

# Checkpoint1 - Cyber: Verificador Imigração - PFA

## Sumario

- Problema
  - Recursos
  - Método
- Código
  - Estruturando passo a passo do código:
    - Instalar as bibliotecas necessárias:
    - Importar as bibliotecas necessárias
    - Importar as APIs
    - Limpar os dados recebido nas APIs
    - Inutar os dados do agente-PFA
    - Limpar os dados inputados
    - Validar os dados inputados
    - Comparar os dados inputados com as listas das APIs
    - Devolver o resultado para o usuário
    - Juntar tudo em uma função e adicionar recursos para poder montar uma Web page
- Testes
  - Teste Passaporte
    - Resultado Passaporte
  - Teste Nome
    - Métricas
    - Threshold 0.85:
      - Resultado
    - Threshold 0.80:
      - Resultado

---

## Problema

A proposta do nosso primeiro Checkpoint da matéria de *CyberSecurity* seria desenvolver uma solução com recursos aprendidos dentro da matéria para a INTERPOL.

A ideia é que essa solução, de Inteligência Artificial, possa ajudar diretamente no combate ao terrorismo, basicamente, identificando dentro de um grupo de pessoas, dentro desse contexto, pessoas que estão chegando no seu país, independente se são estrangeiros ou residentes do mesmo.

---

## Recursos

Para podermos realizar esse projeto, recebemos algumas APIs contendo Passaportes e nomes de pessoas consideradas "Proibidas" ou "Terroristas em potencial".

---

## Método

Uma vez com esses dados, devemos criar um ambiente em que um agente da PFA, possa inserir os dados do passageiro que esta entrando no país, e devolver para ele a resposta pronta, se ele esta liberado, se ele deve ser preso ou apenas investigado para detectar alguma discrepância.

---

## Código

Estruturando passo a passo do código:

- Instalar as bibliotecas necessárias
- Importar as bibliotecas necessárias
- Importar as APIs
- Limpar os dados recebidos nas APIs
- Inutar os dados do agente-PFA
- Limpar os dados inputados
- Validar os dados inputados
- Comparar os dados inputados com as listas das APIs
- Devolver o resultado para o usuário
- Juntar tudo em uma função e adicionar recursos para poder montar uma Web page


### Instalar as bibliotecas necessárias:

```
!pip install -q pyngrok
!pip install -q streamlit
!pip install -q streamlit_ace
!pip install fuzzywuzzy[speedup]
!streamlit run app.py & > / dev / null &
```

### Importar as bibliotecas necessárias

```
import streamlit as st
import requests as req
import re
from fuzzywuzzy import process
from fuzzywuzzy import fuzz
from pyngrok import ngrok
```

### Importar as APIs

 Definindo 2 funções para pegar as duas APIs

```
def api_passaport():
    lista_passaporte = []
    lista_arg = req.get("https://kndhvh.github.io/arg.json")

def api_nome():
    lista_nome = []
    lista_interpol = req.get("https://cspinheiro.github.io/interpol.json")
```


### Limpar os dados recebido nas APIs

 Atualizando as 2 funções para que elas devolvam os dados limpos em forma de uma lista

```
def api_passaporte():
    lista_passaporte = []
    lista_arg = req.get("https://kndhvh.github.io/arg.json")
    arg = lista_arg.json()['arg']
    for i in arg:
        lista_passaporte.append(str(i['arg']))
    return lista_passaporte


def api_nome():
    lista_nome = []
    lista_interpol = req.get("https://cspinheiro.github.io/interpol.json")
    interpol = lista_interpol.json()['interpol']
    for i in interpol:
        lista_nome.append(i['interpol'])
    return lista_nome
```

#### Inputar os dados do agente-PFA

 Inputando usando um comando do *streamlit*

```
input_passaporte = st.text_input('Passaporte')
input_nome = st.text_input('Nome Completo')
```


#### Limpar os dados inputados

 Utilizando REGEX para tirar do passaporte, tudo que não for numero

```
input_passaporte = re.sub('\D', '', input_passaporte)
```

#### Validar os dados inputados

 Criando 2 Funções para poder validar os dados inseridos:

 Para um nome ser valido ele deve:

- Ter no mínimo 5 letras
- Ter no mínimo 2 nomes

 Para um passaporte ser valido ele deve:

- Ter exatamente 9 dígitos sem contar as letras

```
def valid_nome(input_nome):
    if len(input_nome) > 4 and len(re.split('\s',input_nome)) >= 2:
        st.success('[VERIFICADO]')
    else:
        st.warning('Nome Invalido!')

def valid_passa(input_passaporte):
    if len(input_passaporte) == 9:
        st.success('[VERIFICADO]')
    else:
        st.warning('Passaporte Invalido!')
```

### Comparar os dados inputados com as listas das APIs

**i** Definindo 2 funções para verificar os dados inseridos com a lista, com o nome, utilizando fuzzywuzzy, e o passaporte REGEX

```
def getmatch_nome(input,lista):
    search = process.extractOne(input,lista,scorer=fuzz.token_sort_ratio)
    if search[1] > 80:
        return 1
    return 0

def getmatch_passa(input,lista):
    match = re.search(lista,input)

    if match:
        return 1
    else:
        return 0
```

### Devolver o resultado para o usuário

**i** Juntando as 4 ultimas funções para que possam conversar entre si e devolver pro usuário o resultado

```

def getmatch_nome(input,lista):
    search = process.extractOne(input,lista,scorer=fuzz.
token_sort_ratio)
    if search[1] > 80:
        return 1
    return 0

def valid_nome(input_nome,lista_nome):
    if len(input_nome) > 4 and len(re.split('\s',input_nome)) >= 2:
        nome = getmatch_nome(input_nome,lista_nome)
        if nome == 1:
            st.error('[CUIDADO] Nome suspeito, favor investigar!')
        else:
            st.sucess('[VERIFICADO] Passageiro liberado para entrada no
pais!')
    else:
        st.warning('Nome Invalido!')

def getmatch_passa(input,lista):
    match = re.search(lista,input)

    if match:
        return 1
    else:
        return 0

def valid_passa(input_passaporte,lista_passaporte):
    if len(input_passaporte) == 9:
        passa = getmatch_passa(input_passaporte,lista_passaporte)
        if passa == 1:
            st.error('[CUIDADO] Passaporte proibido, deter
imediatamente!')
            return 1
        else:
            st.warning('Passaporte Invalido!')

```

Juntar tudo em uma função e adicionar recursos para poder montar uma Web page

 O resultado pode ser visto [Aqui](#)

## Testes

### Teste Passaporte

O primeiro que vamos testar é a verificação de Passaporte, porem, como utilizamos apenas REGEX, não existe um *threshold*, e o programa só ira acusar se o passaporte bater 100% com o da lista de proibidos.

Dito isso, Geramos 20 passaportes aleatórios, e pegamos 20 passaportes da lista de proibidos e adicionamos países aleatórios para o código lidar, esse foi o resultado:

Passaporte	É Proibido?	Detectou Proibido?
FRA784550431	Não	Não
BRA360380270	Não	Não
PER662478032	Não	Não
CHI239808468	Não	Não
GER296300360	Não	Não
FRA752100814	Não	Não
BRA586579990	Não	Não
PER166483333	Não	Não
CHI170831440	Não	Não
GER140248039	Não	Não
FRA626235250	Não	Não
BRA677646403	Não	Não
PER934816755	Não	Não
CHI157544764	Não	Não
GER665071625	Não	Não
FRA810285156	Não	Não
BRA943619895	Não	Não
PER678232656	Não	Não
CHI832611948	Não	Não
GER643561395	Não	Não
FRA220471986	Sim	Sim
BRA585957612	Sim	Sim
PER357690491	Sim	Sim
CHI884128753	Sim	Sim
GER492953014	Sim	Sim
FRA195117564	Sim	Sim
BRA378797522	Sim	Sim
PER435052858	Sim	Sim
CHI802706043	Sim	Sim
GER948127235	Sim	Sim
FRA894983311	Sim	Sim
BRA986771815	Sim	Sim
PER190726026	Sim	Sim
CHI287905181	Sim	Sim
GER572761935	Sim	Sim

FRA810848397	Sim	Sim
BRA187696913	Sim	Sim
PER709537367	Sim	Sim
CHI508541271	Sim	Sim
GER969725523	Sim	Sim

### Resultado Passaporte

Como o resultado do programa foi perfeito, não faz muita diferença qual métrica iremos utilizar para avaliar, de qualquer forma, essas foram as avaliações:

Precisão	Recall	F1-Score	Acurácia
1,00	1,00	1,00	1,00

### Teste Nome

Geramos 40 nomes aleatórios com base nos nomes e sobrenomes da base de dados, com a adição de alguns nomes e sobrenomes de fora, pegamos 40 nomes da lista e fizemos pequenas modificações em alguns, esses foram os resultados:

### Métricas

Escolhemos como métricas o F1-Score e o Recall pelos seguintes motivos:

- Recall: Dentro do ambiente de um Aeroporto na Argentina, não seriam tomadas nenhuma medidas consideradas drásticas ou severas, levando em consideração que um falso positivo traria apenas um desconforto momentâneo ao passageiro ou uma possível perda de tempo, é válido desconsiderar possíveis falsos positivos.
- F1-Scores: o F1-Score é uma das métricas mais completas, portanto, é imprescindível para ter uma visão geral do desempenho, uma vez que o Recall foca em apenas um aspecto.

### Threshold 0.85:

Nome	Proibido	Detectou Proibido?
Maurício de Pereira	Não	Não
Fernando Ribeiro	Não	Não
Matias Silva	Não	Não
Pietra Anjos Castro	Não	Não
Silvana de Oliveira Braga	Não	Não
Bruna Leão	Não	Não
Marianna de Boas Azevedo	Não	Não
Marianna Pereira	Não	Não
Fernando Ribeiro Castro	Não	Não
Patrícia de Rebouças couto	Não	Não
Armando Pinhei	Não	Não
Fabiana de Castro	Não	Não
Maurício Ribeiro	Não	Não
Bruna Boas	Não	Não
Fernanda de Pereira	Não	Não
Fernanda Ribeiro	Não	Não

Leonardo Silva	Não	Não
Fabiana Pinheiro	Não	Não
Pedro Oliveira Azevedo	Não	Não
Nathalia Ribeiro	Não	Não
Silvana Boas	Não	Não
Nathan Pereira	Não	Não
Fernanda de Ribeiro	Não	Não
Felipe de Pinheiro	Não	Não
Pedro Ribeiro	Não	Não
Bruno Pereira Cabral	Não	Não
Ana Ribeiro	Não	Não
Silvana de Pinheiro	Não	Não
Alice de Pereira Castro	Não	Não
Luzivan Ribeiro Herklotz	Não	Não
Maria Leão	Não	Não
Maurício Pereira Maciel	Não	Não
Marianna Ribeiro	Não	Não
Maurício Leão	Não	Não
Luzivan de Pereira Braga	Não	Não
Maurício Ribeiro	Não	Não
Luana Leão Castro	Não	Não
Pedro Pereira	Não	Não
Bruna de Ribeiro	Não	Não
Alice Leão	Não	Não
Matias Herklotz	Não	Não
Priscila dos Santos	Sim	Não
Bruno de Oliveira Ribeiro	Sim	Sim
Pedro da Silva	Sim	Sim
Luzia dos Anjos	Sim	Sim
Luzivan de Oliveira	Sim	Sim
Ricardo da Silva Leão	Sim	Sim
Bruna de Villas Boas	Sim	Sim
Fernando Santos Pereira	Sim	Sim
Fernando de Oliveira Ribeiro	Sim	Sim
Pietra da Silva Rebouças	Sim	Sim
Fernanda dos Anjos Pinheiro	Sim	Sim
Fernando de Oliveira	Sim	Sim



Patrícia da Silva Ribeiro	Sim	Sim
Fernando de Villas Boas	Sim	Sim
Maurício dos Santos Pereira	Sim	Sim
Marlene de Oliveira Ribeiro	Sim	Sim
Nathan da Silva	Sim	Sim
Maurício dos Anjos Pinheiro	Sim	Sim
Michael Oliveira	Sim	Sim
Maurício Silva Ribeiro	Sim	Sim
Maurício de Villas Boas	Sim	Sim
Isabelle dos Santos Pereira	Sim	Sim
João de Oliveira Ribeiro	Sim	Sim
Isabelle dos Anjos Pinheiro	Sim	Sim
Ana da Silva Ribeiro	Sim	Sim
Marianna dos Santos Pereira	Sim	Sim
Marianna de Ribeiro	Sim	Não
Marianna dos Anjos Pinhei	Sim	Sim
Marianna da Silva Ribeiro	Sim	Sim
Pedro dos Santos Pereira	Sim	Sim
Leonardo de Oliveira Ribeiro	Sim	Sim
Silvana da Silva Leão	Sim	Sim
Maria Clara Santos Pereira	Sim	Sim
Fabia Oliveira Ribeiro	Sim	Sim
Maria Clara Silva	Sim	Sim
Danielly Santos Pereira	Sim	Sim
Bárbara Oliveira Ribeiro	Sim	Sim
Luana Silva Leão	Sim	Sim
Larisa Santos Pereira	Sim	Sim
Larissa Oliveira Ribeiro	Sim	Sim
Lucas Silva Leão	Sim	Sim

## Resultado

Recall	F1-Score
0,95	0,98

Apesar das 2 métricas estarem com valores relativamente altos, temos que considerar 2 pontos:

- Foram feitas poucas alterações nos dados de teste da lista da INTERPOLL;
- Apesar de não necessariamente ser um problema, tivemos um total de 0 falsos positivos;

Visto isso, não podemos descartar a possibilidade de termos um problema de *Overfitting*, portanto, iremos refazer os testes com um *Threshold* um pouco menor, e fazendo mais alterações manuais nos dados

**Threshold 0.80:**

Nome	Proibido	Detectou Proibido?
Fabiana Pereira	Não	Não
Luana de Ribeiro	Não	Não
Fabiana Silva	Não	Não
Armando Anjos	Não	Não
Michael Oliveira	Não	Sim
Lucas Leão	Não	Não
Bruna de Boas	Não	Não
Fernando Pereira	Não	Não
Danielly Ribeiro Cabral	Não	Não
Maria Rebouças	Não	Não
Ana de Pinheiro Azevedo	Não	Não
Pedro de Oliveira	Não	Sim
Maurício Ribeiro	Não	Não
Ana de Boas couto	Não	Não
Maurício Pereira	Não	Não
Luana Ribeiro	Não	Não
Ricardo Silva	Não	Não
Fernando de Pinheiro	Não	Não
Luzivan Oliveira Cabral	Não	Sim
Fernanda Ribeiro	Não	Não
Nathalia de Boas Braga	Não	Não
Fernando de Pereira Azevedo	Não	Não
Silvana de Ribeiro	Não	Não
Bruna Pinheiro couto	Não	Não
Nathalia de Ribeiro	Não	Não
Bruna de Pereira Braga	Não	Não
Felipe de Ribeiro	Não	Não
Maurício de Pinheiro	Não	Não
Bruna de Pereira	Não	Não
Larissa de Ribeiro	Não	Não
Maurício Leão Cabral	Não	Não
Fernanda de Pereira	Não	Não
Larissa Ribeiro Cabral	Não	Não
Luzivan Leão	Não	Não
Lucas Pereira	Não	Não

Michael Ribeiro Herklotz	Não	Não
Maria de Leão	Não	Não
Ana Pereira Castro	Não	Não
Isabelle Ribeiro	Não	Não
Larissa de Leão	Não	Não
Larissa de Castro	Não	Não
Priscila Santos Pereira	Sim	Sim
Bruno de Oliveira	Sim	Sim
Pedro da Silva Jr	Sim	Sim
Luzia dos Anjos	Sim	Sim
Luzivan de Oliveira	Sim	Sim
Ricardo da Silva Leão	Sim	Sim
Bruna de Villas Boas	Sim	Sim
Fernando dos S Pereira	Sim	Sim
Fernando de Oliveira Ribeiro	Sim	Sim
Pietra da Silva Rebouças	Sim	Sim
Fernanda Anjos Pinheiro	Sim	Sim
Fernando de Oliveira	Sim	Sim
Patrícia da Silva Ribeiro	Sim	Sim
Fernando de Villas Boas	Sim	Sim
Maurício dos Pereira	Sim	Sim
Marlene de Oliveira Ribeiro	Sim	Sim
Nathan da Silva	Sim	Sim
Maurício dos Anjos Pinheiro	Sim	Sim
Michael de Oliveira	Sim	Sim
Maurício da Silva Ribeiro	Sim	Sim
Maurício de Villas	Sim	Sim
Isa dos Santos Pereira	Sim	Sim
João Oliver Ribeira	Sim	Sim
Isabelle dos Anjos Pinheiro	Sim	Sim
Ana da Silva Ribeiro	Sim	Sim
Mari dos Santos Pereira	Sim	Sim
Marianna de Oliveira Ribeiro	Sim	Sim
Marianna dos Almeida Pinheiro	Sim	Sim
Marianna Silva Ribeiro	Sim	Sim
Pedro Santos Pereira	Sim	Sim
Leonardo de Oliveira Ribeiro Antonio	Sim	Sim

Silvana da Silva Leão	Sim	Sim
Maria Clara Santos Pereira	Sim	Sim
Fabiana Oliveira Ribeiro	Sim	Sim
Maria Clara Silva Leão	Sim	Sim
Danielly Santos Pereira	Sim	Sim
Bárbara Oliveira Ribeiro	Sim	Sim
Luana Silva Leão Castro	Sim	Sim
Larissa Santos Pereira	Sim	Sim
Larissa Oliveira Ribeiro	Sim	Sim
Lucas Silva Leão Andrade	Sim	Não

Resultado

Recall	F1-Score
0,98	0,95