

RELATÓRIO TÉCNICO - IFSP

Título: SINARM – Registros de armas de fogo com categoria 2020-2022.

Nome: Henrique Munhoz Leite

Prontuário: GU3020185

Disciplina: (TE2D6) Tópicos Especiais II - 2024S1

INTRODUÇÃO E JUSTIFICATIVA

Introdução:

Procura-se responder às seguintes perguntas para elaborar uma introdução:

- a) De que se trata o assunto?

R: Trata-se de uma base de dados sobre os registros de armas de fogo em território nacional dos anos de 2020 até 2022.

- b) Quais as principais motivações para que ele se desenvolva? Como nasceu a ideia de desenvolvê-lo?

R: Além do gosto pessoal, acho uma base interessante devido grandes discussões políticas sobre o assunto. Além de um grande histórico, a base conta com dados interessantes como quais calibres as armas de fogo possuem.

- c) Quais os elementos envolvidos neste estudo?

R:

A base de dados contém registros de emissão de armas de fogo no Brasil entre os anos de 2020 e 2022. Cada linha representa um registro de emissão de arma, incluindo detalhes sobre a arma, o proprietário, e o status do registro.

Ficha técnica da base de dados:

Especificações Técnicas:

Número de Registros (Linhas): 268.985

Número de Campos (Colunas): 11

Tamanho do Arquivo: 22.6 MB

Estrutura da Base de Dados:

ANO_EMISSAO

Descrição: Ano de emissão do registro.

Tipo de Dado: Inteiro (int64)

Exemplos: 2020, 2021, 2022

MES_EMISSAO

Descrição: Mês de emissão do registro.

Tipo de Dado: Inteiro (int64)

Exemplos: 1 (Janeiro), 2 (Fevereiro), ..., 12 (Dezembro)

UF

Descrição: Unidade Federativa onde o registro foi emitido.

Tipo de Dado: Texto (object)

Exemplos: AC (Acre), SP (São Paulo), RJ (Rio de Janeiro)

MUNICIPIO

Descrição: Município onde o registro foi emitido.

Tipo de Dado: Texto (object)

Exemplos: ASSIS BRASIL, BRASILÉIA, BUJARI

STATUS_REGISTRO

Descrição: Status do registro da arma.

Tipo de Dado: Texto (object)

Exemplos: Ativo/Válido, Inativo

ESPECIE_ARMA

Descrição: Tipo ou espécie da arma.

Tipo de Dado: Texto (object)

Exemplos: Pistola, Espingarda, Rifle

MARCA_ARMA

Descrição: Marca da arma.

Tipo de Dado: Texto (object)

Exemplos: TAURUS ARMAS S.A., CBC (COMPANHIA BRASILEIRA DE CARTUCHOS)

CALIBRE_ARMA

Descrição: Calibre da arma.

Tipo de Dado: Texto (object)

Exemplos: .380 ACP, 20, 36, .22 LR

SEXO

Descrição: Sexo do proprietário da arma.

Tipo de Dado: Texto (object)

Exemplos: M (Masculino), F (Feminino)

CATEGORIA

Descrição: Categoria do proprietário da arma.

Tipo de Dado: Texto (object)

Exemplos: Cidadão, Caçador de Subsistência

TOTAL

Descrição: Total de armas registradas para o proprietário.

Tipo de Dado: Inteiro (int64)

Exemplos: 1, 2, 3

Justificativa:

Trabalhar com a base de dados do SINARM é essencial para monitorar e ajustar a posse de armas de fogo, auxiliando na prevenção ao crime, no aumento da segurança pública e na formulação de políticas públicas eficazes. Pesquisadores podem estudar o perfil demográfico dos proprietários, entender tendências de mercado e avaliar o impacto social das armas. Além disso, empresas de segurança podem usar esses dados para desenvolver produtos e serviços direcionados. A análise espacial e temporal dos registros permite intervenções regionais e avaliação da ineficácia das políticas de desarmamento.

FUNDAMENTAÇÃO TEÓRICA

Processo KDD - Planejamento

Objetivo: Entender e analisar a base de dados de registros de armas de fogo de um breve período passado 2020 – 2022 com um governo com tendências a direita política. Identificar padrões em seus dados, como ocorrências em compras por região, e criar, com base no modelo gerado, um motor capaz de prever quais calibres e armas serão mais registrados em determinadas regiões do território nacional.

1 – Obtenção dos dados:

Fonte de dados: REGISTRO COM CATEGORIA 2020-2022.

Base de Dados:

Fonte: Arquivo CSV

Ferramentas: Python, Jupyter lab/Notebook, Word, Excel, IA

Período: 2020 – 2022

Colunas:

ANO_EMISSAO: Ano de emissão do registro.

MES_EMISSAO: Mês de emissão do registro.

UF: Unidade Federativa (Estado).

MUNICIPIO: Município.

STATUS_REGISTRO: Status do registro.

ESPECIE_ARMA: Espécie da arma.

MARCA_ARMA: Marca da arma.

CALIBRE_ARMA: Calibre da arma.

SEXO: Sexo do proprietário.

CATEGORIA: Categoria do registro.

TOTAL: Total de registros.

1. Extração dos Dados

A extração dos dados foi realizada a partir de um arquivo CSV. Utilizei a biblioteca Pandas para carregar os dados no formato tabular.

2. Transformação dos Dados

Durante a transformação, os seguintes passos foram realizados:

Renomear Colunas:

Correção do nome da coluna MES_MISSAO para MES_EMISSAO.

Seleção de Colunas:

Garantir que todas as colunas relevantes estão presentes e organizadas.

Remoção de Valores Ausentes:

Verificação e remoção de registros com valores ausentes.

Conversão de Tipos de Dados:

Conversão das colunas ANO_EMISSAO, MES_EMISSAO e TOTAL para o tipo int.

Remoção de Valores Zero:

Remoção de registros onde ANO_EMISSAO, MES_EMISSAO ou TOTAL eram iguais a zero.

Remoção de Outliers:

Utilização do método do intervalo interquartil (IQR) para remover outliers das colunas numéricas.

Codificação de Variáveis Categóricas:

Transformação de variáveis categóricas em variáveis binárias utilizando one-hot encoding.

Normalização de Variáveis Numéricas:

Normalização das colunas ANO_EMISSAO, MES_EMISSAO e TOTAL utilizando StandardScaler.

3. Análise e Verificação dos Dados

A verificação dos dados processados foi realizada para garantir que os dados estão corretos e prontos para análise.

O código completo de pré-processamento foi anexado ao moodle no formato de um notebook Jupyter (.ipynb).

DESENVOLVIMENTO (METODOLOGIA E ANÁLISE)

1. Definição do Problema

Antes de iniciar o trabalho, é crucial definir claramente os objetivos específicos:

- **Analisar a distribuição de registros de armas por região (estado e município).**
- **Identificar padrões temporais (mensais e anuais) nos registros.**
- **Determinar correlações entre categorias de armas, calibres, marcas e regiões.**
- **Criar um modelo preditivo para prever registros futuros de calibres e armas por região.**

2. Coleta e Carregamento dos Dados

- **Fonte dos Dados:** Utilizamos uma base de dados fornecida, contendo registros de armas de fogo entre 2020 e 2022.

- **Carregamento dos Dados:** Utilizamos a biblioteca Pandas para carregar os dados no ambiente de análise.

3. Limpeza e Pré-processamento dos Dados

- **Remover Valores Ausentes:** Eliminamos linhas com valores ausentes usando `df.dropna()`.
- **Conversão de Tipos de Dados:** Convertendo as colunas de data para o tipo `datetime`.
- **Remoção de Outliers:** Identificação e remoção de outliers para evitar distorções na análise.
- **Codificação de Variáveis Categóricas:** Transformação de variáveis categóricas em variáveis dummy usando one-hot encoding.
- **Normalização de Variáveis Numéricas:** Normalização dos dados numéricos para garantir que todas as variáveis estejam na mesma escala.

4. Análise Exploratória de Dados (EDA)

- **Distribuição Temporal:** Analisamos a distribuição de registros por ano e mês.
- **Distribuição Regional:** Analisamos a distribuição de registros por UF e município.
- **Análise por Categoria:** Examinamos a distribuição de registros por espécie de arma, marca, calibre e categoria.
- **Correlação de Variáveis:** Utilizamos matrizes de correlação para identificar relações entre variáveis.

5. Identificação de Padrões

- **Análise de Tendências Temporais:** Utilizamos gráficos de linha e histogramas para visualizar tendências ao longo do tempo.
- **Mapas de Calor Regional:** Criamos mapas de calor para visualizar a distribuição geográfica dos registros de armas.
- **Análise de Segmentos:** Agrupamos os dados por diferentes categorias para identificar padrões específicos.

6. Criação do Modelo Preditivo

- **Divisão dos Dados:** Dividimos os dados em conjuntos de treino e teste.
- **Seleção do Modelo:** Utilizamos algoritmos de aprendizado de máquina, como `RandomForestClassifier`.
- **Treinamento do Modelo:** Treinamos o modelo utilizando o conjunto de treino.
- **Avaliação do Modelo:** Avaliamos a performance do modelo utilizando métricas como precisão, recall, F1-score e AUC-ROC no conjunto de teste.
- **Ajuste e Otimização:** Ajustamos hiperparâmetros e otimizamos o modelo para melhorar sua performance.

7. Implementação do Motor de Previsão

- **Desenvolvimento do Motor:** Desenvolvemos um motor de previsão utilizando o modelo treinado.
- **Validação e Teste:** Validamos o motor de previsão com dados novos para garantir sua precisão.
- **Implantação:** Implementamos o motor em um ambiente de produção, onde pode ser utilizado para fazer previsões sobre registros futuros de armas.

8. Documentação e Relatórios

- **Documentação Técnica:** Documentamos todo o processo, incluindo código, metodologias e resultados.

CONSIDERAÇÕES FINAIS

Honestamente, não foi fácil e em muitos momentos recorri a ajuda na internet para o desenvolvimento. Mais difícil do que esperado, e desafiador no sentido de entender o que extrair do motor, mais até do que necessariamente a sua construção.

REFERÊNCIAS BIBLIOGRÁFICAS

https://servicos.dpf.gov.br/dadosabertos/SINARM_CSV/REGISTROS/