**1. Policiamento preditivo**

**Tipo:** supervisionado

Este script utiliza O RandomForestClassifier, um algoritmo popular para tarefas de classificação. As etapas são as seguintes:

* Carregamento de dados: Carrega seu conjunto de dados em um Pandas DataFrame.
* Seleção de características: Escolha as características relevantes e a variável de destino. Neste caso, estamos a utilizar as coordenadas do local, a data, a descrição do local, o bairro e a área da comunidade como características e o tipo primário de crime como variável de destino.
* Pré-processamento de dados: Converter variáveis categóricas em um formato numérico usando codificação de um ponto.
* Divisão de treinamento-teste: Dividir os dados em conjuntos de treinamento e teste.
* Inicialização e treinamento do modelo: Inicializar o modelo de aprendizado de máquina (classificador Random Forest) e treiná-lo nos dados de treinamento.
* Previsão: Usar o modelo treinado para fazer previsões no conjunto de teste.
* Avaliação: Avaliar o desempenho do modelo usando métricas como precisão e relatório de classificação.

**2. Previsão de prisões**

**Tipo:** supervisionado

Prevê se uma prisão será feita com base em determinados recursos usando a tarefa de classificação binária.

* Carregamento de dados: Carregue seu conjunto de dados em um DataFrame do pandas.
* Seleção de recursos: Escolha os recursos relevantes e a variável de destino. Nesse caso, estamos usando as coordenadas do local, a data, a descrição do local, o distrito e a área da comunidade como recursos e a coluna "Arrest" como variável de destino.
* Pré-processamento de dados: Converter variáveis categóricas em um formato numérico usando codificação de um ponto.
* Divisão de treinamento e teste: Dividir os dados em conjuntos de treinamento e teste.
* Inicialização e treinamento do modelo: Inicializar o modelo de aprendizado de máquina (classificador Random Forest) e treiná-lo nos dados de treinamento.
* Previsão: Use o modelo treinado para fazer previsões no conjunto de teste.
* Avaliação: Avaliar o desempenho do modelo usando métricas como precisão, relatório de classificação e matriz de confusão.

A matriz de confusão fornecerá informações sobre verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, o que pode ser crucial para entender o desempenho do modelo, especialmente em uma tarefa de classificação binária como a previsão de prisões.

**3. Previsão de tendências temporais (ARIMA - AutoRegressive Integrated Moving Average)**

**Tipo:** majoritariamente não-supervisionado

Para a previsão de tendências temporais, os modelos de séries temporais são comumente usados. Um modelo popular e eficaz é o modelo ARIMA (AutoRegressive Integrated Moving Average). O ARIMA é adequado para dados de séries temporais univariadas e pode capturar padrões, tendências e sazonalidade.

* Preparação de dados: Carregar o conjunto de dados e converter a coluna "Date" para o formato datetime.
* Extração de recursos temporais: Extraia as características temporais relevantes, como ano, mês e dia, da coluna "Date".
* Criar séries temporais: Agrupe os dados por data e crie uma série temporal.
* Train-Test Split (divisão de treinamento e teste): Dividir a série temporal em conjuntos de treinamento e teste.
* Initialize ARIMA Model (Inicializar modelo ARIMA): Escolha os parâmetros de ordem (p, d, q) para o modelo ARIMA. Esses parâmetros podem precisar de ajustes com base nas características específicas de seus dados.
* Ajuste e previsão: Ajuste o modelo ARIMA nos dados de treinamento e preveja os valores futuros.
* Avaliação: Avalie o desempenho do modelo usando uma métrica como a raiz do erro quadrático médio (RMSE).
* Plotar resultados: Visualize os dados de treinamento, os dados reais e a previsão ARIMA.

Ajuste o parâmetro de ordem e considere outros ajustes finos com base nas características de seus dados. Lembre-se de que a previsão de séries temporais pode ser complexa e que etapas adicionais, como o ajuste de hiperparâmetros e a validação do modelo, podem ser necessárias para obter os melhores resultados.

**4. Análise geoespacial**

**Tipo:** majoritariamente não-supervisionado

Trabalhar com dados geoespaciais e criar mapas interativos.

* Carregamento de dados: Carregue seu conjunto de dados geoespaciais usando geopandas. Certifique-se de que o conjunto de dados esteja em um formato geoespacial como o GeoJSON.
* Criar mapa base: Crie um mapa base usando folium com um centro e um nível de zoom especificados.
* Clusters de marcadores: Inicialize um MarkerCluster para agrupar marcadores próximos para melhor visualização.
* Adicionar incidentes ao mapa: Itere em seu conjunto de dados e adicione marcadores de cada incidente ao mapa. Você pode personalizar o conteúdo do pop-up e da dica de ferramenta para exibir informações relevantes.
* Salvar mapa: Salve o mapa como um arquivo HTML para análise ou compartilhamento posterior.

Esse script cria um mapa interativo com marcadores agrupados, facilitando a exploração de padrões geoespaciais. Você pode personalizar o mapa ajustando os parâmetros e acrescentando camadas ou informações adicionais.

Lembre-se de substituir "your\_geospatial\_dataset.geojson" pelo caminho real do arquivo de seu conjunto de dados geoespaciais. Além disso, verifique se as colunas como "Latitude" e "Longitude" estão presentes em seu conjunto de dados e se estão formatadas corretamente.

**5. Padrões de crime em áreas comunitárias**

**Tipo:** não-supervisionado

Criação de visualizações ou resumos de padrões de crimes específicos para diferentes áreas da comunidade.

* Carregamento de dados: Carregue seu conjunto de dados em um DataFrame do Pandas.
* Agrupamento e agregação: Agrupe os dados por "Community Area" e "Primary Type" para obter contagens de cada tipo de crime em cada área da comunidade.
* Cálculos: Calcule o número total de incidentes em cada área da comunidade e a porcentagem de cada tipo de crime.
* Plotagem: Use matplotlib para criar um gráfico de barras empilhadas mostrando a distribuição dos tipos de crime selecionados em cada área da comunidade.
* Personalização: Personalize o título, os rótulos e a legenda do gráfico para melhorar a legibilidade.

Esse script fornece uma representação visual da distribuição de crimes em diferentes áreas da comunidade, que pode ser compartilhada com os membros da comunidade para facilitar o envolvimento e as discussões sobre estratégias de prevenção de crimes. Você pode personalizar ainda mais o script com base nas necessidades específicas da comunidade ou criar visualizações adicionais para transmitir diferentes aspectos dos dados.