

# TRAFFIC DATA ANALYSIS (RECIFE)

INTERLIGENCIA ARTIFICIAL  
TRAFEGO AUTOMOTIVO



# INTEGRANTES



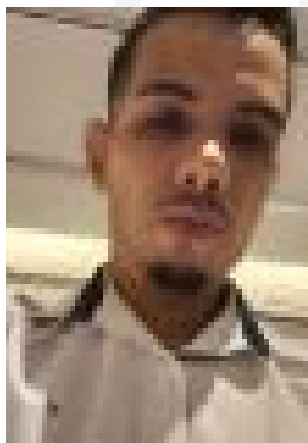
- Nome: Raphael L. Siqueira
- Contato: leitesiqueiraraphael@gmail.com
- Ra: 1352222014
- Formação: Ciência da Computação - UNIFG



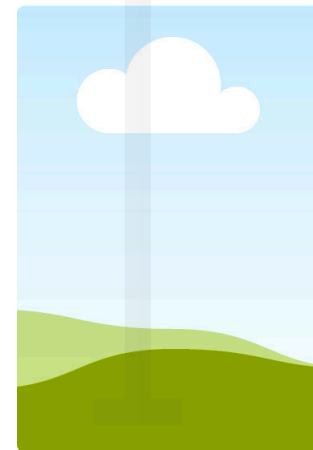
- Nome: João Gabriel de Oliveira Feitosa
- Contato: 1352111818@ulife.com.br
- Ra: 1352111818
- Formação: Ciência da Computação - UNIFG



- Nome: Victor Hugo Silva
- Contato: victor.hugolk0@gmail.com
- Ra: 135211404
- Formação: Ciência da Computação UNIFG

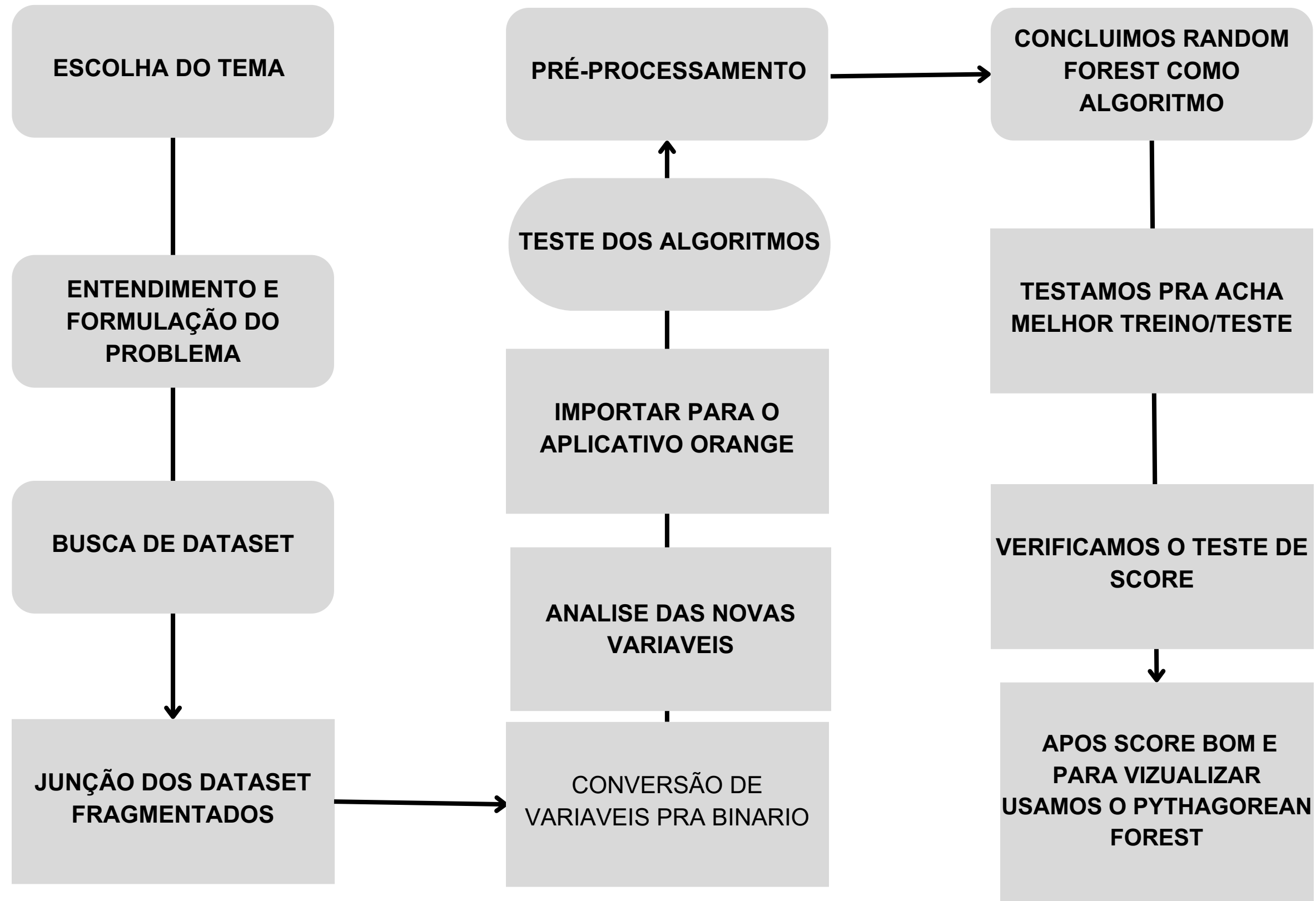


- Nome: José Augusto da Silva Neto
- Contato: 1352110816@ulife.com.br
- Ra: 1352110816
- Formação: Ciência da Computação - UNIFG



- Nome: William Ruan
- Contato: williamruam2424@gmail.com
- Ra: 1352320666
- Formação: Ciência da Computação - UNIFG

# WORKFLOW



## ENTENDIMENTO DO PROBLEMA

Recife enfrenta sérios desafios na mobilidade urbana, com mais de 1,5 milhão de veículos em circulação diariamente e uma infraestrutura inadequada. A topografia acidentada e a prioridade dada ao transporte individual resultam em congestionamentos, dificuldades de estacionamento e ineficiência no sistema de transporte, não atendendo às necessidades de uma população crescente.

Os principais problemas incluem: Altos índices de acidentes de trânsito, refletindo uma falta de segurança viária.

Analisando dados coletados no site da Prefeitura do Recife, é possível oferecer recomendações concretas para a prefeitura e órgãos de trânsito, contribuindo para um planejamento mais eficaz e sustentável da mobilidade no Recife e arredores. Será que isso pode levar a uma redução nos engarrafamentos, melhoria na segurança viária e, em última instância, uma melhor qualidade de vida para os cidadãos?



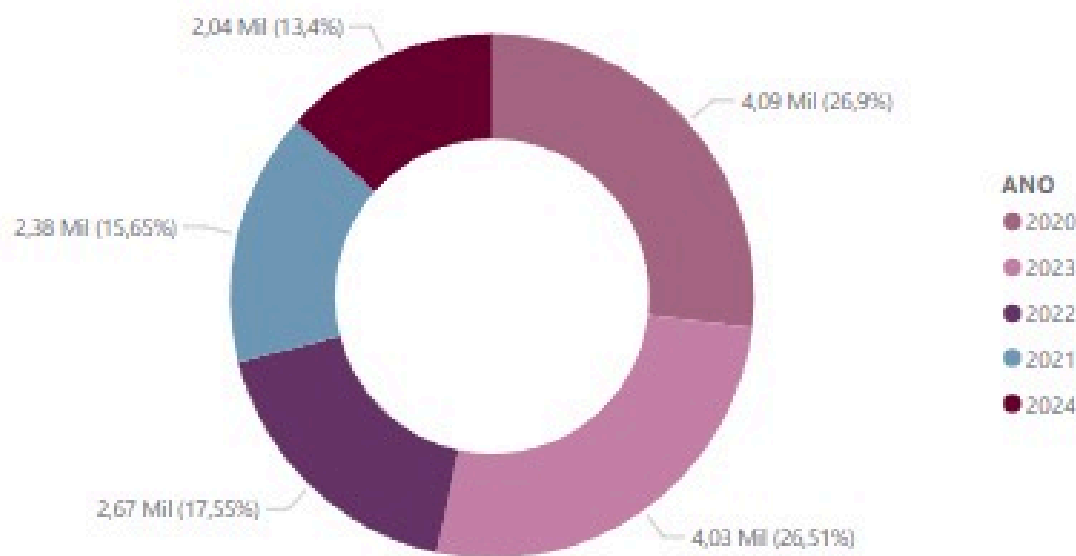
# ENTENDIMENTO DO PROBLEMA

## Acidentes de Trânsito no trânsito do Recife (2020 - 2024)

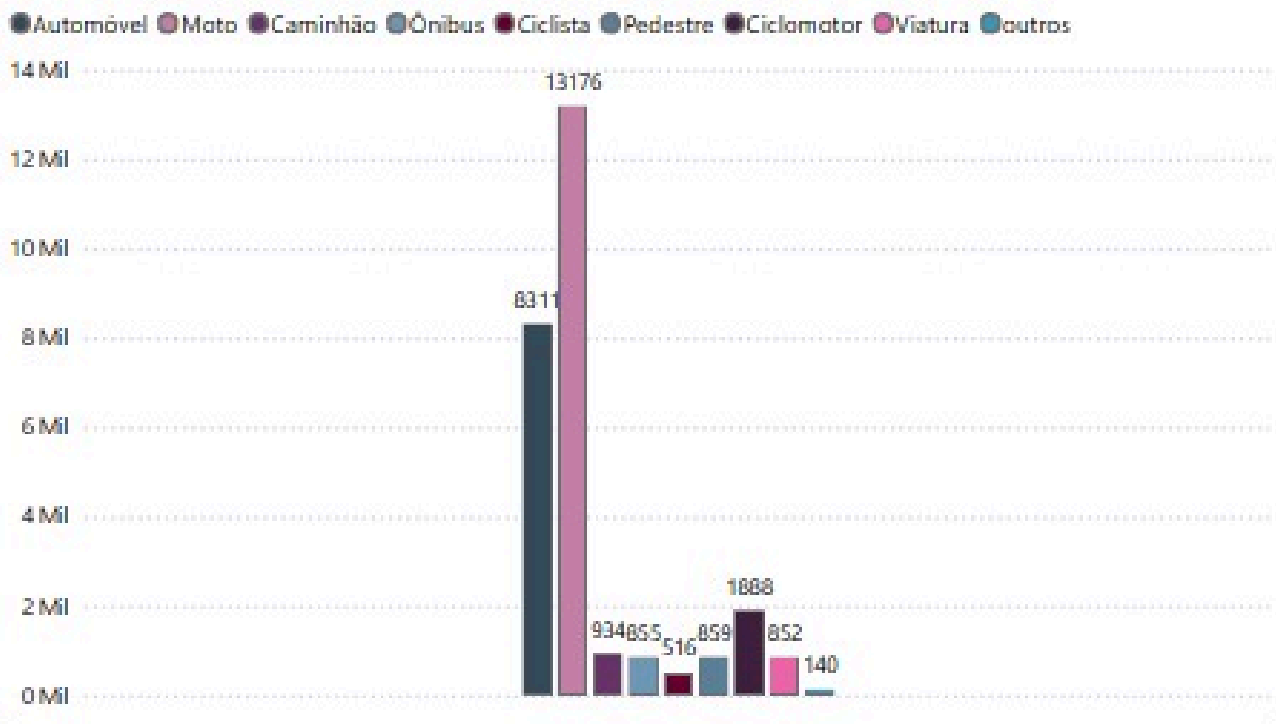
QUANTIDADE TOTAL DE OCORRÊNCIAS

15.211

QTD DE ACIDENTES POR ANO



QTD DE ACIDENTES POR CATEGORIA



QTD POR TIPO

TIPO	QUANTIDADE
ABALROAMENTO LONGITUDINAL	8802
ABALROAMENTO TRANSVERSAL	965
ACID. DE PERCURSO	640
ALAGAMENTO	11
APOIO COMPESA	3
ATROPELAMENTO	1
ATROPELAMENTO ANIMAL	156
ATROPELAMENTO DE ANIMAL	6
ATROPELAMENTO DE PESSOA	8
CAPOTAMENTO	155
CHOQUE	35
CHOQUE OBJETO FIXO	184
CHOQUE VEÍCULO PARADO	147
COLISÃO	101
COLISÃO COM CICLISTA	1102
COLISÃO FRONTAL	287
COLISÃO LATERAL	468
COLISÃO TRANSVERSAL	580
COLISÃO TRASEIRA	290
ENGAVETAMENTO	1078
MONITORAMENTO	37
OUTROS	3
OUTROS APOIOS	6
QUEDA	1
SEMÁFORO	55
TOMBAMENTO	41
Total	49
	15211

QTD DE VÍTIMAS POR ANO

2020	2089	62
ANO	Vítimas	Vítimas fatais
2021	2119	25
ANO	Vítimas	Vítimas fatais
2022	2460	25
ANO	Vítimas	Vítimas fatais
2023	3878	31
ANO	Vítimas	Vítimas fatais
2024	1897	4
ANO	Vítimas	Vítimas fatais

# PRÉ PROCESSAMENTO - JUPYTER NOTEBOOK

```
[ ] import pandas as pd
    from sklearn.preprocessing import LabelEncoder
```

```
[ ] # Carregar e padronizar as colunas
    anos = range(2015, 2025)
    dfs = []
```

```
▶ for ano in anos:
    try:
        df = pd.read_csv(f"acidentes_{ano}.csv", sep=";", encoding='utf-8') # Carregar arquivo
        df.rename(columns={"data": "Data"}, inplace=True) # Padronizar nomes de colunas
        df["Ano"] = ano # Adicionar coluna de ano
        dfs.append(df)
        print(f"Successfully loaded data for year {ano}") # Mensagem de confirmação
    except Exception as e:
        print(f"Error reading file for year {ano}: {e}")

    # Verificar quantidade de DataFrames carregados
print(f"Total DataFrames loaded: {len(dfs)}")
```

↗ Mostrar saída oculta

```
[ ] # Concatenar todos os DataFrames
    df_principal = pd.concat(dfs, ignore_index=True)
    print("Anos disponíveis:", df_principal['Ano'].unique())
```

↗ Anos disponíveis: [2015 2016 2017 2018 2019 2020 2021 2022 2023 2024]

```
[ ] # Padronizar datas no formato ISO para 2018 e 2019
    df_2018_2019 = df_principal[(df_principal['Ano'] == 2018) | (df_principal['Ano'] == 2019)]
    df_2018_2019['Data'] = pd.to_datetime(df_2018_2019['Data'], format='%d/%m/%Y').dt.strftime('%Y-%m-%d')
    df_principal.loc[df_2018_2019.index, 'Data'] = df_2018_2019['Data']
```



# PRÉ PROCESSAMENTO - JUPYTER NOTEBOOK

```
[ ] # Converter a coluna 'vitimas' para numérico, substituindo valores inválidos por 0
df_principal['vitimas'] = pd.to_numeric(df_principal['vitimas'], errors='coerce')

# Aplicar a transformação para definir 1 se >= 1, caso contrário 0
df_principal['vitimas'] = df_principal['vitimas'].apply(lambda x: 1 if x >= 1 else 0)
```

```
[ ] df_principal["tipo"] = df_principal["vitimas"]
```

```
[ ] # Substituir valores nulos restantes por uma string padrão ou outro valor
df_principal.fillna("null", inplace=True)
```

```
# Verificar as primeiras linhas dos dados processados
print(df_principal.head())
```

```
↩
```

	tipo	situacao	Data	hora	bairro \
0	0	CANCELADA	2015-12-03	11:36:00.000	CASA AMARELA
1	0	FINALIZADA	2015-06-01	06:35:00.000	ENCRUZILHADA
2	0	FINALIZADA	2015-06-01	08:04:00.000	BOA VIAGEM
3	0	FINALIZADA	2015-06-01	08:06:00.000	AFOGADOS
4	0	CANCELADA	2015-06-01	08:13:00.000	MADALENA

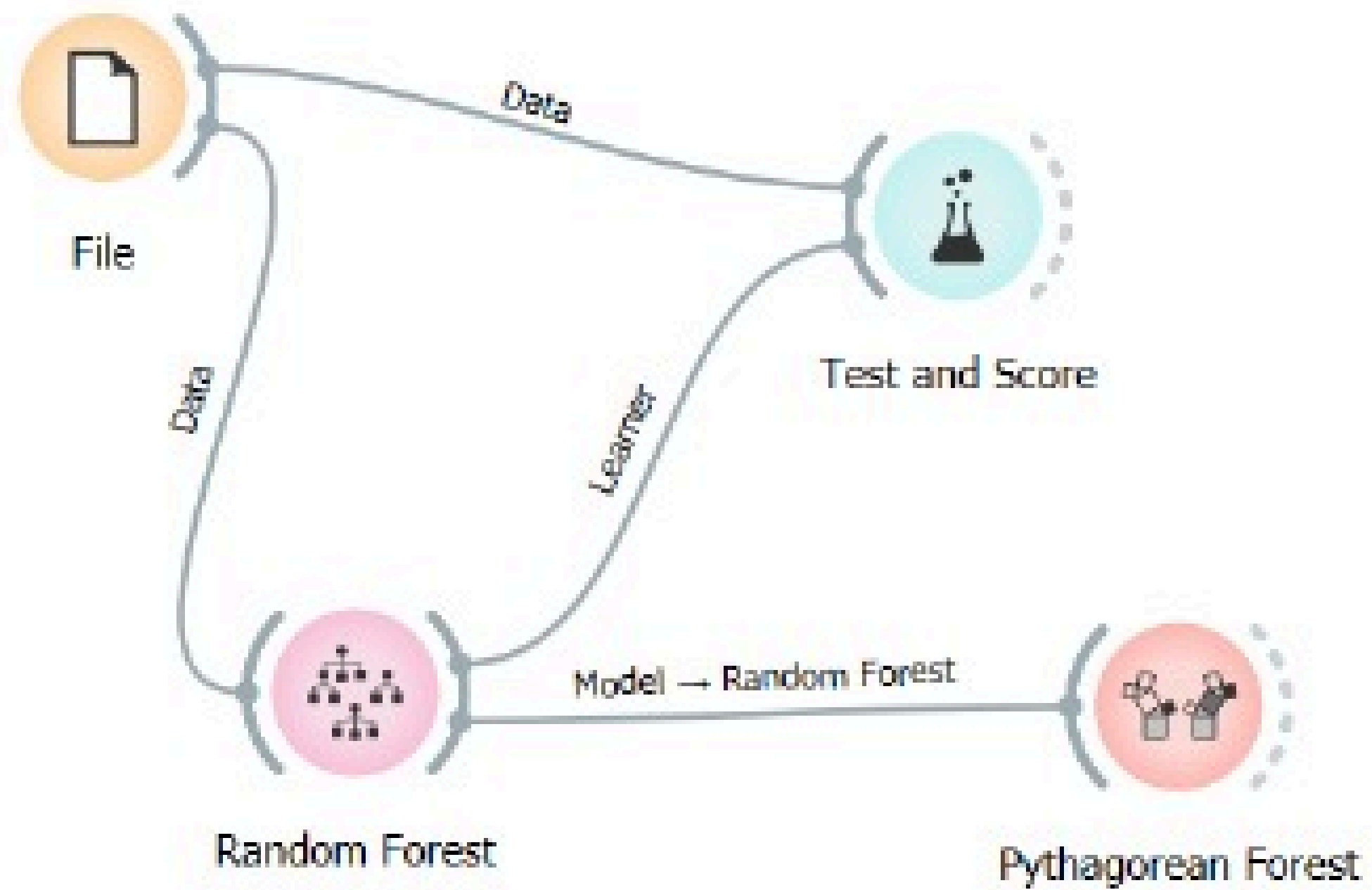
  

	endereco	numero	complemento \
0	AV NORTE	null	SENTIDO SUBURBIO E/F AO SESC
1	EST DE BELEM	null	IGREJA BATISTA MEMORIAL DE BELÉM
2	RUA ERNESTO DE PAULA SANTOS	null	SEMAFORO Nº217
3	AV SUL	null	AO LADO DA ESTAÇÃO LARGO DA PAZ
4	RUA BENFICA	null	EM FRENTE AO CLUBE INTERNACIONAL

	natureza	descricao	condicao_via	conservacao_via \
0	COLISÃO	COLISÃO SEM VITIMAS	...	null null
1	COLISÃO	SEM VÍTIMAS.	...	null null
2	COLISÃO	SEM VITIMA	...	null null

## Visualização do Aplicativo Orange





# Resultado do Teste de Score

Cross validation

Number of folds:

5

☒

Stratified

Cross validation by feature

Random sampling

Repeat train/test:

10

Training set size:

80 %

☒

Stratified

Leave one out

☒

Test on train data

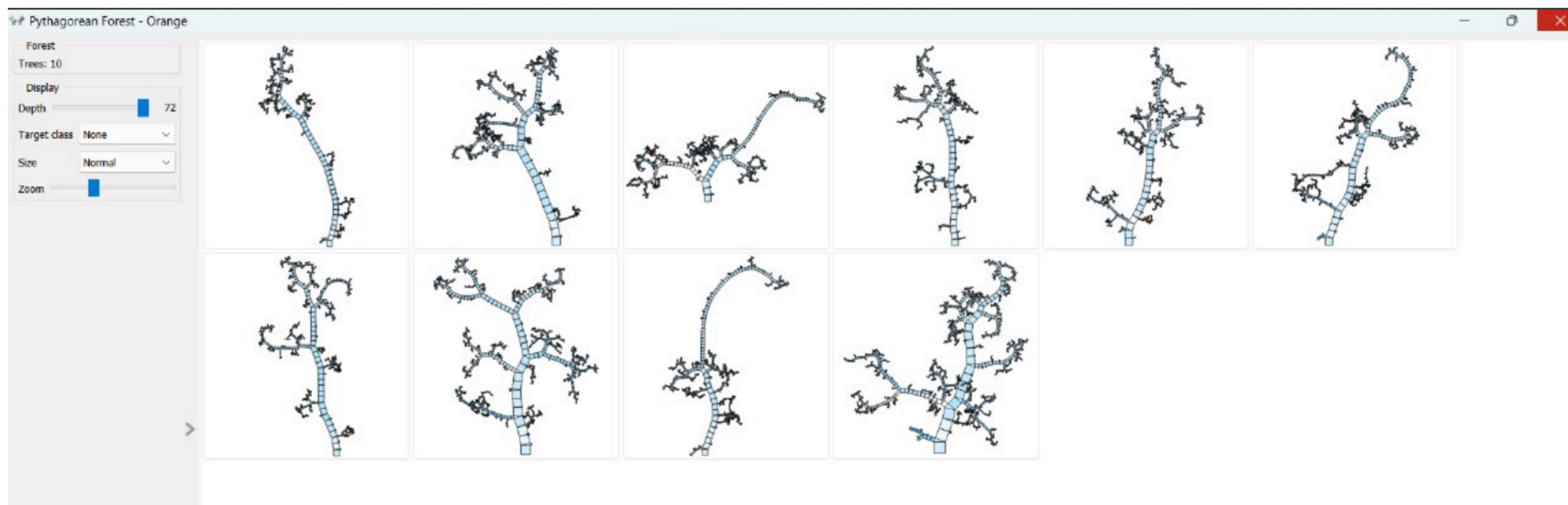
☐

Test on test data

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.903	0.893	0.880	0.894	0.893	0.629

- **AUC (ÁREA SOB A CURVA): 0,903**
  - ESSA MÉTRICA MEDE A CAPACIDADE DO MODELO DE DISTINGUIR ENTRE AS CLASSES. UM VALOR PRÓXIMO DE 1 INDICA UM DESEMPENHO EXCELENTE. NO CASO, O VALOR DE 0,903 MOSTRA QUE O MODELO TEM ALTA CAPACIDADE DE DISCRIMINAÇÃO.
- **ACURÁCIA: 0,893**
  - A ACURÁCIA REPRESENTA A PROPORÇÃO DE CLASSIFICAÇÕES CORRETAS. UM VALOR DE 0,893 SIGNIFICA QUE O MODELO ACERTOU 89,3% DAS PREDIÇÕES, O QUE É UM RESULTADO MUITO BOM.
- **F1-SCORE: 0,880**
  - O F1-SCORE É UMA MEDIDA HARMÔNICA DA PRECISÃO E DO RECALL. UM VALOR ALTO INDICA UM BOM EQUILÍBRIO ENTRE AS DUAS MÉTRICAS. NO CONTEXTO, O VALOR DE 0,880 SUGERE QUE O MODELO ESTÁ CLASSIFICANDO CORRETAMENTE A CLASSE POSITIVA (VÍTIMAS) COM ALTA PRECISÃO E RECALL.
- **PRECISÃO: 0,895**
  - A PRECISÃO MEDE A PROPORÇÃO DE EXEMPLOS POSITIVOS CORRETAMENTE CLASSIFICADOS ENTRE TODOS OS EXEMPLOS CLASSIFICADOS COMO POSITIVOS. UM VALOR DE 0,895 INDICA QUE QUANDO O MODELO PREVÊ UMA INSTÂNCIA COMO POSITIVA, ELE ESTÁ CERTO 89,5% DAS VEZES.
- **RECALL: 0,893**
  - O RECALL MEDE A PROPORÇÃO DE EXEMPLOS POSITIVOS CORRETAMENTE CLASSIFICADOS ENTRE TODOS OS EXEMPLOS POSITIVOS REAIS. UM VALOR DE 0,893 INDICA QUE O MODELO ESTÁ IDENTIFICANDO 89,3% DAS INSTÂNCIAS POSITIVAS.
- **MCC (COEFICIENTE DE MATTHEWS): 0,629**
  - O MCC É UMA MEDIDA MAIS ROBUSTA QUE CONSIDERA TODAS AS COMBINAÇÕES DE VERDADEIROS POSITIVOS, VERDADEIROS NEGATIVOS, FALSOS POSITIVOS E FALSOS NEGATIVOS. UM VALOR DE 0,629 INDICA UM BOM DESEMPENHO GERAL DO MODELO, ESPECIALMENTE CONSIDERANDO POSSÍVEIS DESBALANCEAMENTOS NAS CLASSES.

# Visualização do Resultado Random Forest no Pythagorean Forest



# Bibliografia/Site de Apoio

**CHAMADOS DE SINISTROS (ACIDENTES) DE TRÂNSITO COM E SEM VITIMAS 2015 A 2024 - CONJUNTOS DE DADOS - PORTAL DE DADOS ABERTOS DA CIDADE DO RECIFE**  
<http://dados.recife.pe.gov.br/dataset/acidentes-de-transito-com-e-sem-vitimas>

- **LIU, Y., ET AL. (2020). "PREDICTING TRAFFIC ACCIDENTS USING MACHINE LEARNING MODELS." COMPUTERS, ENVIRONMENT AND URBAN SYSTEMS, 79, 101412.**  
*PREDIZENDO ACIDENTES DE TRÂNSITO USANDO MODELOS DE APRENDIZADO DE MÁQUINA*
- **DIXON, S., & HARTENSTEIN, H. (2019). "ARTIFICIAL INTELLIGENCE IN ROAD TRAFFIC SAFETY: A REVIEW." IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 20(4), 1262-1274.**  
*INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA DO TRÂNSITO: UMA REVISÃO*
- **ZHAO, Z., ET AL. (2021). "REAL-TIME TRAFFIC ACCIDENT PREDICTION AND RISK EVALUATION USING DEEP LEARNING." JOURNAL OF TRANSPORTATION SAFETY & SECURITY, 13(2), 167-182.**  
*PREVISÃO EM TEMPO REAL DE ACIDENTES DE TRÂNSITO E AVALIAÇÃO DE RISCO USANDO APRENDIZADO PROFUNDO*