Key concepts and components of the project

Data Collection and Storage:

The project initially involves collecting and storing data related to crimes. The data includes attributes such as name, age, drug test results, obedience, emotion score, confidence score, consistency score, and gender.

Data Preprocessing:

Data preprocessing is essential to prepare the data for model training. In this project, preprocessing involves handling missing values, scaling numeric features, and one-hot encoding categorical features.

Machine Learning Model:

The core of the project involves training a machine learning model, specifically a logistic regression model. This model is trained on the pre-processed data to make predictions about whether an individual is involved in a crime.

Sentiment Analysis:

The project includes a sentiment analysis module that analyses open-ended responses from suspects and witnesses. It calculates scores related to obedience, confidence, emotion, and consistency based on the responses.

Database Integration:

A MySQL database is introduced to store both the training data and the results of predictions. The database facilitates data storage, retrieval, and tracking of model performance.

Training Data Generation:

A "dataProcessing.py" module is responsible for generating training data by analysing the input data, including age, drug test results, obedience, confidence score, emotion score, consistency score, and gender.

Modularization:

The code is organized into modular components for ease of development and maintenance. Modules include data retrieval,

model training, model deployment, sentiment analysis, database management, and prediction.

Prediction Module:

The "predictionInAction.py" module coordinates the various components. It prompts users to enter their information, utilizes the sentiment analysis module to calculate relevant scores, uses the trained model to make predictions, and stores the results in the database.

Model Deployment:

The project demonstrates the deployment of machine learning models. Trained models are saved to disk and loaded for making real-time predictions.

License and Copyright:

The project includes a GNU license agreement template to define the terms and conditions for the use and distribution of the software.

Error Handling:

Exception handling is implemented to catch SQL exceptions when retrieving data from the database to ensure robustness.

Machine Learning Model Evaluation:

The project assesses model performance using metrics such as accuracy, classification report, and confusion matrix. These metrics help evaluate the model's ability to predict criminal behaviour.

Security and Privacy:

The project emphasizes privacy by not sharing suspect information with third parties. It also follows ethical guidelines when collecting and analyzing data.

Feature Engineering:

Feature engineering includes transforming data to make it suitable for machine learning. It involves scaling numeric features, one-hot encoding categorical features, and generating new features like emotion and confidence scores.

Criteria for Decision Making:

The decision on whether a suspect is involved in a crime is based on various factors, including age, gender, drug test results, obedience, confidence score, emotion score, consistency score, and innocence score.

These key concepts and components illustrate a comprehensive project that integrates data collection, machine learning, sentiment analysis, database management, and ethical considerations to predict criminal behaviour. The project is organized into modular components to ensure maintainability and extensibility.