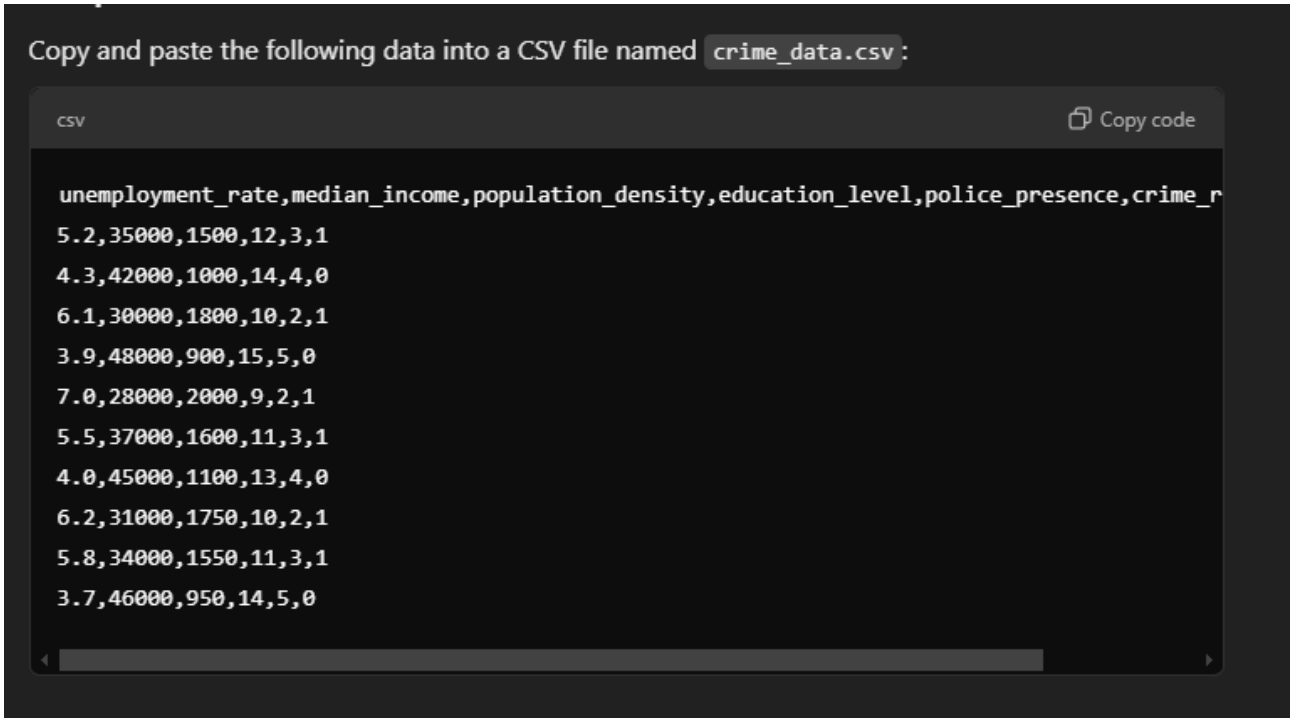
Assignment -1

TOPIC- SDG-16 Peace, Justice, and Strong Institutions: Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable, and inclusive institutions at all levels.

NAME – KRRISHIKA TANEJA
SAP ID – 500120536
ROLL NO – R2142230241
BATCH – 12

dataset named crime_data.csv with features
like unemployment_rate, median_income, population_density, education_level,
and police_presence, and a target variable crime_rate_category



DATA SET CSV

Objective

The primary objective of this project is to develop a machine learning model that can classify areas as "High Risk" or "Low Risk" for crime based on socioeconomic and demographic factors. This model can help law enforcement agencies and policymakers allocate resources effectively to high-risk areas, ultimately contributing to safer communities and supporting SDG-16.

Methodology

1. **Data Collection and Preparation**:
    - Use a dataset (`crime_data.csv`) containing features that impact crime rates, including:
        - **unemployment_rate**: The unemployment percentage in the area.
        - **median_income**: Average income level.
        - **population_density**: Number of people per square mile.
        - **education_level**: Average years of education.
        - **police_presence**: Number of police officers per 1000 people.
    - The target variable, `crime_rate_category`, indicates "High Risk" (1) or "Low Risk" (0) for crime.
    - Handle missing values if any, and preprocess categorical features (if applicable).

Project Description

This project supports **SDG-16: Peace, Justice, and Strong Institutions** by developing a machine learning model to predict crime rates in different areas. Using data on socioeconomic factors, past crime statistics, police presence, and demographic information, the model identifies areas at higher risk of crime. A Random Forest classifier was used to predict crime levels, while interpretability techniques like **LIME** were applied to explain which factors (such as unemployment rate or education level) most influence crime predictions. This model aims to assist policymakers and law enforcement in allocating resources effectively, fostering safer and more just communities.

Data Preprocessing

```python
# Import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix
from imblearn.over_sampling import RandomOverSampler
import joblib

# Step 1: Load Dataset
data = pd.read_csv('/content/crime_data.csv')
print("First 5 rows of the dataset:")
print(data.head())

# Step 2: Preprocess the Data
# Separate features and target variable
X = data.drop(columns=['crime_rate_category'])  # Assuming 'crime_rate_category' is the target column
y = data['crime_rate_category']

# Handle class imbalance using RandomOverSampler
ros = RandomOverSampler(random_state=42)
X_res, y_res = ros.fit_resample(X, y)

# Split data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X_res, y_res, test_size=0.2, random_state=42)

# Scale the features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Save the scaler for future use in the Flask API
joblib.dump(scaler, 'scaler.joblib')
```

```
First 5 rows of the dataset:
   unemployment_rate  median_income  population_density  education_level  \
0                5.2          35000                1500               12
1                4.3          42000                1000               14
2                6.1          30000                1800               10
3                3.9          48000                 900               15
4                7.0          28000                2000                9
```

```python
X_test = scaler.transform(X_test)

# Save the scaler for future use in the Flask API
joblib.dump(scaler, 'scaler.joblib')
```

```
First 5 rows of the dataset:
   unemployment_rate  median_income  population_density  education_level  \
0                5.2          35000                1500               12
1                4.3          42000                1000               14
2                6.1          30000                1800               10
3                3.9          48000                 900               15
4                7.0          28000                2000                9

   police_presence  crime_rate_category
0                3                    1
1                4                    0
2                2                    1
3                5                    0
4                2                    1
['scaler.joblib']
```

```python
# Step 3: Train the Model
rf_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train, y_train)

# Cross-validation to check model reliability
cv_scores = cross_val_score(rf_model, X_train, y_train, cv=5, scoring='accuracy')
print(f"Cross-Validation Accuracy Scores: {cv_scores}")
print(f"Mean Cross-Validation Accuracy: {cv_scores.mean()}")

# Save the trained model for Flask deployment
joblib.dump(rf_model, 'crime_risk_model.joblib')
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_split.py:776: Use
  warnings.warn(
Cross-Validation Accuracy Scores: [1. 1. 1. 1. 1.]
Mean Cross-Validation Accuracy: 1.0
['crime_risk_model.joblib']
```

```python
# Step 4: Model Evaluation
# Make predictions on the test set
predictions = rf_model.predict(X_test)

# Calculate evaluation metrics
accuracy = accuracy_score(y_test, predictions)
precision = precision_score(y_test, predictions)
recall = recall_score(y_test, predictions)
f1 = f1_score(y_test, predictions)
conf_matrix = confusion_matrix(y_test, predictions)

print("Model Evaluation Metrics:")
print(f"Accuracy: {accuracy}")
print(f"Precision: {precision}")
print(f"Recall: {recall}")
print(f"F1 Score: {f1}")
print(f"Confusion Matrix:\n{conf_matrix}")
```

```
Model Evaluation Metrics:
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
Confusion Matrix:
[[2 0]
 [0 1]]
```

Project Title: **Crime Risk Prediction Using Machine Learning**

Brief Description:

This project supports **SDG-16: Peace, Justice, and Strong Institutions** by developing a machine learning model to classify areas as "High Risk" or "Low Risk" for crime based on factors like unemployment rate, median income, population density, education level, and police presence. Using a Random Forest model, the project aims to help law enforcement and policymakers identify high-risk areas, enabling more effective resource allocation and contributing to safer communities. Interpretability techniques such as LIME provide insights into key factors affecting crime risk, ensuring transparency and practical use in real-world applications.