

**Elysson Alves**  
**Gustavo Almeida**  
**Kalmax dos Santos**

## Entrega 1

### 1. Abordar o Problema e Analisar o Panorama em Geral

#### **Objetivos em termos de negócios**

O objetivo é prever o número de ocorrências de crimes em uma **região** ou **município** para o próximo mês ou ano, a fim de otimizar a alocação de recursos de segurança pública, como policiais, viaturas ou iluminação pública. Isso pode resultar em uma redução da criminalidade e maior segurança para a população.

#### **Como a solução será usada**

A solução será utilizada por gestores públicos, como as secretarias de segurança e prefeituras, para criar estratégias proativas de combate ao crime. Por exemplo:

- Reforçar policiamento em regiões com previsão de alta criminalidade.
- Melhorar a infraestrutura urbana (iluminação, câmeras de vigilância, etc.).

#### **Soluções/alternativas atuais**

- Métodos manuais ou baseados em experiência local para alocação de recursos, como reuniões de planejamento.
- Uso de relatórios estáticos baseados em dados históricos.
- [Policiamento Preditivo](#)

#### **Como o problema deve ser abordado**

- **Tipo de aprendizado:** Aprendizado supervisionado para prever o número de crimes baseado em variáveis históricas.
- **Modo de operação:** Offline inicialmente, com atualizações regulares no modelo conforme novos dados são inseridos.

#### **Como o desempenho será medido**

- **Métricas primárias:**

- **Erro absoluto médio (MAE):** Para avaliar precisão das previsões em valores absolutos.
- **Erro quadrático médio (RMSE):** Para penalizar erros maiores, particularmente indesejáveis, já que grandes diferenças entre os valores previstos e reais têm um impacto maior na métrica.
- **Métricas secundárias:**
  - **R<sup>2</sup>:** Para medir o ajuste do modelo aos dados, variando entre 0 e 1, onde 1 indica que o modelo explica a variabilidade do modelo.

### **A medida de desempenho alinhada com o objetivo do negócio**

Sim, métricas como MAE e RMSE refletem diretamente a precisão do modelo e ajudam a prever com maior “certeza” a quantidade de crimes, essencial para alocação de recursos, por exemplo.

### **Qual seria o desempenho mínimo necessário para alcançar o objetivo do negócio?**

- Um **MAE** inferior à média de crimes por mês em regiões críticas.
  - Por exemplo, se uma região tem em média **100 crimes/mês**, a expectativa é que o erro médio do modelo (MAE) fique abaixo desse número. Isso indica que o modelo está sendo mais útil do que simplesmente usar a média histórica como previsão.
- O **RMSE** deve ser significativamente menor do que a variação padrão dos crimes mensais na região. Isso assegura que o modelo não está apenas replicando a variabilidade dos dados históricos, mas fornecendo previsões úteis.
- O **R<sup>2</sup>** acima de 0,7 indica que mais de 70% da variabilidade dos dados de crime é explicada pelo modelo, um exemplo, quanto maior for, o modelo possui um bom ajuste dos dados, capturando os padrões e tendências

---

## **2. Análise Comparativa e Planejamento**

### **Problemas que podem ser comparados**

- Previsão de demanda em cadeias de suprimentos (similitudes no comportamento temporal).
- Previsão de congestionamentos ou acidentes em cidades, utilizando variáveis ambientais e comportamentais.

## Reutilização de experiências e ferramentas

- Dados similares de estudos sobre criminalidade em outras regiões.

## Expertise Humana

- Colaborações com secretarias de segurança pública podem fornecer conhecimento sobre os dados e os padrões criminais.
- Consultores em estatística, ciência de dados ou criminologia são úteis para análise e validação do modelo.

## Como o problema poderia ser resolvido de forma manual

- Analisar de forma contínua as tendências de crescimento ou redução de crimes em dados históricos.
- 

## 3. Discussões

### Suposições feitas:

1. Dados históricos de criminalidade são confiáveis e completos, eles abordam as reais ocorrências dos anos que foram captadas.
  2. O comportamento criminal é parcialmente previsível com base em padrões históricos.
  3. As intervenções propostas podem impactar positivamente para os índices de criminalidade, visando reduzi-los.
  4. A distribuição da criminalidade não é igualitária, existem crimes que quase não ocorrem em alguns casos temporais.
  5. Ser uma espécie de [Policiamento Preditivo](#), onde é possível prever ocorrências criminais.
  6. Vamos tratar muito a parte de pré-processamento.
- 

## 4. Obter os Dados

- Liste os dados de que você precisa e de quanto precisa.  
**Aproximadamente uma base de dados com mais de 1 milhão de linhas e features que especifiquem os dados de um crime específico e sua**

**frequência e assim aplicar em vários outros crimes em séries temporais diversas.**

- Encontre e documente onde você pode obter esses dados.

**Secretaria de Segurança Pública e Defesa Social e Instituto de Segurança Pública de Minas Gerais.**

- Verifique quanto espaço esses dados ocuparão.

**Depende das séries temporais escolhidas**

- Verifique as obrigações legais e obtenha autorização, se necessário.

**Os dados são públicos.**

- Obtenha permissões de acesso.

**Os dados são públicos.**

- Crie um workspace (com espaço de armazenamento suficiente).

**Google Colab.**

- Obtenha os dados.

**Dados já obtidos com sucesso.**

- Converta os dados em um formato que você possa manipular com facilidade (sem alterar os próprios dados).
- Assegure que as informações confidenciais sejam excluídas ou protegidas (por exemplo, deixando-as anônimas).
- Verifique o tamanho e o tipo de dados (série temporal, amostrados, geográficos etc.)

**Séries temporais diversas**

- Amostre um conjunto de teste, deixe-o de lado e nem coloque a mão nele (sem data snooping, hein?).

## Entrega 2

### 5. Explore os Dados

- Nota: procure um especialista de campo para estas etapas.
  - Crie uma cópia dos dados para exploração (amostragem até um tamanho gerenciável, se necessário).
  - Crie um Jupyter notebook para manter um registro de sua exploração de dados.
  - Estude cada atributo e suas propriedades:
    - Nome.
    - Tipo (categórico, int/float, bounded/unbounded, texto, estruturado etc.).
    - % de valores ausentes.

- Ruído e tipo de ruído (estocástico, outliers, erros de arredondamento etc.).
  - Utilidade para a tarefa.
  - Tipo de distribuição (gaussiana, uniforme, logarítmica etc.).
- Para tarefas de aprendizado supervisionado, identifique o(s) atributo(s)-alvo.
- Visualize os dados.
- Estude as correlações entre os atributos.
- Estude como você resolveria o problema manualmente.
- Identifique as transformações promissoras que você pode querer aplicar.
- Identifique dados extras que seriam úteis (confira o Capítulo 2).
- Documente o que você aprendeu.

## 6. Prepare os Dados

- Notas:
  - Trabalhe em cópias dos dados (mantenha o conjunto de dados original intacto).
  - Escreva funções para todas as transformações de dados que você aplicar, por cinco motivos:
  - Desse modo, você consegue preparar facilmente os dados da próxima vez que obtiver um novo conjunto de dados.
  - Assim, você pode aplicar essas transformações em projetos futuros para:
  - Limpar e preparar o conjunto de teste.
  - Limpar e preparar novas instâncias de dados assim que sua solução estiver em produção.
  - Facilitar o tratamento de suas escolhas de preparação como hiperparâmetros.
- Limpeza de dados:
  - Corrige ou remove outliers (opcional).
  - Preenche os valores ausentes (por exemplo, com zero, média, mediana...) ou elimina suas linhas (ou colunas).
- Seleção de características (opcional):
  - Dropa os atributos que não fornecem informações úteis para a tarefa.
- Feature engineering, quando apropriado:
  - Discretize características contínuas.
  - Decompõe as características (por exemplo, categóricas, data/hora etc.).
  - Adiciona transformações promissoras de características (ex.:  $\log(x)$ ,  $\sqrt{x}$ ,  $x^2$  etc.).
  - Agrega características em novas características promissoras.
- Escalonamento de características

- Padronizar e normalizar as características.