

PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A

Este projeto faz parte dos requisitos da atividade de conclusão da disciplina de Projeto Aplicado II, do Curso de Tecnologia em Ciência de Dados. Realizado no 1º Semestre de 2024, Turma 03A. O objetivo é aplicar os conhecimentos adquiridos ao longo do curso para abordar uma problemática relevante, neste caso, a análise e prevenção da violência contra a mulher, com recorte no estado de Minas Gerais.

Docente: Anderson Adaime de Borba

Discentes: Ana Cláudia de Lima Aleixo
Ana Carolina Martins Chung ti Kam
Maressa Bonassoli
Iasmin Melo

Brasil-2024

Introdução

A violência contra a mulher é uma questão alarmante e persistentemente presente na sociedade brasileira, exigindo abordagens inovadoras para prevenção e intervenção. Neste contexto, este projeto visa analisar os números de violência contra a mulher, com recorte no estado de Minas Gerais, utilizando dados disponíveis no "Sistema Nacional de Estatísticas de Segurança Pública - Municípios". A meta é entender a extensão do problema, e estudar a possibilidade de criar uma ferramenta preditiva que identifique locais mais propensos à violência contra a mulher, proporcionando alertas para a implementação de medidas preventivas eficazes.

Índice

1. Primeira Etapa do Projeto

- 1.1.1. Grupo de trabalho.
- 1.1.2. Premissas do projeto: definição da organização escolhida, área de atuação e apresentação dos dados que serão utilizados (imagem ou texto).
- 1.1.3. Objetivos e metas.
- 1.1.4. Cronograma de atividades (Estimativa).

2. Segunda Etapa do Projeto

- 2.1.1. Definição da linguagem de programação usada no projeto.
- 2.1.2. Análise exploratória da base de dados escolhida.
- 2.1.3. Tratamento da base de dados (Preparação e treinamento).
- 2.1.4. Definição e descrição das bases teóricas dos métodos.
- 2.1.5. Definição e descrição de como será calculada a acurácia.

3. Terceira Etapa do Projeto

- 3.1.1. O método analítico definido na etapa anterior aplicado à base de dados escolhida.
- 3.1.2. Medidas de acurácia, usando os métodos definidos na etapa anterior.
- 3.1.3. Medidas de acurácia, usando os métodos definidos na etapa anterior.
Descrição dos resultados preliminares, apresentando um produto gerado e rascunhando um possível modelo de negócios.
- 3.1.4. Esboço do storytelling.

4. Quarta Etapa do Projeto

- 4.1.1.1. O relatório técnico com todos as fases do projeto.
- 4.1.1.2. A apresentação do *storytelling*. Pode ser um arquivo PPT ou similares.
- 4.1.1.3. Definição do repositório no GitHub com todos os arquivos do projeto.
- 4.1.1.4. Disponibilização de um vídeo com a apresentação no YouTube.

PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A

Este projeto faz parte dos requisitos da atividade de conclusão da disciplina de Projeto Aplicado II, do Curso de Tecnologia em Ciência de Dados. Realizado no 1º Semestre de 2024, Turma 03A. O objetivo é aplicar os conhecimentos adquiridos ao longo do curso para abordar uma problemática relevante, neste caso, a análise e prevenção da violência contra a mulher nos municípios brasileiros.

Docente: Anderson Adaime de Borba

Discentes: Ana Cláudia de Lima Aleixo
Ana Carolina Martins Chung ti Kam

Brasil-2024




Introdução

A violência contra a mulher é uma questão alarmante e persistentemente presente na sociedade brasileira, exigindo abordagens inovadoras para prevenção e intervenção. Neste contexto, este projeto visa analisar os números de violência contra a mulher nos municípios brasileiros, utilizando dados disponíveis no "Sistema Nacional de Estatísticas de Segurança Pública - Municípios". A meta é entender a extensão do problema, e estudar a possibilidade de criar uma ferramenta preditiva que identifique locais mais propensos à violência contra a mulher, proporcionando alertas para a implementação de medidas preventivas eficazes.

Exigências da atividade do A1 - Aplicando Conhecimento (atividade em grupo)

Para atender aos requisitos da primeira aula, compartilhamos as informações essenciais para o desenvolvimento do projeto:

1. Grupo de Trabalho

- Membros da equipe:
 -  Ana Claudia de Lima Aleixo
 -  Ana Carolina Martins Chung ti Kam
 -  Maressa Bonassoli

2. Premissas do Projeto

- **Empresa:** Instituição de segurança pública, uma ONG ou uma empresa de consultoria especializada em segurança.
- **Área de Atuação:** Análise de estatísticas de segurança pública com base nos dados do Sistema Nacional de Estatísticas de Segurança Pública - Municípios.
- **Apresentação dos Dados:** Os dados serão obtidos em texto e números, a partir do conjunto disponível em [<https://dados.gov.br/dados/conjuntos-dados/sistema-nacional-de-estatisticas-de-seguranca-publica>] (<https://dados.gov.br/dados/conjuntos-dados/sistema-nacional-de-estatisticas-de-seguranca-publica>).
- <https://www.gov.br/mdh/pt-br/ondh/painel-de-dados>
- <https://www.gov.br/mdh/pt-br/ondh/painel-de-dados/segundo-semester-de-2023>

3. Objetivos e Metas

▪ **Objetivos Gerais**

- a) Desenvolver um modelo de ciência de dados para análise de segurança pública.
- b) Aumentar a eficiência na classificação e análise de dados de segurança pública.
- c) Melhorar a precisão das previsões sobre o comportamento de segurança nos municípios.
- d) Metas Específicas
 - (a) Coletar um conjunto de dados representativo do Sistema Nacional de Estatísticas de Segurança Pública até o final de fevereiro de 2024.
 - (b) Realizar o pré-processamento dos dados, incluindo limpeza e normalização, até meados de março.
 - (c) Desenvolver e treinar um modelo inicial para prever tendências de segurança até o final de abril.
 - (d) Atingir uma taxa de acurácia mínima de X% na previsão de eventos de segurança até meados de maio.
 - (e) Realizar ajustes no modelo com base nos resultados da avaliação até o final de abril.
- e) Implementar o modelo no ambiente de produção até meados de maio.
- f) Preparar uma documentação abrangente do projeto até o final de maio.
- g) Realizar uma apresentação eficaz do projeto até 24 de maio, comunicando os resultados e conclusões de maneira clara.

4. Cronograma de Atividades

CRONOGRAMA DE ATIVIDADES		
FASE	ATIVIDADE	PRAZO
1	Coleta de Dados	8 fev. - 28 fev
2	Pré-processamento de Dados	1 mar - 14 mar
3	Desenvolvimento e Treinamento do Modelo	15 mar - 18 abr
4	Avaliação e Ajustes do Modelo	19 abr. - 30 abr
5	Implementação no Ambiente de Produção	1 mai. - 15 mai.
6	Documentação e Preparação para Apresentação	16 mai. - 22 mai
7	Apresentação do Projeto	23 mai. - 24 mai.



PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A

Este projeto faz parte dos requisitos da atividade de conclusão da disciplina do Projeto Aplicado II, do Curso de Tecnologia em Ciência de Dados. Realizado no 1º Semestre de 2024, Turma 03A. O objetivo é aplicar os conhecimentos adquiridos ao longo do curso para abordar uma questão relevante, neste caso, a análise e prevenção da violência contra a mulher, com recorte no estado de Minas Gerais.

INTRODUÇÃO

A violência contra a mulher é uma questão alarmante e persistentemente presente na sociedade brasileira, exigindo abordagens inovadoras para prevenção e intervenção. Neste contexto, este projeto visa analisar os números de violência contra a mulher, com recorte no estado de Minas Gerais, utilizando dados disponíveis no "Sistema Nacional de Estatísticas de Segurança Pública - Municípios". A meta é entender a extensão do problema, e estudar a possibilidade de criar uma ferramenta preditiva que identifique locais mais propensos à violência contra a mulher, proporcionando alertas para a implementação de medidas preventivas eficazes.

Exigências da atividade do A2 - Aplicando Conhecimento (atividade em grupo)

Para atender aos requisitos da segunda aula, compartilhamos as informações essenciais para o desenvolvimento do projeto:

PROJETO-APLICADO-II-Grupo-Cassandra's 🚀

- INTEGRANTES
 - Ana Lima Aleixo
 - Maressa Bonassoli
 - Ana Chung ti Kam
 - lasmin Melo

FASE II DO PROJETO APLICADO II

Na Fase II do Projeto, iremos cumprir os seguintes requisitos para a entrega da Atividade II:

- 1. **Gerenciamento de Código**
 - Definiremos as bibliotecas (pacotes) da linguagem Python que serão utilizadas.
 - Realizaremos a implantação de um repositório no Github para iniciar a execução colaborativa do trabalho.
- 2. **Análise da Base de Dados**
 - Definiremos a base de dados a ser utilizada.
 - Realizaremos uma análise exploratória para compreender melhor os dados disponíveis.
- 3. **Tratamento da Base de Dados**
 - Prepararemos e trataremos a base de dados para garantir sua adequação ao propósito do projeto.
- 4. **Fundamentação Teórica**
 - Definiremos e descreveremos as bases teóricas dos métodos analíticos que serão aplicados.
- 5. **Avaliação da Precisão**
 - Descreveremos como será calculada a acurácia do modelo, fornecendo transparência sobre os critérios de avaliação.

1.0 DEFINIÇÃO DA LINGUAGEM E BIBLIOTECA

- Iniciaremos com o uso das seguintes bibliotecas da linguagem Python.
 - import pandas as pd
 - import matplotlib.pyplot as plt
 - import seaborn as sns
 - import display
- GitHub do Projeto
<https://github.com/AnaAleixo/PROJETO-APLICADO-II-Grupo-Cassandra-s>

2.0 ANALISE DA BASE DE DADOS

- Analisaremos as seguintes bases de dados
[/content/violencia_domestica_2014.csv](#) [/content/violencia_domestica_2015.csv](#) [/content/violencia_domestica_2016.csv](#)
[/content/violencia_domestica_2017.csv](#) [/content/violencia_domestica_2018-1.csv](#) [/content/violencia_domestica_2019.csv](#)
[/content/violencia_domestica_2020.csv](#) [/content/violencia_domestica_2021.csv](#) [/content/violencia_domestica_2022.csv](#)
[/content/violencia_domestica_2023.csv](#)

3.0 TRATAMENTO DAS BASES DE DADOS

Iniciamos fazendo uma limpeza nos dados, eliminando linhas em branco e caracteres estranhos aos dados.

Em seguida combinamos os arquivos CSV em um único arquivo para iniciarmos a Análise Exploratória dos dados

3.1 Instalando as Bibliotecas

```
!pip install pandas

Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (1.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
```

3.2 Combinando os arquivos em um único DataFrame.

Combinamos cada arquivo CSV, utilizando o ponto e vírgula como delimitador para combinar todos DataFrame em um único DataFrame chamado dados_combinados. Em seguida, exportamos o novo DataFrame para um arquivo CSV chamado violencia_domestica_combinado.csv.

```
import pandas as pd

# Lista com os caminhos dos arquivos CSV
caminhos_arquivos = [
    '/content/violencia_domestica_2014.csv',
    '/content/violencia_domestica_2015.csv',
    '/content/violencia_domestica_2016.csv',
    '/content/violencia_domestica_2017.csv',
    '/content/violencia_domestica_2018-1.csv',
    '/content/violencia_domestica_2019.csv',
    '/content/violencia_domestica_2020.csv',
    '/content/violencia_domestica_2021.csv',
    '/content/violencia_domestica_2022.csv',
    '/content/violencia_domestica_2023.csv'
]

# Lista para armazenar DataFrames de cada arquivo CSV
dfs = []

# Itera sobre os arquivos CSV e os combina
for caminho_arquivo in caminhos_arquivos:
    try:
        # Utilizamos o ponto e vírgula como delimitador
        df = pd.read_csv(caminho_arquivo, delimiter=';')
        dfs.append(df)
        print(f'Arquivo {caminho_arquivo} lido com sucesso.')
    except Exception as e:
        print(f'Erro ao ler o arquivo {caminho_arquivo}: {e}')

# Combina os DataFrames em um único DataFrame
dados_combinados = pd.concat(dfs, ignore_index=True)

# Caminho para o novo arquivo CSV combinado
caminho_saida = '/content/violencia_domestica_combinado.csv'

try:
    # Exporta o DataFrame combinado para um arquivo CSV
    dados_combinados.to_csv(caminho_saida, index=False)
    print(f'Arquivo combinado gerado com sucesso em {caminho_saida}.')
except Exception as e:
    print(f'Erro ao gerar o arquivo combinado: {e}')

Arquivo /content/violencia_domestica_2014.csv lido com sucesso.
Arquivo /content/violencia_domestica_2015.csv lido com sucesso.
Arquivo /content/violencia_domestica_2016.csv lido com sucesso.
Arquivo /content/violencia_domestica_2017.csv lido com sucesso.
Arquivo /content/violencia_domestica_2018-1.csv lido com sucesso.
Arquivo /content/violencia_domestica_2019.csv lido com sucesso.
Arquivo /content/violencia_domestica_2020.csv lido com sucesso.
Arquivo /content/violencia_domestica_2021.csv lido com sucesso.
Arquivo /content/violencia_domestica_2022.csv lido com sucesso.
Arquivo /content/violencia_domestica_2023.csv lido com sucesso.
Arquivo combinado gerado com sucesso em /content/violencia_domestica_combinado.csv.
```

3.2 Fazendo Análises Exploratórias na Base de Dados

Nessa fase iremos explorar os dados com o objetivo de identificar padrões, tendências e possíveis anomalias nos dados. Vamos usar visualizações gráficas (gráficos, histogramas, mapas etc.) para entender a distribuição e relações entre as variáveis e considerar questões como sazonalidade, correlações entre variáveis e outros insights relevantes.

- Começamos visualizando as primeiras linhas do DataFrame do novo arquivo gerado para ter uma ideia inicial dos dados.

```
import pandas as pd
from IPython.display import display, HTML

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

# Visualiza as primeiras linhas do DataFrame
# dados_combinados.head()

# Exibe as primeiras linhas do DataFrame em uma tabela HTML
display(HTML(dados_combinados.head().to_html(index=False)))
```


municipio_fato	mes	ano	rmbh	natureza_delito	tentado_consumado	qtde_vitimas
ABADIA DOS DOURADOS	1.0	2014.0	3) Interior de MG	DANO	CONSUMADO	1.0
ABRE-CAMPO	1.0	2014.0	3) Interior de MG	LESAO CORPORAL	CONSUMADO	1.0
ACAIAÇA	1.0	2014.0	3) Interior de MG	AMEACA	CONSUMADO	1.0
AGUA BOA	1.0	2014.0	3) Interior de MG	AMEACA	CONSUMADO	1.0
AGUA BOA	1.0	2014.0	3) Interior de MG	VIAS DE FATO / AGRESSAO	CONSUMADO	1.0

3.3 Análise Descritiva

Iniciaremos fazendo uma análise descritiva, com o objetivo de melhor compreender os dados e assim, identificar padrões preliminares e obter insights antes de avançar para a modelagem de classificação.

> 3.3.1 Estatísticas Descritivas

Calcular as variáveis relevantes, como médias, medianas, desvios padrão, mínimos e máximos. A fim de obter uma visão geral das características dos dados.

↳ 1 célula oculta

▽ 3.4 Contagem de Valores Únicos

Objetivo é obter a contagem de valores únicos para variáveis categóricas, como 'natureza_delito', para entender a diversidade dos incidentes.

```
# Contagem de valores únicos para variáveis categóricas
contagem_categorias = dados_combinados['tentado_consumado'].value_counts()
print(contagem_categorias)
```

```
CONSUMADO          978162
TENTADO             11986
PREENCHIMENTO OPCIONAL    21
Name: tentado_consumado, dtype: int64
```

▽ 3.5 Médias Mensal de Vítimas

```
import pandas as pd
from IPython.display import display, HTML

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

# Lista para armazenar os DataFrames de médias por ano
medias_por_ano = []

# Loop de 2014 a 2023
for ano in range(2014, 2024):
    dados_ano = dados_combinados[dados_combinados['ano'] == ano]
    media_crimes_ano = dados_ano.groupby('mes')['qtde_vitimas'].mean().reset_index()
    # Adiciona uma coluna para indicar o ano
    media_crimes_ano['Ano'] = ano
    medias_por_ano.append(media_crimes_ano)

# Concatena os DataFrames ao longo do eixo das colunas
resultado_final = pd.concat(medias_por_ano, axis=1)

# Exibe o resultado final em uma tabela HTML
display(HTML(resultado_final.to_html(index=False)))
```

mes	qtde_vitimas	Ano	mes	qtde_vitimas	Ano	mes	qtde_vitimas	Ano	mes	qtde_vitimas	Ano	mes	qtde_vitimas	Ano	mes	qtde_vitimas	Ano
1.0	1.523330	2014	1.0	1.479754	2015	1.0	1.444683	2016	1.0	1.444709	2017	1.0	1.460466	2018	1.0	1.477923	2019
2.0	1.475578	2014	2.0	1.453023	2015	2.0	1.460164	2016	2.0	1.434470	2017	2.0	1.423039	2018	2.0	1.448170	2019
3.0	1.478061	2014	3.0	1.468457	2015	3.0	1.439959	2016	3.0	1.455874	2017	3.0	1.453824	2018	3.0	1.464651	2019
4.0	1.470699	2014	4.0	1.455805	2015	4.0	1.450290	2016	4.0	1.462690	2017	4.0	1.427516	2018	4.0	1.444827	2019
5.0	1.475342	2014	5.0	1.429878	2015	5.0	1.434427	2016	5.0	1.441300	2017	5.0	1.415454	2018	5.0	1.421504	2019
6.0	1.464120	2014	6.0	1.428794	2015	6.0	1.405541	2016	6.0	1.417653	2017	6.0	1.419909	2018	6.0	1.415340	2019
7.0	1.455099	2014	7.0	1.413709	2015	7.0	1.455619	2016	7.0	1.411221	2017	7.0	1.429604	2018	7.0	1.389851	2019
8.0	1.468254	2014	8.0	1.441519	2015	8.0	1.432276	2016	8.0	1.426528	2017	8.0	1.431499	2018	8.0	1.432416	2019
9.0	1.477660	2014	9.0	1.466827	2015	9.0	1.446404	2016	9.0	1.419688	2017	9.0	1.444610	2018	9.0	1.449862	2019
10.0	1.490353	2014	10.0	1.467534	2015	10.0	1.466116	2016	10.0	1.466967	2017	10.0	1.462624	2018	10.0	1.456860	2019
11.0	1.498011	2014	11.0	1.470770	2015	11.0	1.429966	2016	11.0	1.438370	2017	11.0	1.432442	2018	11.0	1.470953	2019
12.0	1.469541	2014	12.0	1.471680	2015	12.0	1.440098	2016	12.0	1.455734	2017	12.0	1.452435	2018	12.0	1.468927	2019

▽ 3.6 Médias Aritméticas de Vítimas por Ano

```
import pandas as pd
from IPython.display import display, HTML

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

# Calcula a média aritmética dos meses para cada ano
media_aritmetica_anual = dados_combinados.groupby(['ano', 'mes'])['qtde_vitimas'].mean().reset_index()

# Calcula as médias aritméticas anuais para cada ano
medias_aritmeticas_por_ano = media_aritmetica_anual.groupby('ano')['qtde_vitimas'].mean().reset_index()

# Exibe o DataFrame com as médias aritméticas anuais
print("Médias Aritméticas Anuais:")
display(HTML(medias_aritmeticas_por_ano.to_html(index=False)))
```

Médias Aritméticas Anuais:

ano	qtde_vitimas
2014.0	1.478837
2015.0	1.453979
2016.0	1.442129
2017.0	1.439600
2018.0	1.437785
2019.0	1.445107
2020.0	1.414753
2021.0	1.403863
2022.0	1.374622
2023.0	1.393180

3.7 Média de Vítimas nos 10 anos

```
import pandas as pd

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

# Calcula a média de vítimas para todo o período
media_total = dados_combinados['qtde_vitimas'].mean()

# Exibe a média total
print(f'Média total de vítimas ao longo dos 10 anos: {media_total:.2f}')
```

Média total de vítimas ao longo dos 10 anos: 1.43

3.8 Médias Anuais e Acumulada ao Longo de 10 Anos

```
import pandas as pd

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

# Calcula as médias anuais
medias_anuais = dados_combinados.groupby('ano')['qtde_vitimas'].mean().reset_index()

# Calcula a média total ao longo dos 10 anos
media_total = dados_combinados['qtde_vitimas'].mean()

# Cria um DataFrame com as informações
df_medias = pd.DataFrame({
    'Ano': medias_anuais['ano'].tolist() + ['Média Total'],
    'Média de Vítimas': medias_anuais['qtde_vitimas'].tolist() + [media_total]
})

# Exibe o DataFrame
print(df_medias)
```

	Ano	Média de Vítimas
0	2014.0	1.479253
1	2015.0	1.455071
2	2016.0	1.442619
3	2017.0	1.440350
4	2018.0	1.438509
5	2019.0	1.446432
6	2020.0	1.415549
7	2021.0	1.404275
8	2022.0	1.375063
9	2023.0	1.394057
10	Média Total	1.430635

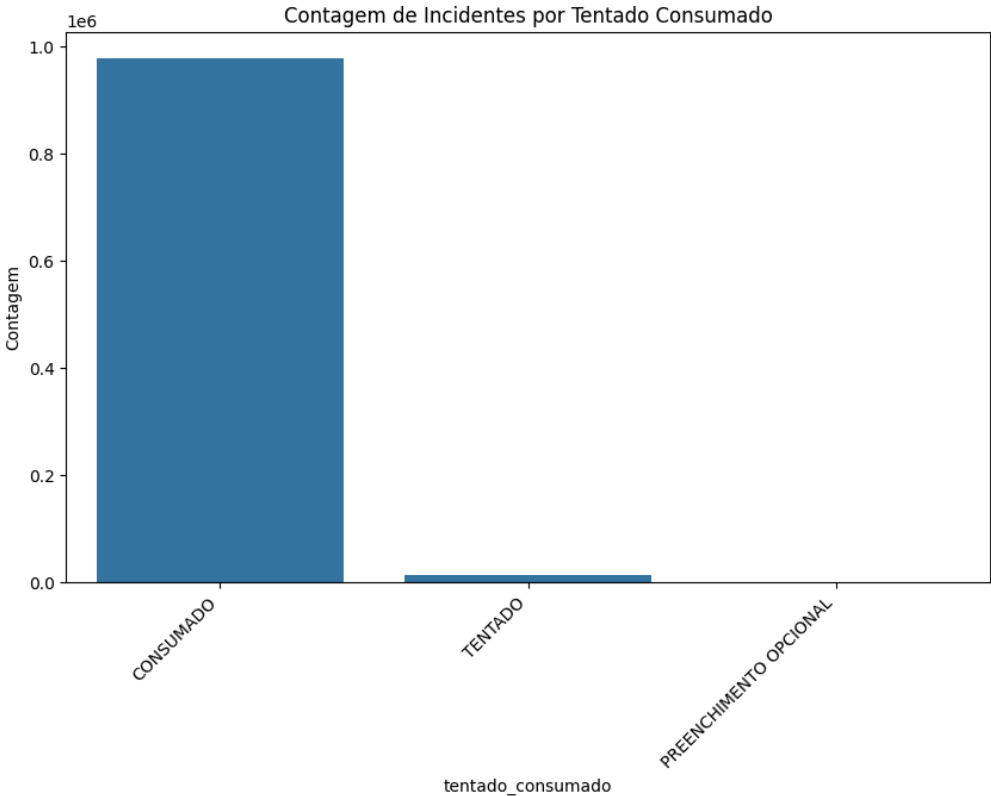
3.9 Visualizações Gráficas

Objetivo visualizar gráficos, como gráficos de barras para contagens de categorias ou gráficos de linha para tendências temporais.

3.10.1 Incidentes Tentado Consumado

```
import matplotlib.pyplot as plt
import seaborn as sns

# Gráfico de barras para contagem de categorias
plt.figure(figsize=(10, 6))
sns.countplot(x='tentado_consumado', data=dados_combinados)
plt.title('Contagem de Incidentes por Tentado Consumado')
plt.xlabel('tentado_consumado')
plt.ylabel('Contagem')
plt.xticks(rotation=45, ha='right')
plt.show()
```



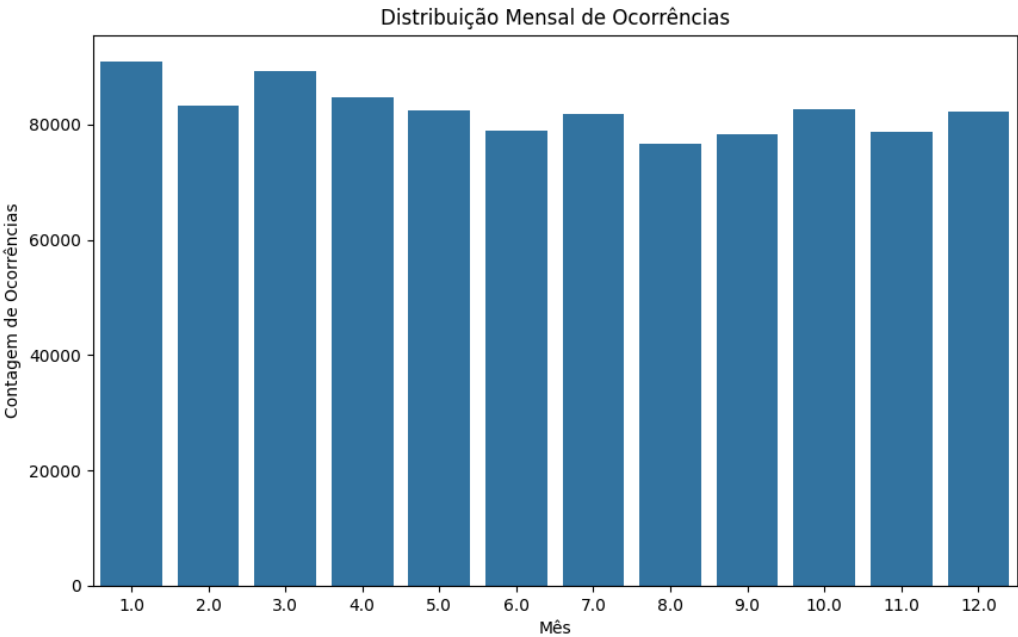
3.10.2 Distribuição Mensal de Ocorrências

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

plt.figure(figsize=(10, 6))
sns.countplot(data=dados_combinados, x='mes')
plt.title('Distribuição Mensal de Ocorrências')
plt.xlabel('Mês')
plt.ylabel('Contagem de Ocorrências')
plt.show()
```



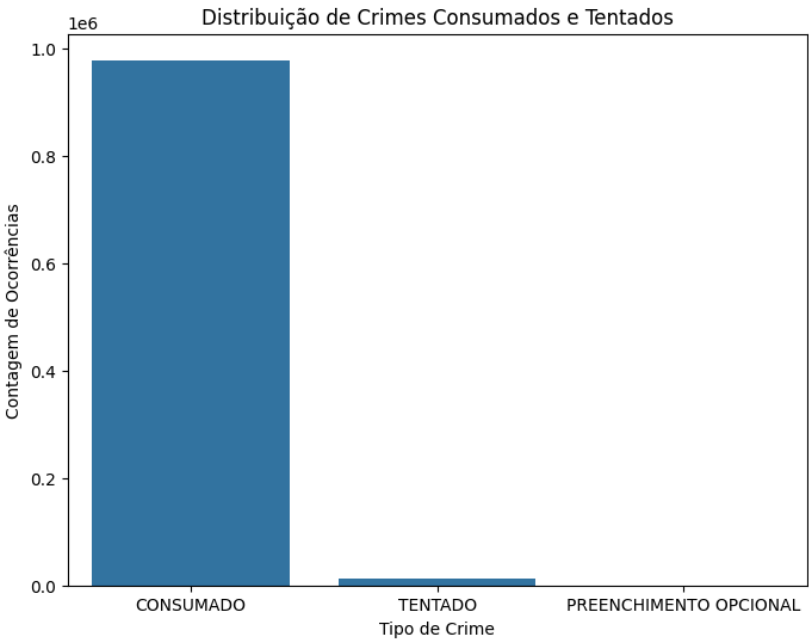
3.10.3 Distribuição de Crimes Consumados e Tentados

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

plt.figure(figsize=(8, 6))
sns.countplot(data=dados_combinados, x='tentado_consumado')
plt.title('Distribuição de Crimes Consumados e Tentados')
plt.xlabel('Tipo de Crime')
plt.ylabel('Contagem de Ocorrências')
plt.show()
```



3.10.4 Médias Anuais de Vítimas de Violência Doméstica

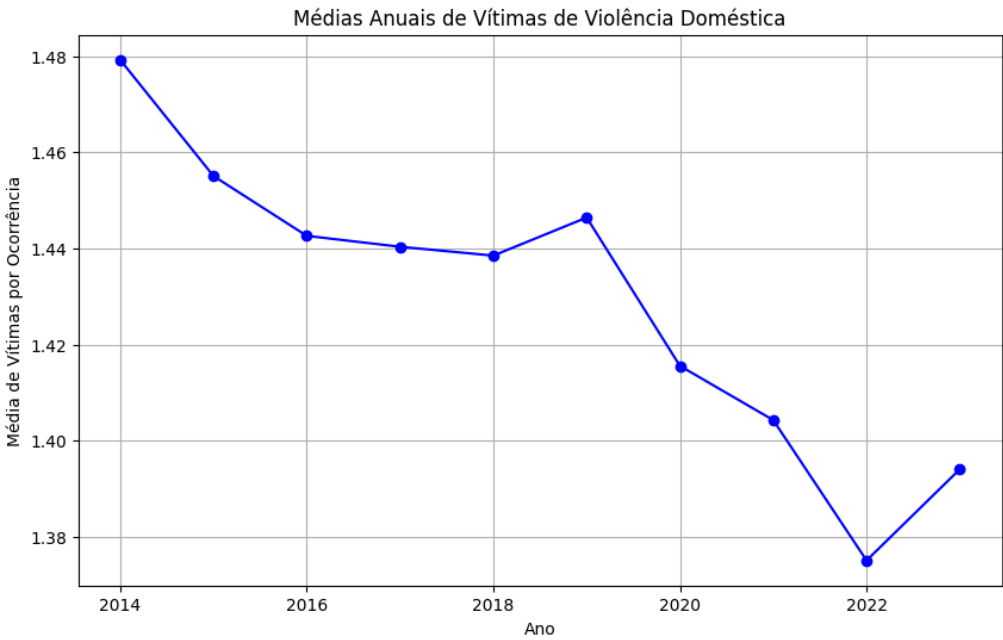
```
import pandas as pd
import matplotlib.pyplot as plt

# Caminho para o arquivo CSV combinado
caminho_combinado = '/content/violencia_domestica_combinado.csv'

# Lê o arquivo CSV
dados_combinados = pd.read_csv(caminho_combinado)

# Calcula a média anual de vítimas
media_vitimas_anual = dados_combinados.groupby('ano')['qtde_vitimas'].mean()

# Plota o gráfico de médias anuais de vítimas
plt.figure(figsize=(10, 6))
media_vitimas_anual.plot(marker='o', linestyle='-', color='b')
plt.title('Médias Anuais de Vítimas de Violência Doméstica')
plt.xlabel('Ano')
plt.ylabel('Média de Vítimas por Ocorrência')
plt.grid(True)
plt.show()
```



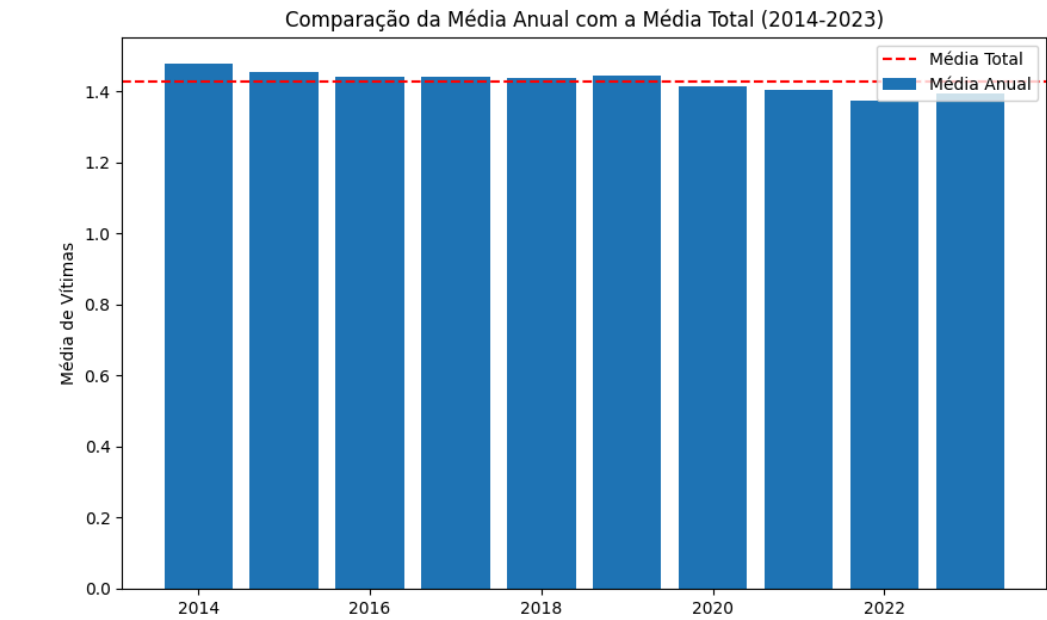
3.10.5 Comparação da Média Anual com a Média Total (2014-2023)

```
import pandas as pd
import matplotlib.pyplot as plt

# Cria um DataFrame com as médias anuais
medias_anuais = pd.DataFrame({
    'Ano': [2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023],
    'Média Anual': [1.478837, 1.453979, 1.442129, 1.439600, 1.437785, 1.445107, 1.414753, 1.403863, 1.374622, 1.393180]
})

# Calcula a média total ao longo dos 10 anos
media_total = medias_anuais['Média Anual'].mean()

# Cria o gráfico de barras agrupadas
plt.figure(figsize=(10, 6))
plt.bar(medias_anuais['Ano'], medias_anuais['Média Anual'], label='Média Anual')
plt.axhline(y=media_total, color='red', linestyle='--', label='Média Total')
plt.title('Comparação da Média Anual com a Média Total (2014-2023)')
plt.xlabel('Ano')
plt.ylabel('Média de Vítimas')
plt.legend()
plt.show()
```



4.0 FUNDAMENTAÇÃO E DEFINIÇÃO DO MÉTODO A SER APLICADO

- Método a ser Aplcado: Analítico de Classificação

A Análise Exploratória dos dados, nas detalhadas informações importantes que nos orientou para as próximas etapas do projeto.

Um ponto que nos chamou atenção foi a média mensal de ocorrências (6.38/m), esse dado nos revela uma distribuição relativamente uniforme ao longo do ano, indicando uma dinâmica estável nas incidências. Notavelmente, os meses de março e setembro se destacam como períodos em que ocorrem os picos mais significativos de ocorrências.

Também, observamos uma predominância de ocorrências em cidades do interior, ressaltando a necessidade de estratégias específicas para abordar questões locais nessas áreas. A diversidade de crimes é impressionante, com aproximadamente 361 tipos distintos de violência contra a mulher identificados, sendo a ameaça o tipo mais frequente. Uma descoberta intrigante desta análise inicial é a proporção significativa de crimes consumados. Esta observação sugere a urgência de estratégias preventivas eficazes para interromper a concretização desses atos violentos.

Outro ponto de destaque é a média de vítimas por incidente, que se estabelece em 1.43. Isso indica predominantemente casos envolvendo uma única vítima.

Essas informações nos forneceram insights valiosos para a escolha do método analítico a ser usado, pois a pequena flutuação das médias do número de vítimas de violência doméstica ao longo de 10 anos, demonstra que as políticas de prevenção, se existirem, não estão sendo efetivas em proteger e diminuir os números de mulheres vítimas de violência. Essa constatação nos direcionou a aplicar o método analítico de classificação, e assim alinhar a condução desse projeto ao objetivo principal que é desenvolver uma ferramenta preditiva que possa auxiliar o sistema de segurança pública no monitoramento preventivo de vítima de violência.

5.0 AVALIAÇÃO DA PRECISÃO- DEFINIÇÃO DO CÁLCULO DA ACURÁCIA

Definição do Percentual da Acurácia: Intervalo entre 50% e 60%.

Não existe um percentual padrão de acurácia, uma vez que o desempenho de um modelo pode variar significativamente, dependendo da natureza específica do problema e dos objetivos do projeto. Além disso, outros fatores, como diferentes cenários e contextos, devem ser criteriosamente considerados durante as análises. Neste contexto, decidimos que uma acurácia situada entre 50% e 60% representam um patamar considerável de precisão. Essa avaliação ganha relevância ao ser contextualizada diante da complexidade e diversidade dos dados relacionados à violência contra a mulher.

Divisão das Bases de Dados: Serão divididas em 70/30, podendo ser 80/20.

Para conduzir a análise preditiva, optaremos pela técnica de avaliação cruzada temporal. Nesse processo, treinaremos o modelo com dados de um período anterior e o testaremos com dados mais recentes, simulando situações de previsão real. A divisão das bases será efetuada na proporção de 70/30, podendo ser ajustada para 80/20 conforme necessário. As bases mais recentes serão reservadas para o conjunto de teste, enquanto as mais antigas serão utilizadas no treinamento do modelo. Essa abordagem visa garantir uma avaliação robusta e realista do desempenho preditivo do modelo ao longo do tempo.



PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A

Este projeto faz parte dos requisitos da atividade de conclusão da disciplina do Projeto Aplicado II, do Curso de Tecnologia em Ciência de Dados. Realizado no 1º Semestre de 2024, Turma 03A. O objetivo é aplicar os conhecimentos adquiridos ao longo do curso para abordar uma questão relevante, neste caso, a análise e prevenção da violência contra a mulher, com recorte no estado de Minas Gerais.

INTRODUÇÃO

A violência contra a mulher é uma questão alarmante e persistentemente presente na sociedade brasileira, exigindo abordagens inovadoras para prevenção e intervenção. Neste contexto, este projeto visa analisar os números de violência contra a mulher, com recorte no estado de Minas Gerais, utilizando dados disponíveis no "Sistema Nacional de Estatísticas de Segurança Pública - Municípios". A meta é entender a extensão do problema, e estudar a possibilidade de criar uma ferramenta preditiva que identifique locais mais propensos à violência contra a mulher, proporcionando alertas para a implementação de medidas preventivas eficazes.

Exigências da atividade do A3 - Aplicando Conhecimento (atividade em grupo)

Para atender aos requisitos da terceira aula, compartilhamos as informações essenciais para o desenvolvimento do projeto:

PROJETO-APLICADO-II-Grupo-Cassandra's 🚀

- INTEGRANTES
 - Ana Lima Aleixo
 - Maressa Bonassoli
 - Ana Chung ti Kam
 - Iasmin Melo

GITHUB DO PROJETO

- <https://github.com/AnaAleixo/PROJETO-APLICADO-II-Grupo-Cassandra-s>

FASE III DO PROJETO APLICADO II

Na Fase III do Projeto, iremos cumprir os seguintes requisitos para a entrega da Atividade III:

- 1.0) Consolidar os resultados do método analítico, definido na etapa anterior, aplicado à base de dados escolhida com padrão.
- 2.0) Aplicar as medidas de acurácia para verificar o desempenho dos métodos definidos na etapa anterior.
- 3.0) Descrever os resultados preliminares, apresentando um produto gerado, e rascunhando um possível modelo de negócios.
- 4.0) Esboço do storytelling.

1.0) CONSOLIDAR OS RESULTADOS DO MÉTODO ANALÍTICO, DEFINIDO NA ETAPA ANTERIOR, APLICADO À BASE DE DADOS ESCOLHIDA COM PADRÃO

✓ 1.1) SEPARANDO AS BASES EM CONJUNTOS DE TREINO E TESTE

```
# Importar as bibliotecas necessárias
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score

# Carregar o DataFrame combinado
df = pd.read_csv("/content/violencia_domestica_combinado.csv")

# Verificar o formato do DataFrame
print("Formato do DataFrame:", df.shape)

# Visualizar as primeiras linhas do DataFrame
print(df.head())
```

Formato do DataFrame: (668942, 7)

	municipio_fato	mes	ano	rmbh	\
0	ABADIA DOS DOURADOS	1.0	2014.0	3)	Interior de MG
1	ABRE-CAMPO	1.0	2014.0	3)	Interior de MG
2	ACAIACA	1.0	2014.0	3)	Interior de MG
3	AGUA BOA	1.0	2014.0	3)	Interior de MG
4	AGUA BOA	1.0	2014.0	3)	Interior de MG

	natureza_delito	tentado_consumado	qtde_vitimas
0	DANO	CONSUMADO	1.0
1	LESAO CORPORAL	CONSUMADO	1.0
2	AMEACA	CONSUMADO	1.0
3	AMEACA	CONSUMADO	1.0
4	VIAS DE FATO / AGRESSAO	CONSUMADO	1.0

1.2) CRIANDO MODELO DE CLASSIFICAÇÃO USANDO O
RANDOMFORESTCLASSIFIER

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.impute import SimpleImputer

# Carregar o DataFrame combinado
df = pd.read_csv("/content/violencia_domestica_combinado.csv")

# Remover colunas de texto que não são importantes para o modelo
X = df.drop(columns=['municipio_fato', 'rmbh', 'natureza_delito', 'tentado_consumado', 'qtde_vitimas'])

# Preencher valores ausentes na coluna 'qtde_vitimas' com zero
df['qtde_vitimas'] = df['qtde_vitimas'].fillna(0)

# Converter 'qtde_vitimas' para tipo inteiro
y = df['qtde_vitimas'].astype(int)

# Dividir os dados em conjuntos de treino e teste (70% treino, 30% teste)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Criar um imputador para preencher os valores ausentes com a média
imputer = SimpleImputer(strategy='mean')

# Preencher os valores ausentes em X_train
X_train_imputed = imputer.fit_transform(X_train)

# Converter a matriz resultante de volta para um DataFrame
X_train_imputed = pd.DataFrame(X_train_imputed, columns=X_train.columns)

# Verificar se há valores ausentes em X_test e preencher, se necessário
if X_test.isnull().any().any():
    X_test_imputed = imputer.transform(X_test)
    X_test_imputed = pd.DataFrame(X_test_imputed, columns=X_test.columns)
else:
    X_test_imputed = X_test.copy()

# Criar e treinar o modelo RandomForestClassifier
model = RandomForestClassifier(random_state=42)
model.fit(X_train_imputed, y_train)

# Prever os valores do conjunto de teste
y_pred = model.predict(X_test_imputed)

# Calcular a acurácia do modelo
accuracy = accuracy_score(y_test, y_pred)
print("Acurácia do modelo:", accuracy)
```

Acurácia do modelo: 0.801468756941593

2.0) APLICAR AS MEDIDAS DE ACURÁCIA PARA VERIFICAR O DESEMPENHO DOS MÉTODOS DEFINIDOS NA ETAPA ANTERIOR

✓ 2.1) APLICAÇÃO DE MEDIDAS DE ACURÁCIA ADICIONAIS

- Recall
- F1-score

```
from sklearn.metrics import recall_score, f1_score

# Calcular o recall
recall = recall_score(y_test, y_pred, average='weighted')

# Calcular o F1-score
f1 = f1_score(y_test, y_pred, average='weighted')

print("Recall do modelo:", recall)
print("F1-score do modelo:", f1)
```

```
Recall do modelo: 0.801468756941593
F1-score do modelo: 0.7131457088697434
```

2.2) EXPLICANDO O MODELO E OS RESULTADOS OBTIDOS

- Começamos carregando os dados do arquivo CSV usando a função `pd.read_csv()` do pandas. Em seguida, realizamos um tratamento nos dados removendo as colunas de texto que não são relevantes para o modelo, mantendo apenas as features essenciais para a classificação. Observamos que a coluna 'qtde_vitimas' continha valores ausentes, optando por preenchê-los com zero utilizando o método `fillna(0)`.
- Dividimos os dados em conjuntos de treino e teste, atribuindo 70% dos dados para treinamento e 30% para teste, utilizando a função `train_test_split()` do scikit-learn. Em seguida, aplicamos o `SimpleImputer` do scikit-learn para preencher os valores ausentes nas features com a média dos dados, garantindo assim que o modelo possa ser treinado e testado adequadamente.
- Criamos uma instância do `RandomForestClassifier` e procedemos ao treinamento do modelo utilizando os dados de treinamento por meio do método `fit()`. Posteriormente, utilizamos o modelo treinado para fazer previsões sobre os dados de teste usando o método `predict()`.
- Para avaliar o desempenho do modelo, calculamos sua acurácia comparando as previsões feitas com os valores reais do conjunto de teste, utilizando a função `accuracy_score()` do scikit-learn. Obtivemos uma acurácia do modelo de aproximadamente 80.15%, o que indica que o modelo classificou corretamente cerca de 80.15% dos casos no conjunto de teste.
- Além disso, buscando uma avaliação mais completa, utilizamos outras métricas como o recall e o F1-score. O recall do modelo foi de aproximadamente 80.15%, indicando sua capacidade de identificar corretamente cerca de 80.15% dos casos positivos (vítimas de violência doméstica) no conjunto de teste. Já o F1-score do modelo foi de aproximadamente 0.713, revelando uma boa harmonia entre precisão e recall.
- Esses resultados sugerem que o modelo apresenta um desempenho sólido, capaz de classificar corretamente tanto os casos positivos quanto os negativos, o que o torna uma opção viável para uso em produção.

3.0) DESCREVER OS RESULTADOS PRELIMINARES, APRESENTANDO UM PRODUTO GERADO, E RASCUNHANDO UM POSSÍVEL MODELO DE NEGÓCIOS

3.1) DESCRIÇÃO DOS RESULTADOS PRELIMINARES

Produto Gerado

Descrição do Produto

- Nosso produto é um modelo preditivo de violência contra a mulher, desenvolvido a partir da análise estatística da média dos casos de violência em **Minas Gerais**. Utilizando técnicas de aprendizado de máquina e baseando-se nos dados fornecidos pelo **"Sistema Nacional de Estatísticas de Segurança Pública - Municípios"**, o modelo tem como objetivo identificar áreas com maior probabilidade de ocorrência de violência contra a mulher. Além disso, busca avaliar a eficácia das políticas de prevenção existentes, proporcionando cenários para o aprimoramento dessas medidas e para a promoção de uma sociedade mais segura e igualitária para as mulheres.

Funcionalidades

- **Previsão através da medição da média de incidentes:** O modelo é capaz de prever através da média de incidência de violência, se as políticas de prevenção estão sendo eficientes, tendo como base em dados históricos e características demográficas, socioeconômicas e geográficas
- **Alertas de Prevenção:** O produto fornece alertas para autoridades policiais, órgãos governamentais e organizações de defesa dos direitos das mulheres sobre os locais identificados como de maior risco, permitindo a implementação de medidas preventivas e intervenções oportunas.

Potenciais Usuários

- **Autoridades Policiais:** Para o direcionamento de recursos e esforços de policiamento preventivo.
- **Governos Estaduais e Municipais:** Para a formulação de políticas públicas e alocação de recursos para prevenção da violência contra a mulher.
- **ONGs e Instituições de Defesa dos Direitos das Mulheres:** Para promover a conscientização e mobilização da sociedade civil na prevenção e combate à violência doméstica.

Benefícios Esperados

- **Redução da Incidência de Violência:** Ao identificar e intervir precocemente em áreas de maior risco, espera-se uma redução na incidência de violência contra a mulher.
- **Melhor Alocação de Recursos:** A alocação eficiente de recursos e esforços para prevenção pode maximizar o impacto das políticas públicas na proteção das mulheres.
- **Conscientização e Mobilização:** O produto pode contribuir para aumentar a conscientização sobre o problema da violência contra a mulher e mobilizar a sociedade para ações de prevenção e apoio às vítimas.

3.2) MODELO DE NEGÓCIO

Segmento de Cliente

- **Autoridades de Segurança Pública**
- **Governos Estaduais e Municipais**
- **ONGs e Instituições de Defesa dos Direitos das Mulheres**

Proposta de Valor

- **Prevenção e Intervenção Precoces:** Capacidade de identificar áreas de maior risco e tomar medidas preventivas antes que ocorram incidentes de violência.
- **Eficiência na Alocação de Recursos:** Otimização da alocação de recursos para prevenção da violência contra a mulher com base em análises preditivas.
- **Conscientização e Engajamento:** Contribuição para a conscientização e mobilização da sociedade civil na prevenção e combate à violência doméstica.

Canais de Distribuição

- **Plataforma Online:** Disponibilização do produto por meio de uma plataforma online acessível a autoridades, governos e organizações interessadas.
- **Parcerias Governamentais:** Estabelecimento de parcerias estratégicas com órgãos governamentais para a distribuição e implementação do produto.

Relacionamento com o Cliente

- **Suporte Técnico:** Oferta de suporte técnico para a implementação e utilização do produto.
- **Treinamento:** Realização de treinamentos para capacitar os usuários na interpretação e utilização dos resultados gerados pelo modelo.
- **Atualizações Regulares:** Fornecimento de atualizações regulares do produto com base em novos dados e avanços tecnológicos.

Fontes de Receita

- **Vendas Diretas:** Venda do produto para clientes interessados, com base em assinaturas ou licenciamento de uso.
- **Parcerias Estratégicas:** Estabelecimento de parcerias estratégicas com governos e organizações para o financiamento e implementação do produto.

Este modelo de negócios tem como objetivo principal fornecer uma solução eficaz e inovadora para a prevenção da violência contra a mulher em Minas Gerais, contribuindo para a construção de uma sociedade mais segura e justa para as mulheres.

4.0) ESBOÇO DO STORYTELLING

4.1) STORYTELLING DO PROJETO

Combatendo a Violência contra a Mulher em Minas Gerais

Era uma vez Minas Gerais, um estado rico em história e cultura, mas também marcado por um problema sério: a violência contra a mulher. Diante desse desafio, um grupo de pesquisadoras e especialistas em Ciência de Dados decidiu unir forças para enfrentar essa questão de frente.

Tudo começou com a **fase I do projeto**, onde mergulharam fundo nos dados disponíveis sobre violência doméstica e de gênero. Elas estudaram relatórios oficiais, analisaram estatísticas e ouviram histórias de sobreviventes. Cada número, cada relato humano, ajudou a pintar um quadro mais claro da situação.

Munidas com o conhecimento adquirido na fase I, passam para a **fase II**, onde exploram soluções tecnológicas. Elas desenvolvem algoritmos sofisticados, criam modelos preditivos e experimentam diferentes abordagens de machine learning. O objetivo é claro: usar a tecnologia para identificar padrões, prever ocorrências e, assim, ajudar na prevenção da violência.

Chega então a **fase III**, onde consolidam os resultados obtidos e partem para a ação. Criam um modelo preditivo de violência contra a mulher, baseado em dados reais e análises estatísticas. Este modelo não só prevê aumento nas ocorrências, mas também fornece insights para autoridades, governos e organizações de defesa dos direitos das mulheres.

O produto final é mais do que uma simples ferramenta. É uma promessa de esperança, para tantas mulheres que sofrem em silêncio. Com ele, as autoridades podem direcionar recursos de forma mais eficaz, os governos podem formular políticas públicas mais assertivas e as ONGs podem ampliar seu alcance.

À medida que o projeto avançar, almejamos que seu impacto seja sentido em todo o estado. A incidência de violência diminua, as políticas de prevenção se tornem mais eficazes e a conscientização da sociedade cresça. E que Minas Gerais caminhe em direção a um futuro mais seguro e igualitário para todas as suas mulheres.

E assim, uma jornada que começou com dados e tecnologia se transforma em uma história de esperança e mudança. Graças ao trabalho árduo e à dedicação daquelas que se recusaram a aceitar a violência como algo inevitável, Minas Gerais irá trilhar um novo caminho, rumo a um futuro onde todas as mulheres possam viver livres do medo e da opressão.

**PROJETO APLICADO II - CURSO
TECNOLOGIA EM CIENCIA DE DADOS
ANO 2024 - 1º SEMESTRE - TURMA 03A**

IV ETAPA DO PROJETO

**Combate a Violência
Contra a Mulher!**

STORYTELLING DO PROJETO



PROJETO APLICADO II

CURSO TECNOLOGIA EM

CIENCIA DE DADOS ANO

2024 - 1º SEMESTRE

TURMA 03A

Este projeto faz parte dos requisitos da atividade de conclusão da disciplina do Projeto Aplicado II, do Curso de Tecnologia em Ciência de Dados. Realizado no 1º Semestre de 2024, Turma 03A. O objetivo é aplicar os conhecimentos adquiridos ao longo do curso para abordar uma questão relevante, neste caso, a análise e prevenção da violência contra a mulher, com recorte no estado de Minas Gerais.

GRUPO CASSANDA'S

- ❖ Ana Cláudia de Lima Aleixo
- ❖ Maressa Bonassoli
- ❖ Ana Chung ti Kam
- ❖ Iasmin Melo



Ana Lima Aleixo



Maressa Bonassoli



Ana_Chung.ti.Kam

Combatendo a Violência contra a Mulher

Era uma vez Minas Gerais, um estado rico em história e cultura, mas também marcado por um problema sério: a violência contra a mulher. Diante desse desafio, um grupo de pesquisadoras e especialistas em Ciência de Dados decidiu unir forças para enfrentar essa questão de frente.



Esta Foto de Autor Desconhecido está licenciado em CC BY

Combatendo a Violência contra a Mulher

Tudo começou quando um grupo de discentes de Ciência de Dados, mergulharam fundo nos dados disponíveis sobre violência doméstica e de gênero. Elas estudaram relatórios oficiais, analisaram estatísticas e ouviram histórias de sobreviventes. Cada número, cada relato humano, ajudou a pintar um quadro mais claro da situação.



Combatendo a Violência contra a Mulher

Munidas com o conhecimento adquirido da análise inicial, passaram para a fase de explorarem soluções tecnológicas. Elas desenvolveram algoritmos, criaram modelos preditivos e experimentaram diferentes abordagens de machine learning. O objetivo era usar a tecnologia para identificar padrões, prever ocorrências e, assim, ajudar na prevenção da violência.



Combatendo a Violência contra a Mulher

Ao fim da jornada, consolidaram os resultados obtidos e partiram para a ação. Criaram um modelo preditivo, baseado em dados reais e análises estatísticas. Modelo que não só prevê aumento nas ocorrências, mas também fornece insights para autoridades, governos e organizações de defesa dos direitos das mulheres. O produto final é mais do que uma simples ferramenta. É uma promessa de esperança, para tantas mulheres que sofrem em silêncio. Com ele, as autoridades podem direcionar recursos de forma mais eficaz, os governos podem formular políticas públicas mais assertivas e as ONGs podem ampliar seu alcance.



Combatendo a Violência contra a Mulher

Almeja-se que esse projeto tenha impacto em todo o estado. Que a incidência de violência diminua, que as políticas de prevenção se tornem mais eficazes e que a conscientização da sociedade cresça. E que Minas Gerais caminhe na direção de futuro mais seguro e igualitário para todas as suas mulheres.



Combatendo a Violência contra a Mulher

E assim, uma jornada que começou com dados e tecnologia se transforma em uma história de esperança e mudança. Graças ao trabalho árduo e à dedicação daquelas que se recusaram a aceitar a violência como algo inevitável, Minas Gerais irá trilhar um novo caminho, rumo a um futuro em que todas as mulheres possam viver livres do medo e da opressão.



PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS ANO 2024 - 1º SEMESTRE - TURMA 03A

IV ETAPA DO PROJETO

Link de acesso ao GitHub

<https://github.com/AnaAleixo/PROJETO-APLICADO-II--Grupo-Cassandra-s>

STORYTELLING DO PROJETO





Bibliografia

A1 - Aplicando Conhecimento (atividade em grupo). Disponível em: <https://ava.mackenzie.br/courses/6638/assignments/45039?module_item_id=232347>. Acesso em: 15 maio. 2024.

A2 - Objetivo da etapa: PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A. Disponível em: <https://ava.mackenzie.br/courses/6638/pages/a2-objetivo-da-etapa?module_item_id=232348>. Acesso em: 15 maio. 2024.

A3 - Objetivo da etapa: PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A. Disponível em: <https://ava.mackenzie.br/courses/6638/pages/a3-objetivo-da-etapa?module_item_id=232353>. Acesso em: 15 maio. 2024.

A4 - Objetivo da etapa: PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A. Disponível em: <https://ava.mackenzie.br/courses/6638/pages/a4-objetivo-da-etapa?module_item_id=232358>. Acesso em: 15 maio. 2024.

Módulos do curso: PROJETO APLICADO II - CURSO TECNOLOGIA EM CIENCIA DE DADOS - ANO 2024 - 1º SEMESTRE - TURMA 03A. Disponível em: <<https://ava.mackenzie.br/courses/6638/modules>>. Acesso em: 15 maio. 2024.

Portal de Dados Abertos. Disponível em: <<https://dados.gov.br/dados/conjuntos-dados/sistema-nacional-de-estatisticas-de-seguranca-publica>>. Acesso em: 15 maio. 2024.

CG_FCOOnline_PredictiveAnalyticsFD_2Ed.pdf. , [s.d.]. Disponível em: <https://altabooks.com.br/wp-content/uploads/2019/07/CG_FCOOnline_PredictiveAnalyticsFD_2Ed.pdf>. Acesso em: 15 maio. 2024