

    # Example independent variables (you may need to adjust based on your actual dataset)

    X = df[['speed', 'weather', 'road\_condition', 'vehicle\_type']]

    y = df['accident\_severity']

     # Step 3: Create the Linear Regression Model

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

    model = LinearRegression()

    model.fit(X\_train, y\_train)

    y\_pred = model.predict(X\_test)

    # Evaluate the model

    mse = mean\_squared\_error(y\_test, y\_pred)

    r2 = r2\_score(y\_test, y\_pred)

    print(f'Mean Squared Error: {mse}')

    print(f'R^2 Score: {r2}')

    # Step 4: Save the Model for Future Use

    joblib.dump(model, 'c:/Users/hp480/OneDrive/Desktop/Road\_analysis/models/linear\_regression\_model.pkl')

    # Step 5: Predict Accident Severity for a Hypothetical Set of Independent Variables

    # hypthetical data

    hypothetical\_data = pd.DataFrame({

        'speed': [45],

        'weather': [1],

        'road\_condition': [2],

        'vehicle\_type': [1]

    })

    # Predict accident severity

    predicted\_severity = model.predict(hypothetical\_data)

    print(f'Predicted Accident Severity: {predicted\_severity[0]}')

    # Step 6: Explain the Benefits in Underdeveloped Countries

    benefits = """

    A linear regression model for predicting road accident severity can help in underdeveloped countries by:

    1. Identifying High-Risk Factors: Understanding which factors (e.g., high speeds, poor road conditions) contribute most to severe accidents.

    2. Implementing Preventive Measures: Authorities can focus on improving road conditions, enforcing speed limits, and raising awareness about safe driving practices.

    3. Resource Allocation: Efficiently allocating resources like ambulances and traffic police to areas with high predicted accident severity.

    4. Policy Making: Assisting policymakers in creating data-driven regulations to enhance road safety.

    """

    print(benefits)

except FileNotFoundError:

    print("CSV file not found. Please check the file path and name.")

except pd.errors.EmptyDataError:

    print("CSV file is empty or not properly formatted.")

except KeyError as e:

    print(f"KeyError: {e}. Please check if the column names are correct.")