**Documentação da Ferramenta MoBAt: Uma Ferramenta de Monitoramento e Análise de Dados de Bases de Ameaças.**

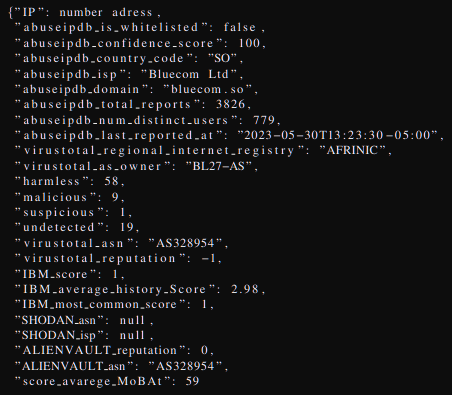
**Visão Geral**

MoBAt é uma ferramenta abrangente para coleta, processamento, análise e visualização de dados de reputação de endereços IP. A ferramenta é composta por dois módulos principais:

* **Módulo de Coleta de Metadados:**
  + Coleta dados de diversas fontes, incluindo AbuseIPDB, SHODAN, Virus Total, Alien Vault, IBM X-Force e relatórios de cibersegurança.
  + Processa e junta as características de cada base de dados.
  + Detecta intrusão.
* **Módulo de Processamento e Análise:**
  + Processa arquivos JSON em um diretório específico e gera um arquivo CSV unificado com os dados processados.
  + Permite a visualização e exploração interativa de dados de endereços IP coletados ao longo de diferentes períodos de tempo.
  + Exibe diversos tipos de visualizações, como gráficos de comportamento de IP, mapeamento de recursos, clusters, seleção de recursos, importância de recursos para aprendizado de máquina, score average Mobat para principais IPs, reputação por país, heatmap de ocorrência por país e tabela de acurácia do modelo e tempo de treinamento.

**Funcionalidades**

* **Coleta de dados:**
  + Coleta dados de diversas fontes de inteligência de ameaças.
  + Processa e junta as características de cada base de dados.
  + Detecta intrusão.
* **Processamento de dados:**
  + Processa arquivos JSON em um diretório específico.
  + Gera um arquivo CSV unificado com os dados processados.
  + Realiza diversas operações de limpeza e tratamento nas colunas do DataFrame.
  + Exporta o Data Frame final para um arquivo CSV.
* **Análise de dados:**
  + Permite a visualização e exploração interativa de dados de endereços IP.
  + Exibe diversos tipos de visualizações.
  + Calcula o valor médio para colunas-chave.
  + Permite a seleção do semestre e do tipo de visualização.
* **Outras funcionalidades:**
  + Permite a otimização do código.
  + Fornece uma interface simples para o usuário no terminal.



**Figura 1: Exemplo de vetor de características.**

**Estrutura do Código**

O código da ferramenta MoBAt está dividido em dois scripts principais:

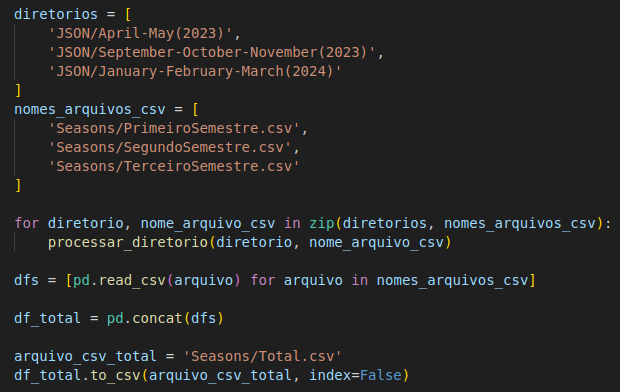
* **Tratament.py:**
  + Implementa as funcionalidades de processamento de dados.
* **Monitoring.py:**
  + Implementa as funcionalidades de análise de dados.

Dependências

* **Pandas:** biblioteca para lidar com dataframes.
* **Matplotlib:** biblioteca para plotar gráficos.

**Uso**

* **Tratament.py:**
  + Execute o script com o comando python Tratament.py.
  + Especifique o diretório que contém os arquivos JSON.
  + Especifique o nome do arquivo CSV de saída.
  + O código **Tratament.py** realiza o processamento de arquivos JSON em um diretório específico e gera um arquivo CSV unificado com os dados processados. Aqui está uma descrição das principais funcionalidades e como elas são implementadas:
    - **Funcionalidade**:
      * **Processamento de Arquivos JSON**: O código percorre um diretório fornecido, extrai arquivos JSON e os carrega em um dicionário de dados unificados.
      * **Criação do Data Frame**: Um DataFrame do Pandas é criado a partir do dicionário de dados unificados.
      * **Tratamento de Colunas**: O código realiza diversas operações de limpeza e tratamento nas colunas do Data Frame, como:
        + Extração de IP e horário a partir dos índices dos dados.
        + Separação de colunas relacionadas à reputação de IPs (como 'harmless', 'malicious', 'suspicious', 'undetected') e valores numéricos relacionados.
        + Conversão de valores 'None' para nulo e preenchimento de valores nulos com 0 em colunas específicas.
        + Inversão de valores em colunas específicas ('abuseipdb\_confidence\_score', 'undetected', 'malicious', 'suspicious') para calcular a média da pontuação.
        + Conversão de valores booleanos para representações corretas ('True' ou 'False') em colunas específicas.
      * **Seleção de Colunas**: Seleção das colunas relevantes para o Data Frame final.
      * **Exportação para CSV**: O Data Frame final é exportado para um arquivo CSV especificado.
      * **Processamento de Múltiplos Diretórios**: O código processa dados de vários diretórios e gera arquivos CSV separados para cada diretório, além de um arquivo CSV unificado com todos os dados.
  + **Exemplo de Uso**
  + O seguinte código exemplifica como usar o script **Tratament.py** para processar dados de um diretório específico:



* + O código acima processa os dados do diretório “**JSON/April-May(2023)**” e gera um arquivo CSV chamado “**Seasons/PrimeiroSemestre.csv**” com os dados processados.
* **Monitoring.py:**
  + Execute o script com o comando python Monitoring.py.
  + Selecione o semestre e o tipo de visualização que deseja ver no menu interativo.
  + Siga as instruções na tela para interagir com as visualizações.
  + **Monitoring.py** é um script projetado para visualização e exploração interativa de dados de endereços IP coletados ao longo de diferentes períodos de tempo. O script permite que os usuários carreguem arquivos CSV contendo dados de endereços IP para diferentes semestres e visualizem vários aspectos dos dados, incluindo padrões de comportamento, mapeamento de recursos, clusters, seleção de recursos e muito mais.

**Funcionalidades**

* + - Carregamento de Dados: O script carrega arquivos CSV para cada semestre, representando dados de endereços IP para esse período. Os arquivos são carregados em DataFrames separados (df1, df2, df3, df4) para cada semestre.
    - Cálculo do Valor Médio: Os valores médios para colunas-chave (abuseipdb\_confidence\_score, abuseipdb\_total\_reports, etc.) são calculados para cada semestre e armazenados em dicionários (mean\_values\_df1, mean\_values\_df2, mean\_values\_df3, mean\_values\_df4).
    - Menu Interativo: O script apresenta um menu interativo ao usuário, permitindo que escolham o semestre e o tipo de visualização que desejam ver.
    - Opções de Visualização:
      * Gráficos de Comportamento de IP: Permite que os usuários visualizem o comportamento de endereços IP específicos.
      * Mapeamento de Recursos: Visualiza o mapeamento de recursos no conjunto de dados.
      * Clusters: Exibe clusters nos dados.
      * Visualização da Seleção de Recursos: Visualiza técnicas de seleção de recursos.
      * Importância de Recursos para Aprendizado de Máquina: Exibe a importância dos recursos para aprendizado de máquina.
      * Score Average Mobat para Principais IPs: Visualiza o Score Average Mobat para os principais IPs.
      * Reputação por País: Exibe a reputação dos IPs por país.
      * Download de Dados de IP: Permite que os usuários baixem dados de IP para o período selecionado.
      * Heatmap de Ocorrência por País: Exibe um heatmap de ocorrências de IP por país.
      * Tabela de Acurácia do Modelo e Tempo de Treinamento: Exibe uma tabela de acurácia do modelo e tempo de treinamento.

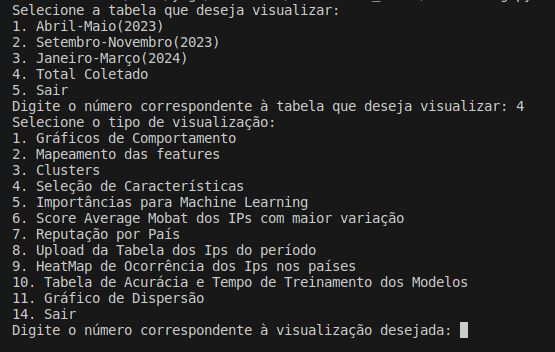
**Estrutura do Código**

* + - O código está estruturado em três partes, cada uma focando em diferentes aspectos de visualização e exploração de dados.
    - Parte 1: Carregamento de dados e cálculo do valor médio para cada semestre.
    - Parte 2: Menu interativo para seleção de opções de visualização.
    - Parte 3: Implementação de funções de visualização para diferentes aspectos dos dados.

**Dependências**

* + - O script requer a biblioteca pandas para lidar com dataframes.
    - As funções de visualização utilizam o matplotlib para plotar gráficos.

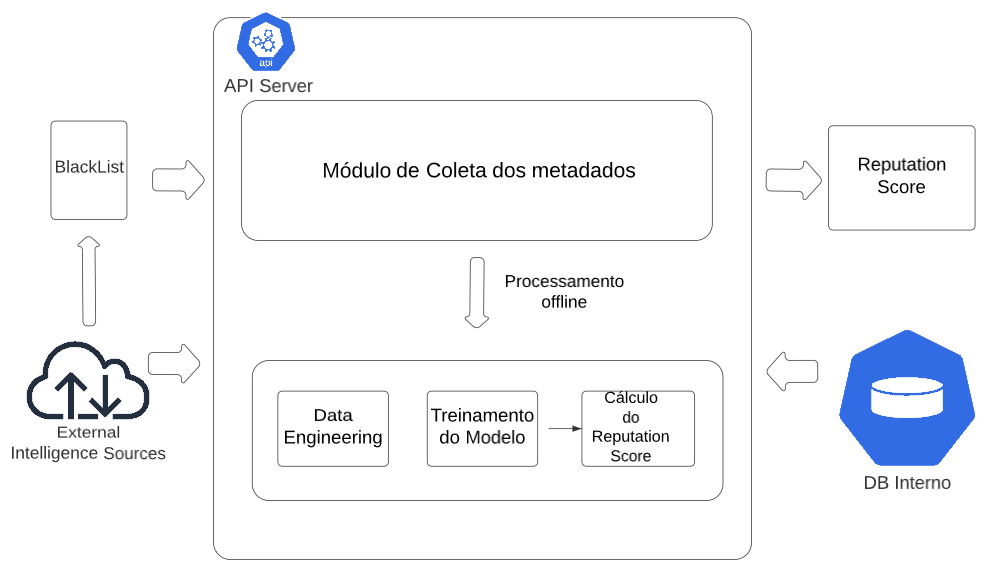
**Exemplo de Uso**

****

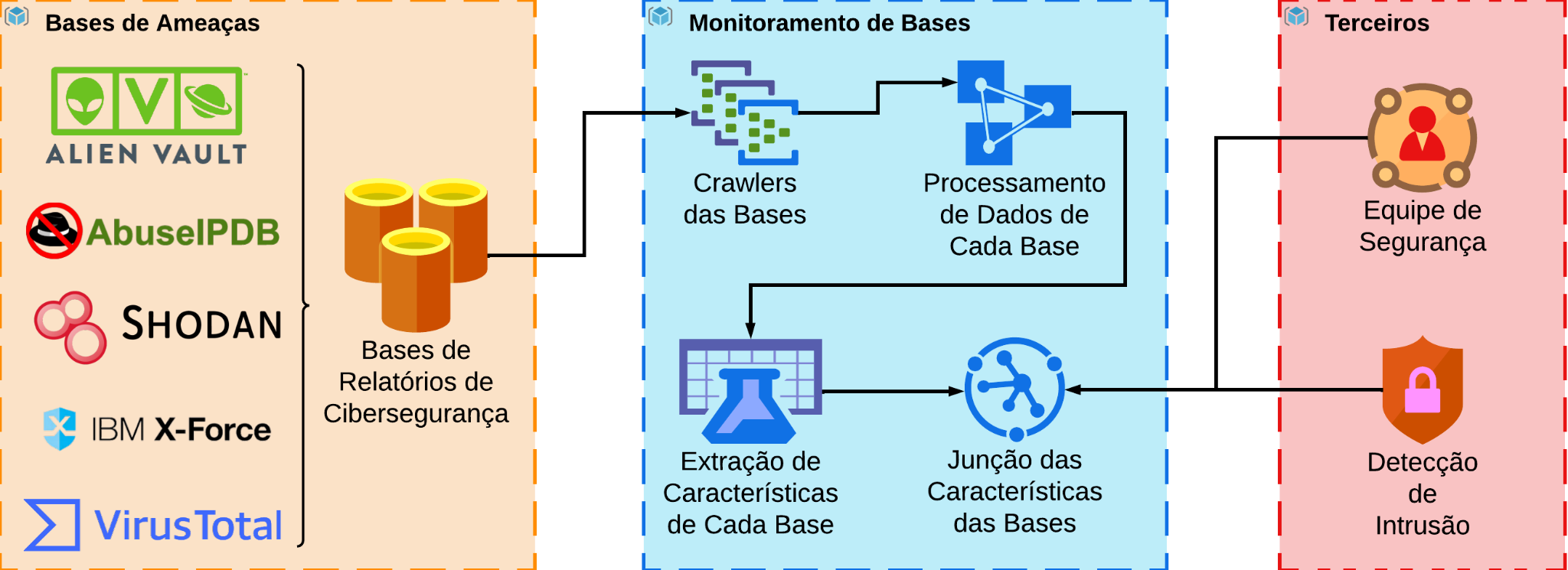
* + - Selecione o semestre e o tipo de visualização que deseja ver no menu interativo.
    - Siga as instruções na tela para interagir com as visualizações.

**Observações**

* O código pode ser otimizado ainda mais, reduzindo a duplicação de código, melhorando a validação de entrada do usuário e consolidando funções de plotagem.
* A ferramenta MoBAt é uma ferramenta poderosa para analisar dados de reputação de endereços IP.
* A ferramenta pode ser utilizada por profissionais de segurança da informação para identificar e mitigar ameaças cibernéticas.



**Figura 1: Visão geral da ferramenta MoBAt.**



**Figura 2: Arquitetura da ferramenta MoBAt.**