



INSIGHT

Data Science Laboratory
Federal University of Ceará





INSIGHT
DATA SCIENCE LAB

OPORTUNIDADE

CIÊNCIA DE
DADOS E
INTELIGÊNCIA
ARTIFICIAL

- **[JOB01] Cientista de Dados**
Python, Bibliotecas (Scikitlearn, Pandas, Matplotlib), Algoritmos ML, Jupyter Notebook, Banco de Dados Relacionais, NoSQL (Cassandra / MongoDB / Redis)
- **[JOB02] Engenheiro de Dados**
ETL, DW, OLAP, BD Relacional, NoSQL (Cassandra, MongoDB, Redis), Modelagem de dados
- **[JOB03] Engenheiro de Big Data**
Spark, Flink, Elasticsearch, Kafka, RabbitMQ, YARN, NoSQL, Hortonworks HDP/HDF
- **[JOB04] Engenheiro de Cloud**
AWS, Azure, Google Cloud Platform, OpenStack
- **[JOB05] Engenheiro de Software Full-Stack**
Scala, Akka, Lagom, ReactJS, React Native, Android, Swift
- **[JOB06] Engenheiro DevOps**
Kubernetes, Docker, Continuous Integration, Ansible, Monitoramento
- **[JOB07] UI/UX Designer**
Identidade Visual, Usabilidade, Design de Interfaces, HTML5, Javascript, CSS3
- **[JOB08] Editor de Vídeo**
Montagem e finalização de vídeo, Produção de vinhetas e vídeos educativos, Final Cut Pro, Adobe Premiere, Adobe After Effects
- **[JOB09] Instrutor e Desenvolvedor de Software Web Full-Stack**
Python, Java, Scala, Javascript, React

INTERESSADO?

Envia um email com o código do perfil no título e
currículo para

JOBS@INSIGHTLAB.UFC.BR



Ciência de Dados Aplicada à Segurança Pública

Prof. Dr. José Macedo
Universidade Federal do Ceará



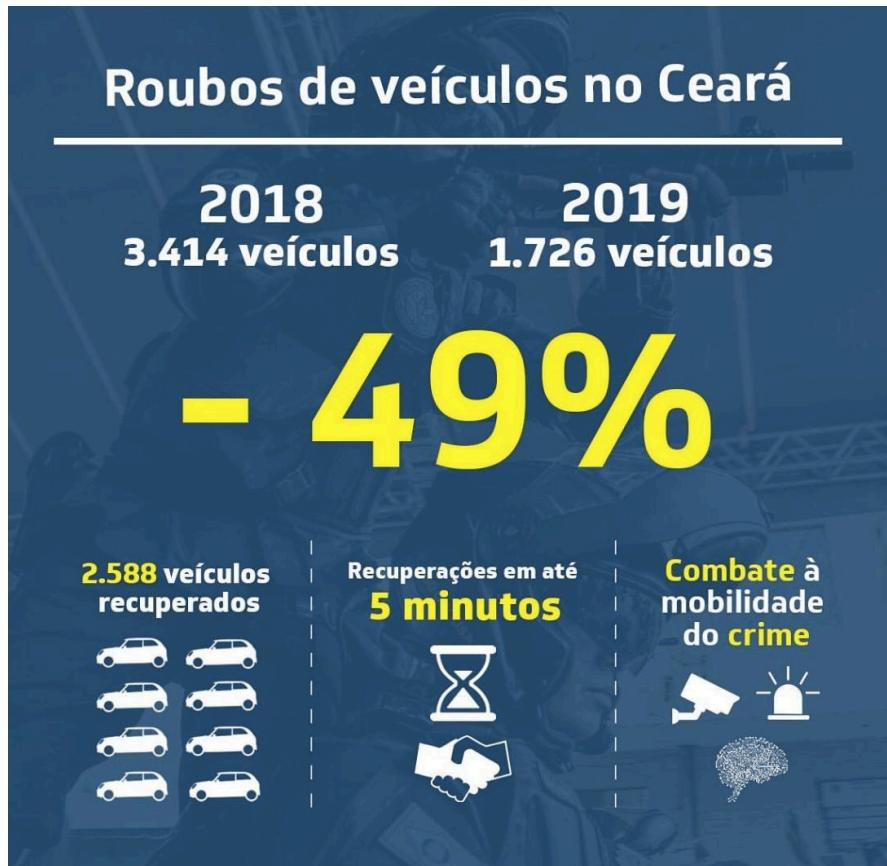
Como a inteligência artificial vai auxiliar no combate ao crime no Ceará

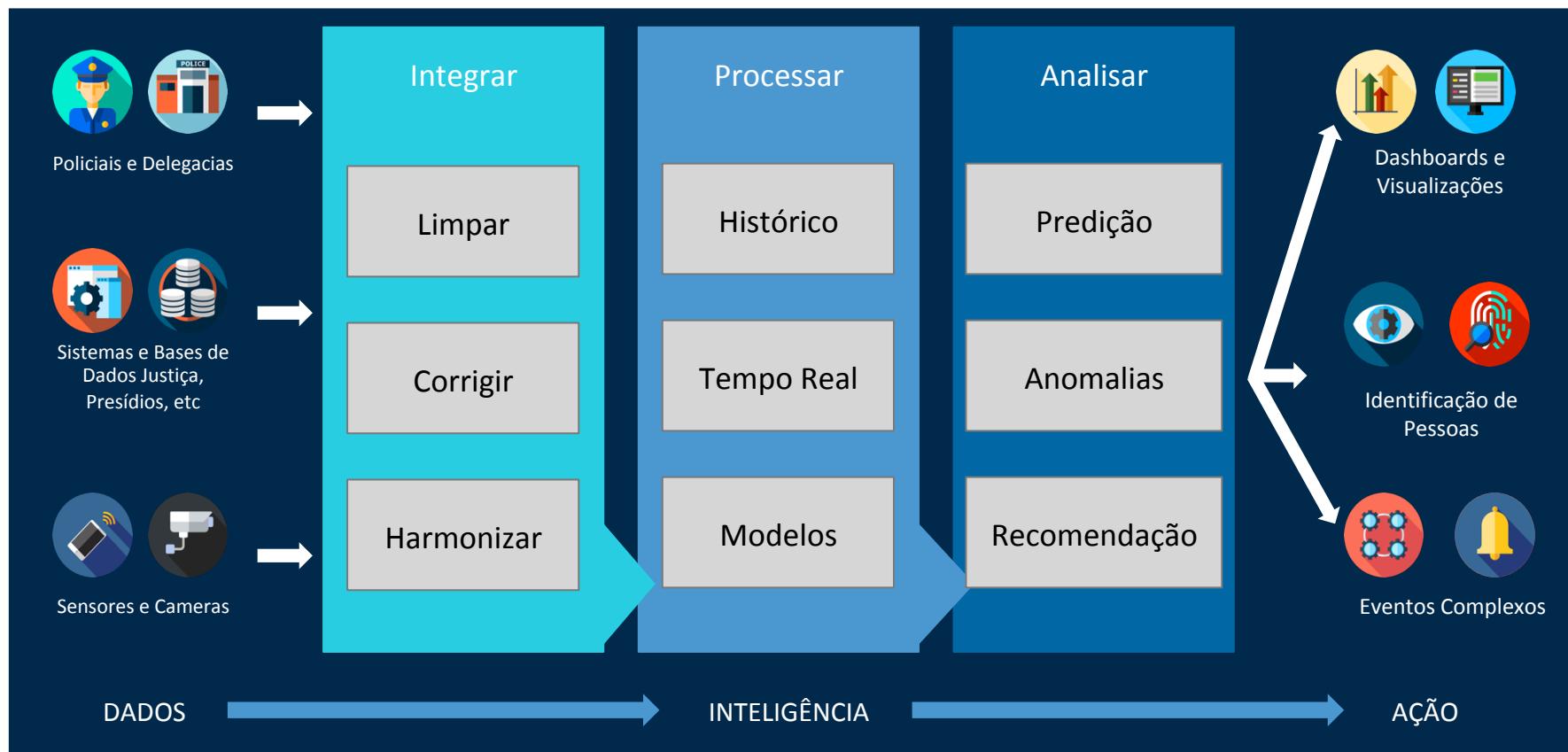
Governo estadual investe R\$ 7,5 milhões em pesquisa de inteligência artificial para auxiliar na segurança pública. Estratégia é parte de um embrião para criação do Big Data Ceará

06/05/2019 02:28:16

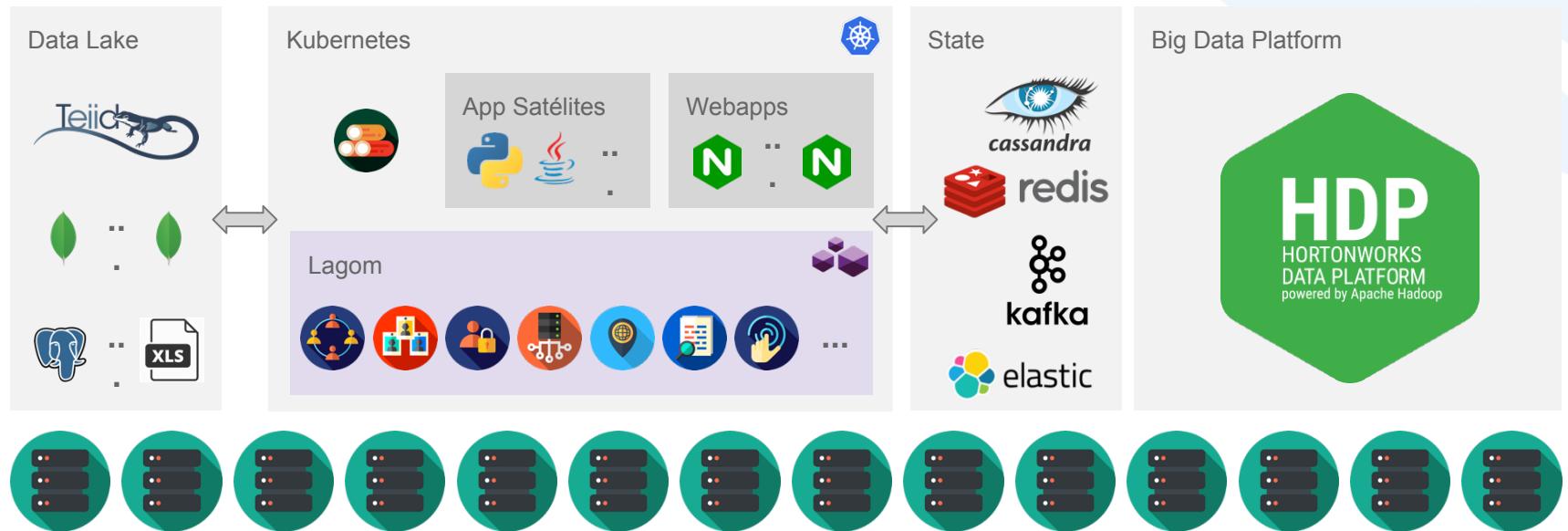
984 2

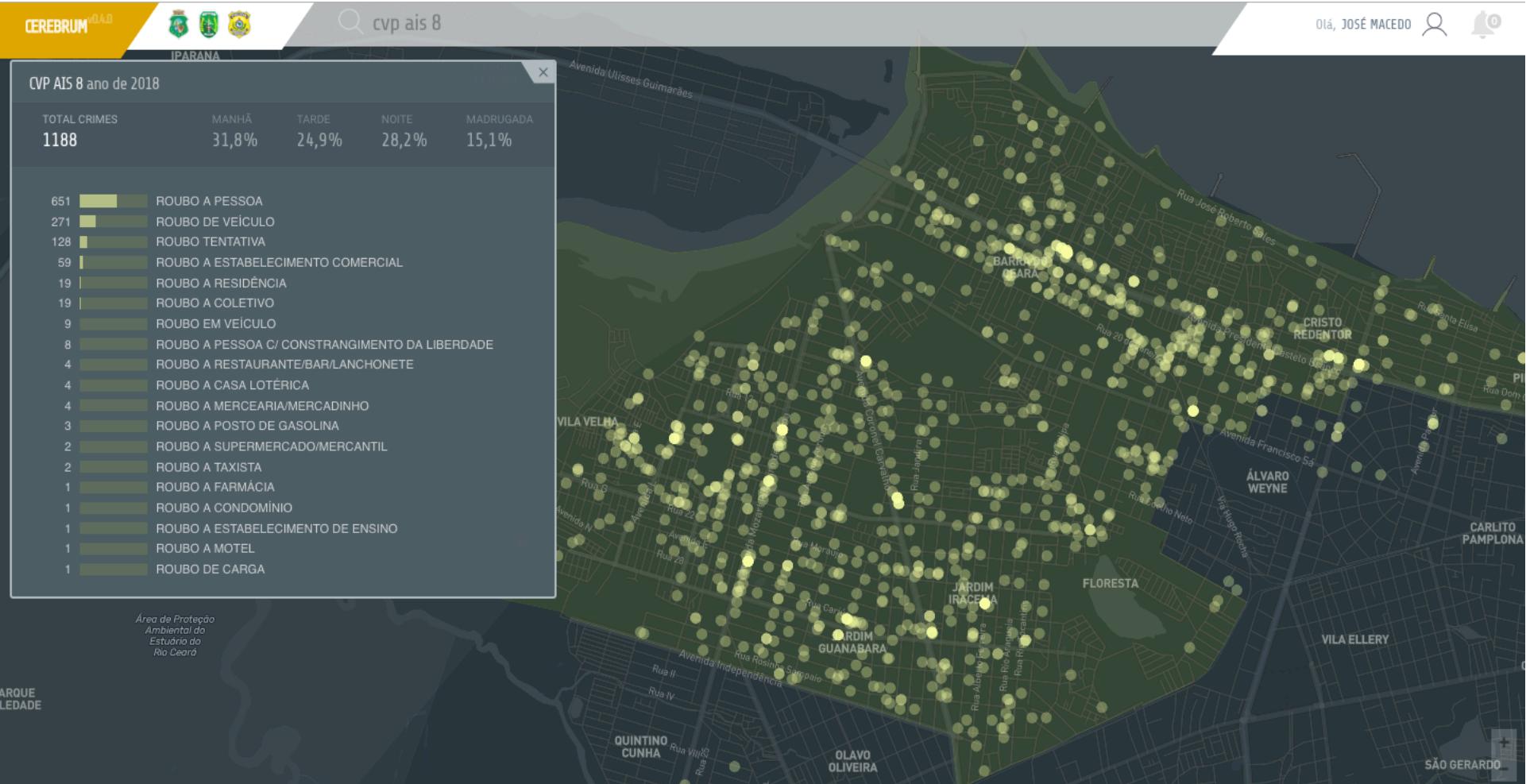






PLATAFORMA BIG DATA SEGURANÇA PÚBLICA





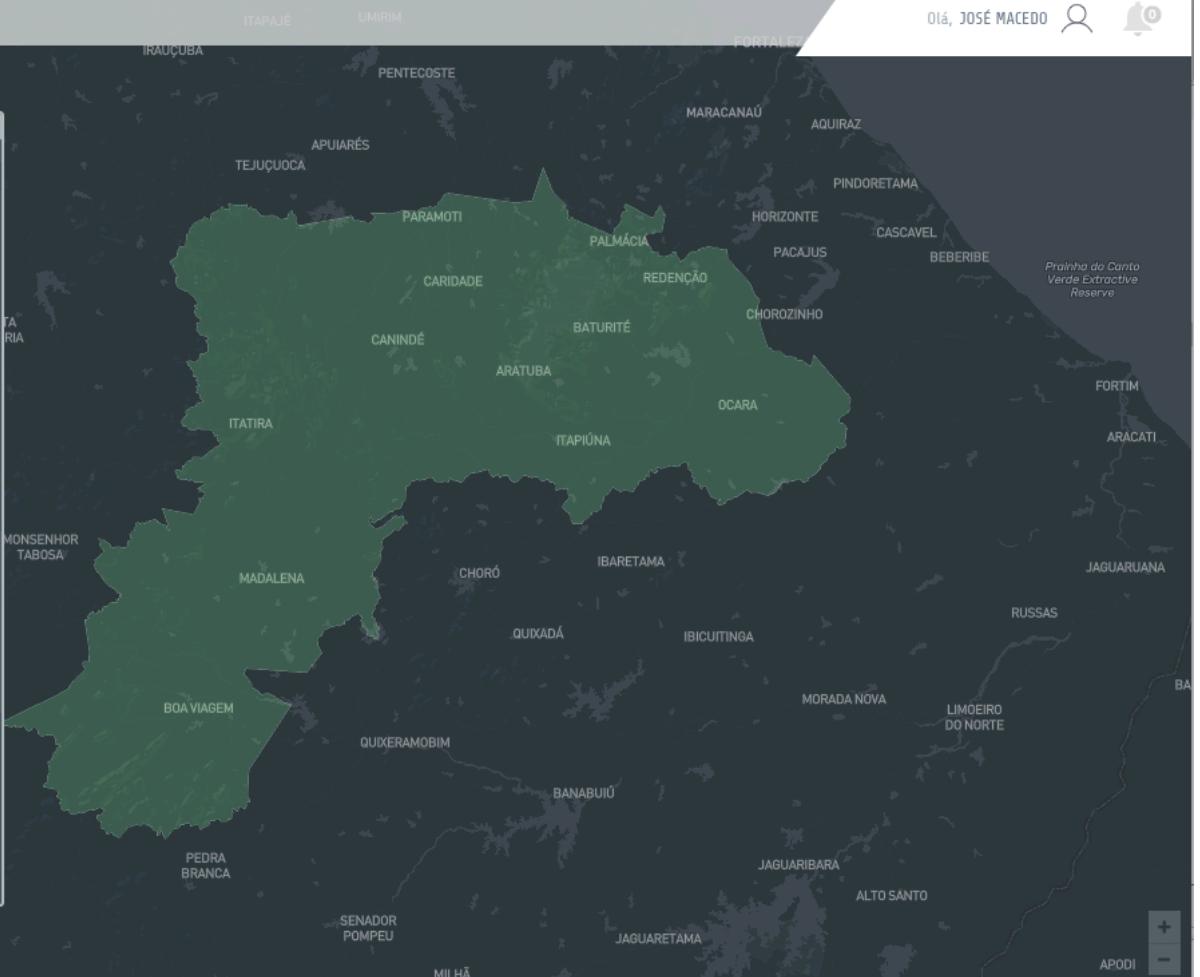


ICA
Área de Proteção
Ambiental Serra
da Ibiapaba

PEDRO II

CVLI SITUAÇÃO ATUAL 2018→2019			
	A-Z	PERCENTUAL	CRIMES
AIS 01	-50%	4 → 2	-2 ▲
AIS 02	-69,5%	23 → 7	-16 ▲
AIS 03	-78,2%	23 → 5	-18 ▲
AIS 04	-91,6%	12 → 1	-11 ▲
AIS 05	-50%	10 → 5	-5 ▲
AIS 06	-78,2%	23 → 5	-18 ▲
AIS 07	-81,5%	38 → 7	-31 ▲
AIS 08	-57,1%	21 → 9	-12 ▲
AIS 09	-80%	20 → 4	-16 ▲
AIS 10	-80%	5 → 1	-4 ▲
AIS 11	-66%	53 → 18	-35 ▲
AIS 12	-79,5%	49 → 10	-39 ▲
AIS 13	-98%	52 → 1	-51 ▲
AIS 14	-69,2%	13 → 4	-9 ▲
AIS 15	-100%	7 → 0	-7 ▲
2018 ● 2019 NENHUM REGISTRO DE CVLI PARA O ANO 2019			
AIS 16	-100%	5 → 0	-5 ▲
AIS 17	-94,1%	17 → 1	-16 ▲
AIS 18	-100%	8 → 0	-8 ▲
AIS 19	-20%	10 → 8	-2 ▲
AIS 20	-100%	5 → 0	-5 ▲
AIS 21	+0%	0 → 0	+0
AIS 22	-100%	4 → 0	-4 ▲

1 VISUALIZAÇÃO



CEREBRUM^{4.8.0}

Tabo dos Anões

viaturas online

Olá, JOSÉ MACEDO

Área de Proteção Ambiental do Estuário do Rio Ceará

Terra Indígena Tapera

FORTALEZA

Parque Estadual do Cocó

VIATURAS online

Somente recentes

Filtro...

209 / 209

ID	Placa	Status	Velocidade	Tempo
38321	RP20092	IGNIÇÃO LIGADA	0 km/h	+0:00:21
209071	RP18401	IGNIÇÃO LIGADA	0 km/h	+0:00:21
285001	CPC24	IGNIÇÃO LIGADA	0 km/h	+0:00:21
34115	RP6181	IGNIÇÃO LIGADA	0 km/h	+0:00:22
200793	RP17551	IGNIÇÃO LIGADA	0 km/h	+0:00:22
285023	RP18361	IGNIÇÃO LIGADA	0 km/h	+0:00:22
92722	RP8531	IGNIÇÃO LIGADA	0 km/h	+0:00:23
209782	RP5152	IGNIÇÃO LIGADA	0 km/h	+0:00:23
223509	RP16113	IGNIÇÃO LIGADA	0 km/h	+0:00:23
283960	RP12251	IGNIÇÃO LIGADA	0 km/h	+0:00:23
284394	RP12142	IGNIÇÃO LIGADA	0 km/h	+0:00:23
284500	RP12261	IGNIÇÃO LIGADA	0 km/h	+0:00:23
284515	R28	IGNIÇÃO LIGADA	0 km/h	+0:00:23
285052	RP1503	IGNIÇÃO LIGADA	0 km/h	+0:00:23
285059	R33	IGNIÇÃO LIGADA	0 km/h	+0:00:23
285153	CCPM06	IGNIÇÃO LIGADA	0 km/h	+0:00:23
283932	RP3281	IGNIÇÃO LIGADA	0 km/h	+0:00:24

1 VISUALIZAÇÃO

+

-

696798 - CP17082



CÂMERAS DE MONITORAMENTO

Filtro...

13 / 13

CÓDIGO	AIS	CÂMERA
173	1	69 - Av. Zezé Diogo - Areninha Praia do Futuro
176	1	265 - Av. Rui Barbosa - Rua João Carvalho
177	1	266 - Av. Rui Barbosa - Rua Deputado Moreira da Rocha
179	1	278 - Av. Luis Vieira - Rua Roberto Pedrosa
180	1	279 - Av. Luis Vieira - Rua Herminia Bonavides
181	1	280 - Rua Herminia Bonavides - Rua Oliveira Filho
182	1	285 - Av. César Cals - Rua Professor Henrique Firmeza
183	1	286 - Av. César Cals - Rua Professor Alvaro Costa
184	1	287 - Av. César Cals - Av. Renato Braga
185	1	307 - Av. Vicente de Castro - Av. José Saboia
186	1	308 - Av. Vicente de Castro - Av. Zezé Diogo
781	2	375
2	6	349

CÂMERA 2

10-23-2018 Tue 23:46:39

349
01 LPR

MARACANAÚ





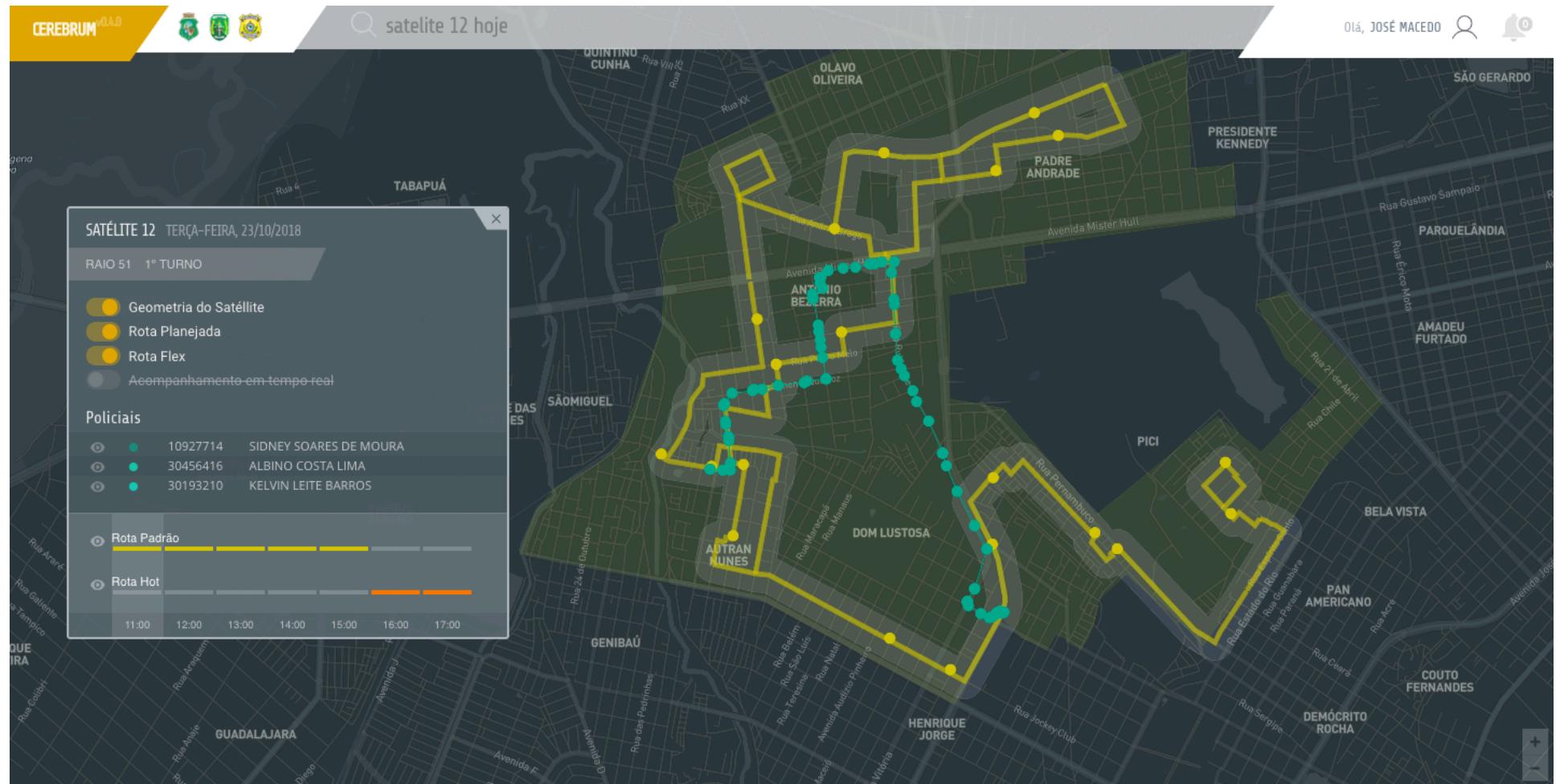
SATÉLITE 12 TERÇA-FEIRA, 23/10/2018

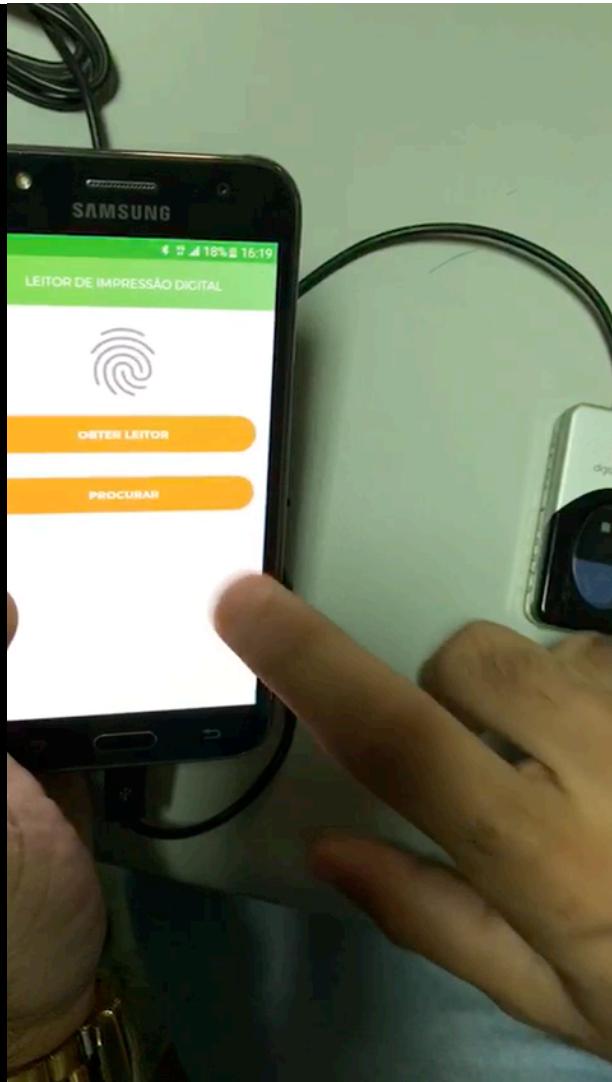
RAIO 51 1º TURNO

- Geometria do Satélite
- Rota Planejada
- Rota Flex
- Acompanhamento em tempo real

Policiais

- | | | |
|----------------------------------|----------|------------------------|
| <input checked="" type="radio"/> | 10927714 | SIDNEY SOARES DE MOURA |
| <input checked="" type="radio"/> | 30456416 | ALBINO COSTA LIMA |
| <input checked="" type="radio"/> | 30193210 | KELVIN LEITE BARROS |





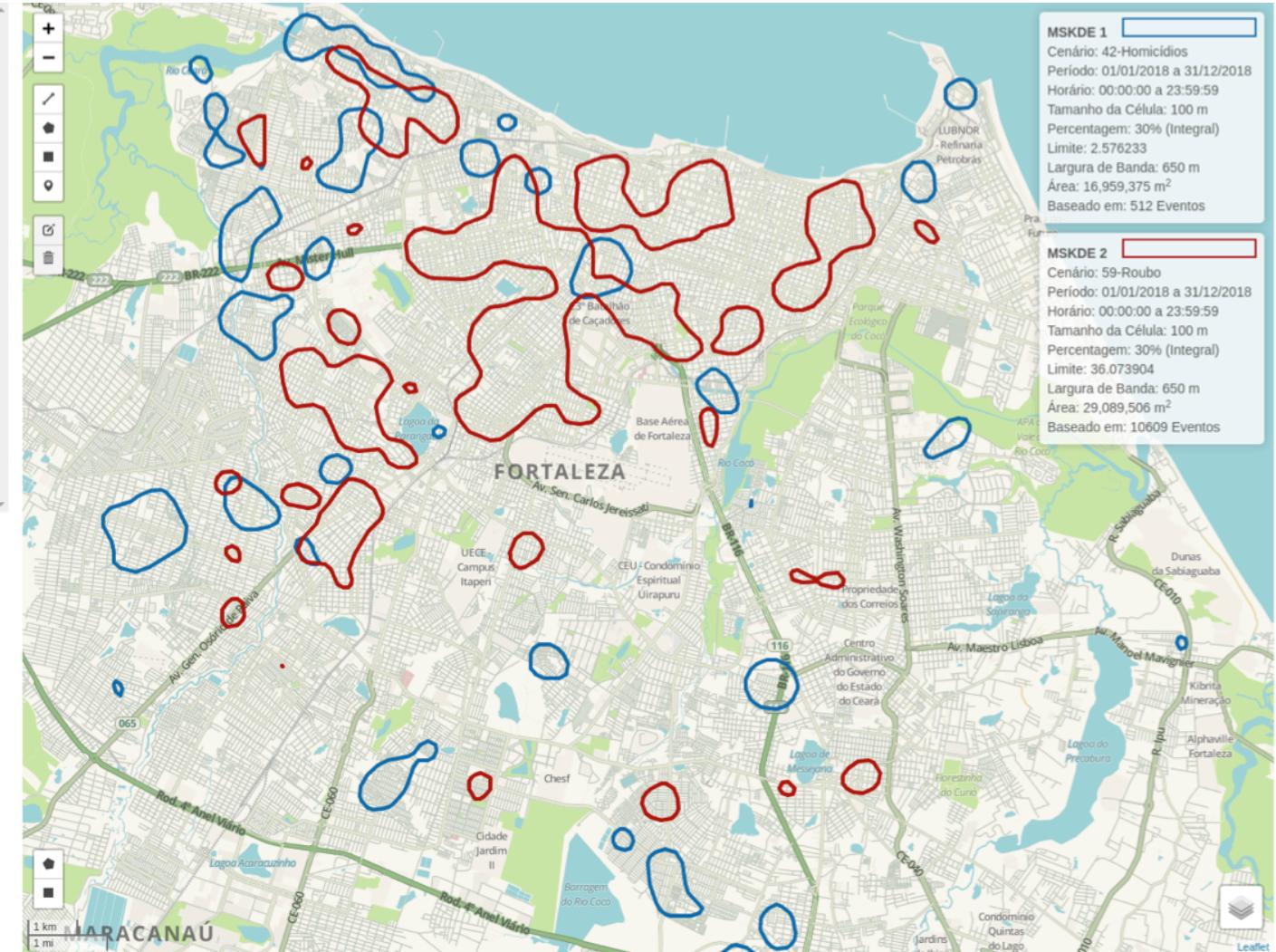
PROBLEMA:

Otimização do Patrulhamento



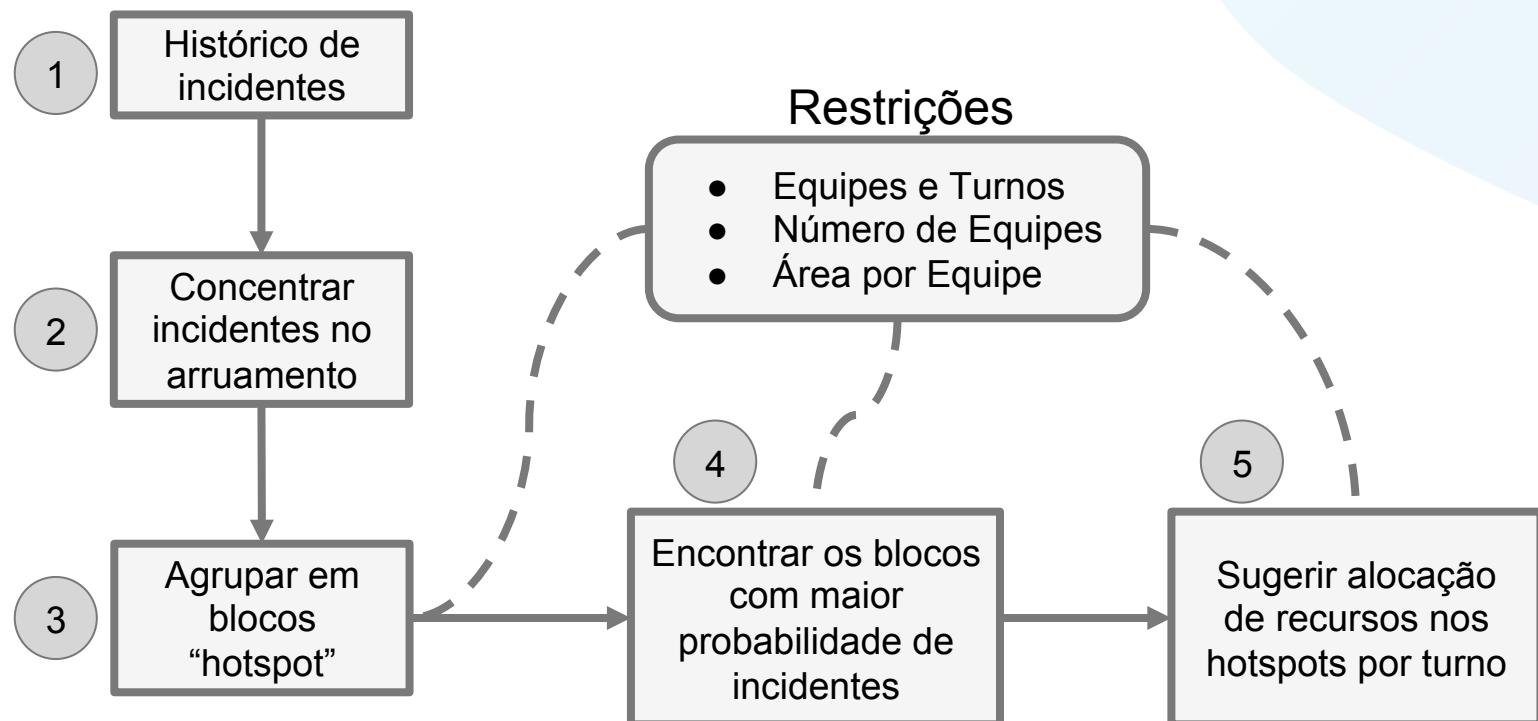
Desenvolver Técnicas para Predição de Crimes para ajudar no Patrulhamento

- [Salvar Análise](#)
- [Seleção de Cenário](#)
- [Seleção de Janela de Tempo](#)
- [Eventos](#)
- [Pontos de Interesse](#)
- [Divisões Administrativas](#)
- [Estimativa de Densidade K](#)
- [MSKDE](#)
- [Filtro Espacial](#)
- [Anotações](#)
- [Camadas](#)
- [Base](#)

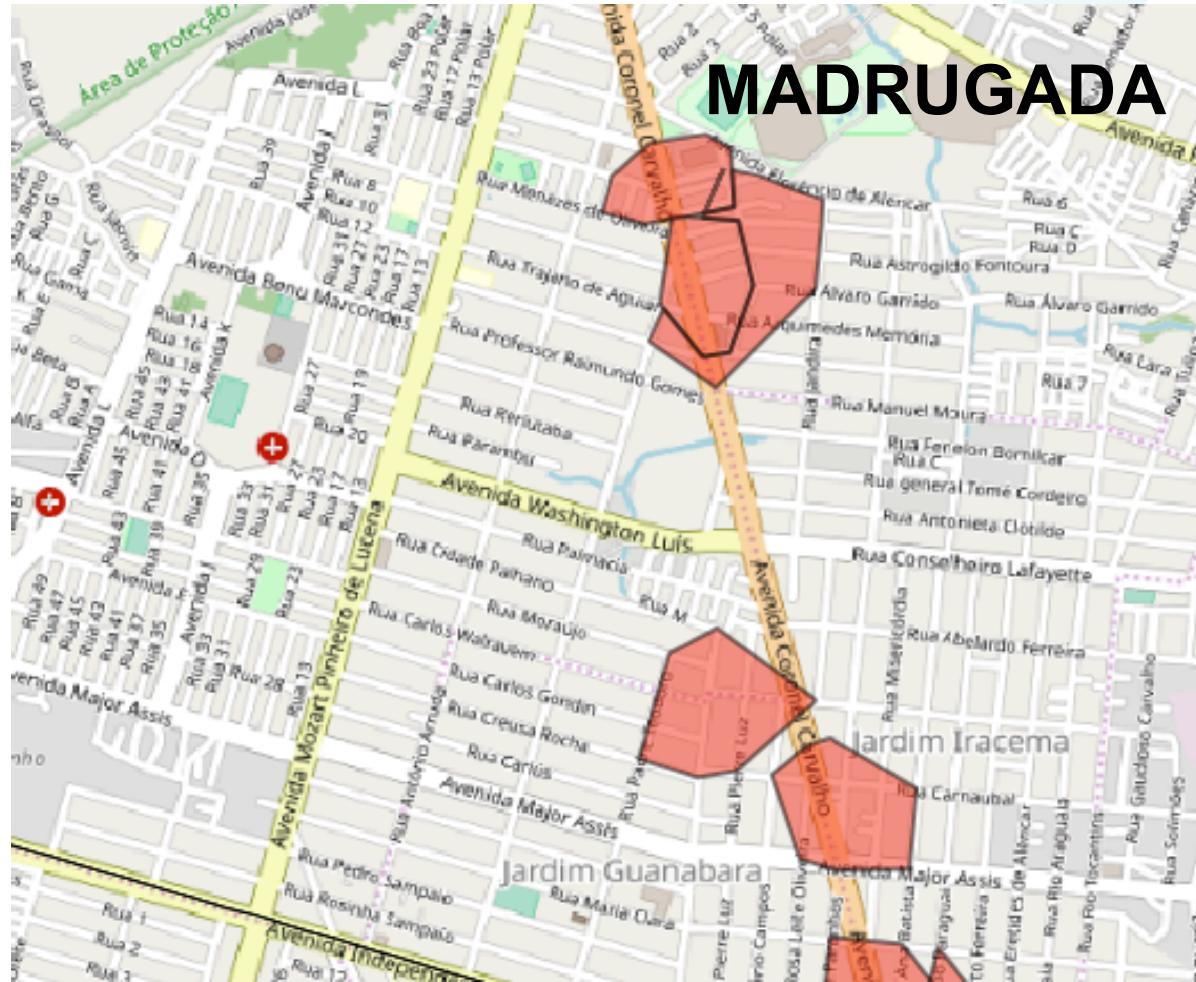


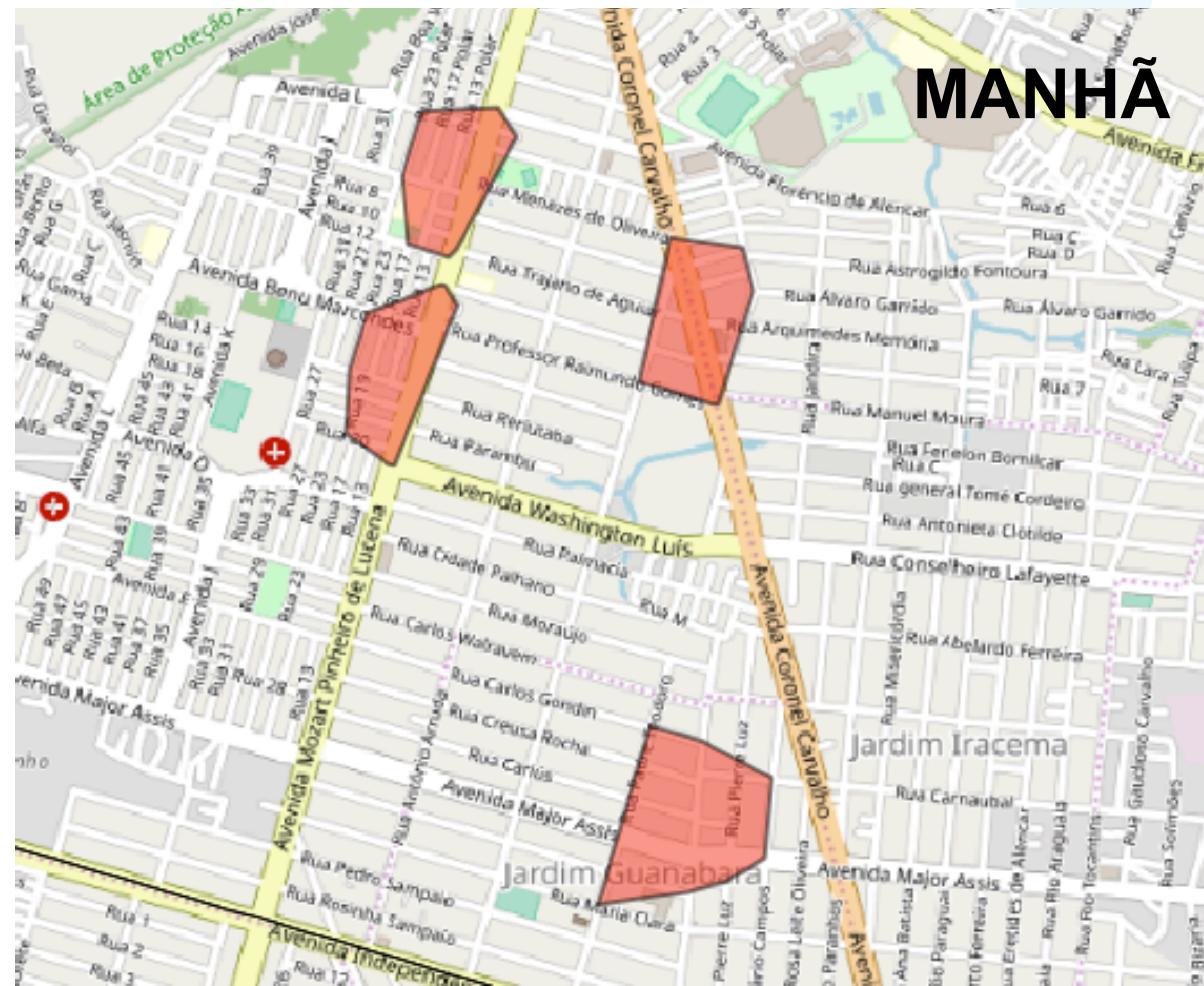
1

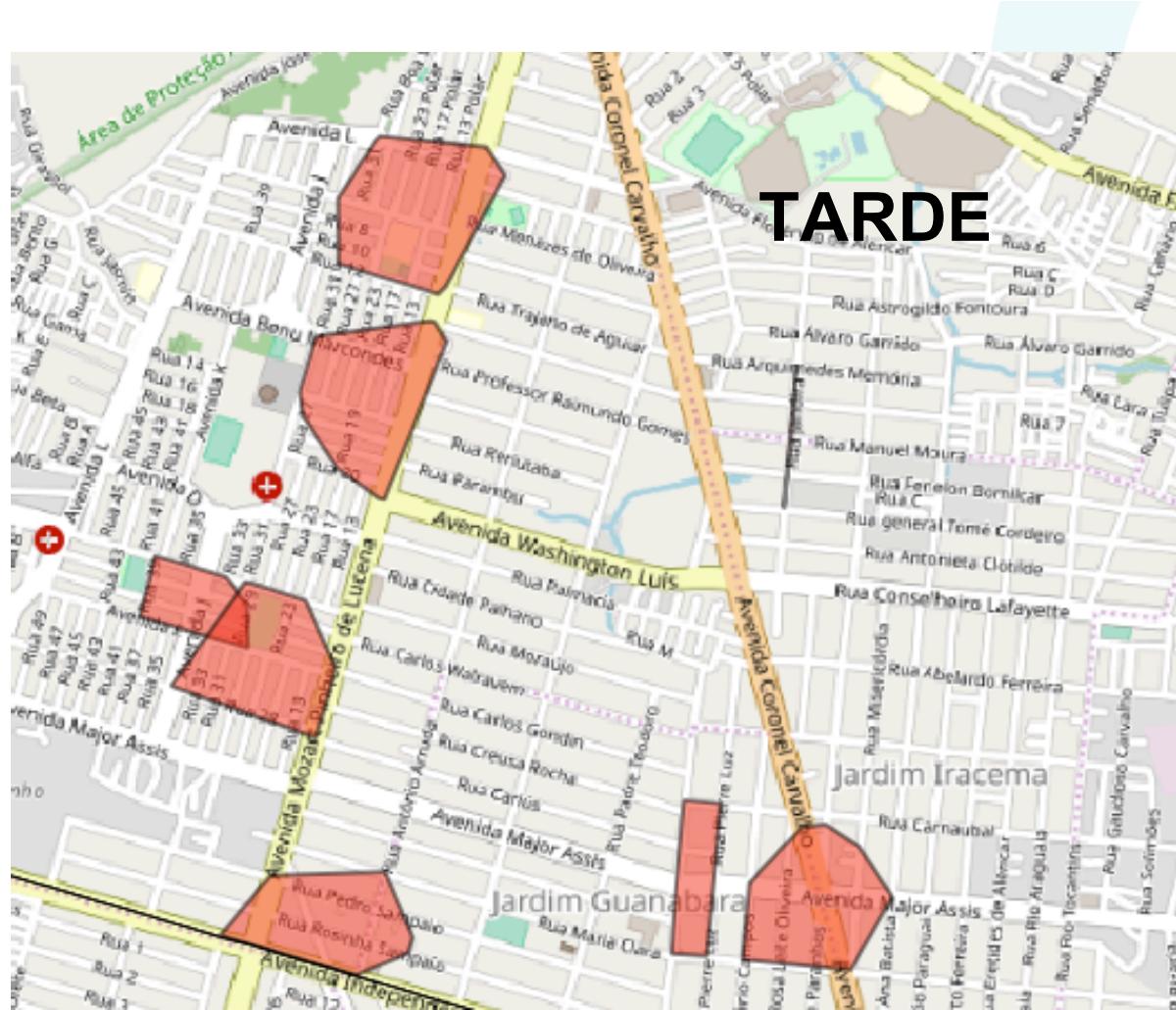
Módulo para patrulhamento baseado em ruas



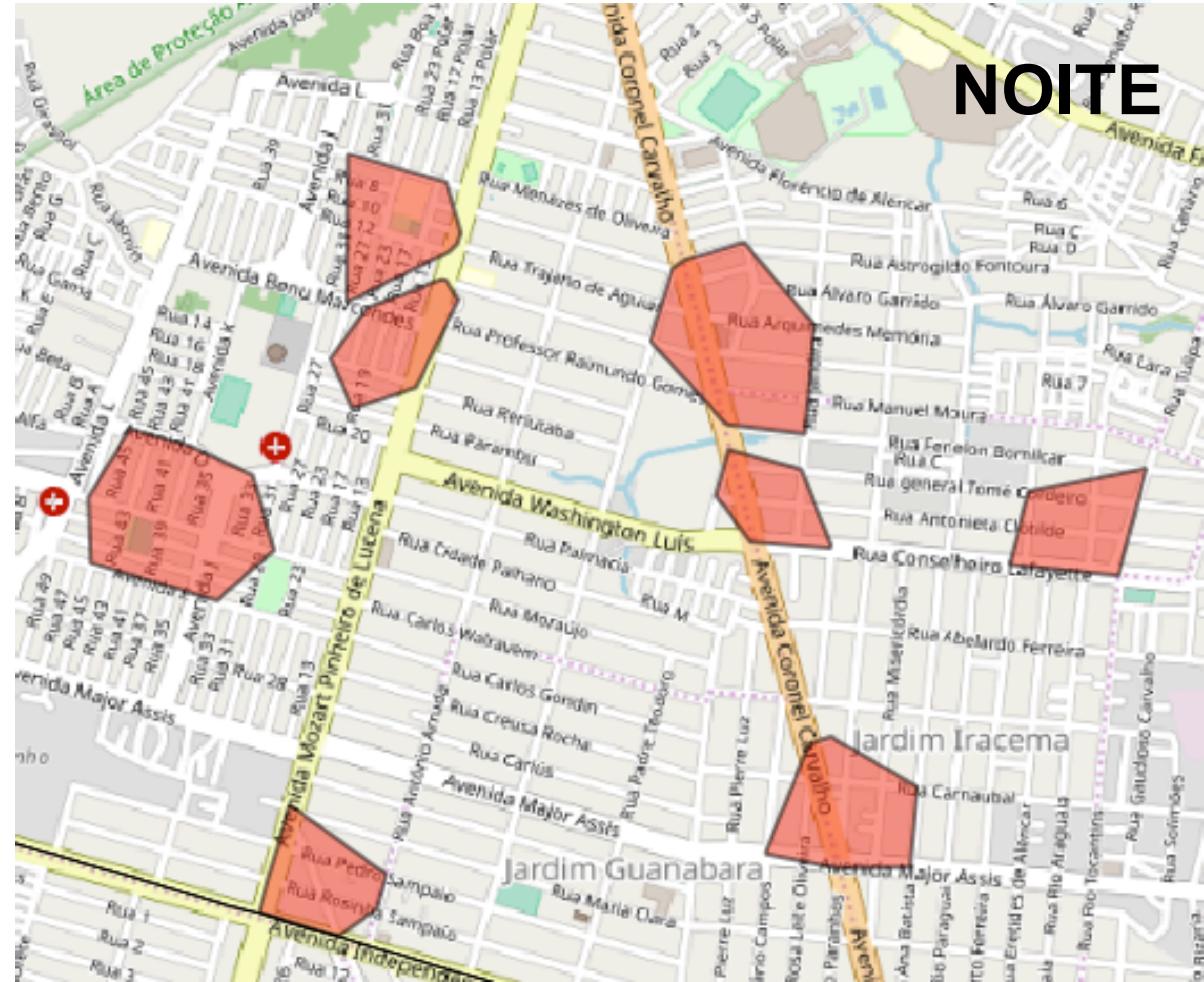
MADRUGADA







NOITE



5

O PADRÃO CRIMINOLÓGICO

HETEROGENEIDADE DE RISCO

Certas localidades são mais propensas a ocorrências de crimes que outras. Essa propriedade é geralmente estática e muda lentamente com o tempo.

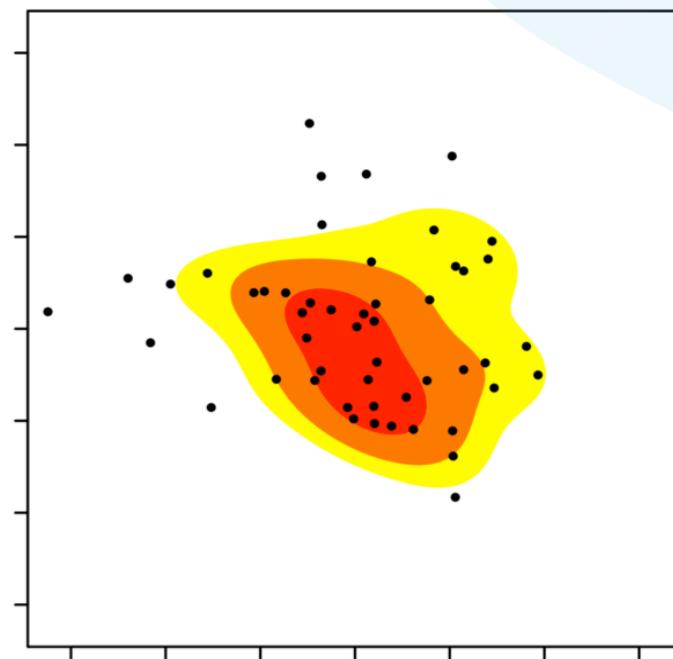
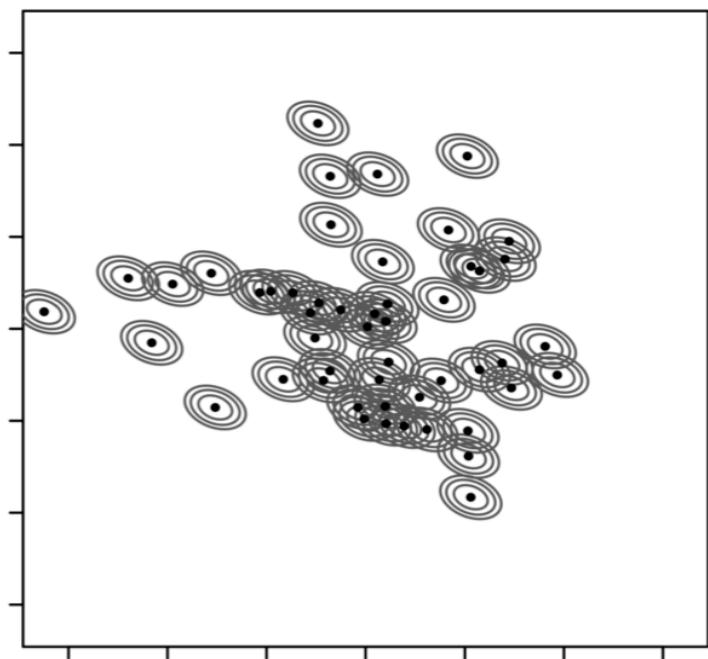
DEPENDÊNCIA DE EVENTO

A ocorrência de um crime aumenta a probabilidade de futuros incidentes na vizinhança. Este efeito é temporário, mas pode durar por semanas.

Referência: CPC: Crime, Policing and Citizenship – Intelligent policing and big data

24

TÉCNICA: Kernel Density Estimation (KDE)



MÉTRICA: PREDICTION ACCURACY INDEX (PAI)

$$PAI = \frac{\text{hit rate}}{\text{proportion of hot spot area}} = \frac{n/N}{a/A}$$

n = number of crimes in predicted crime area

N = number of crimes in study area

a = total area of predicted crime

A = area of the study area

Técnica Usando Grid

Criar uma grid regular na área de estudo.

ETAPAS PRINCIPAIS

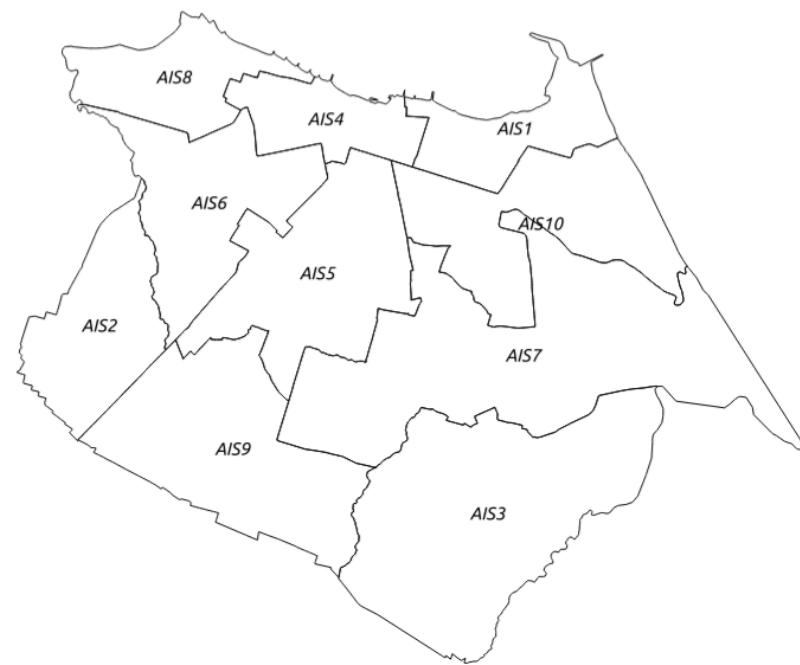
1

2

3

4

SELECCIONAR A ÁREA DE ESTUDO



ETAPAS PRINCIPAIS

1

2

3

4

CRIAR A GRID REGULAR



ETAPAS PRINCIPAIS

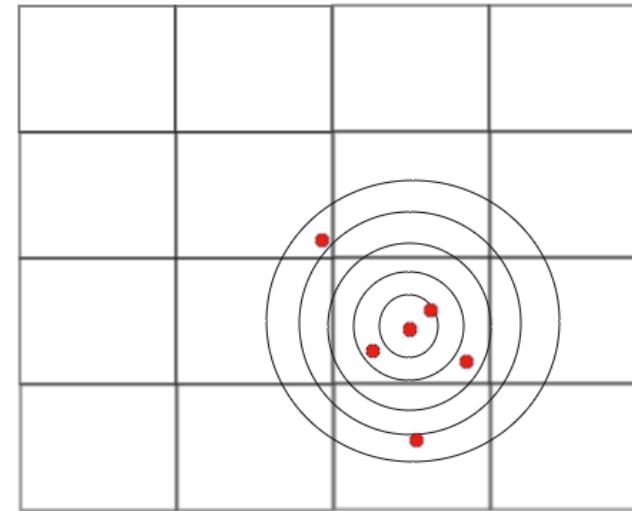
1

2

3

4

PROJETAR O KDE DOS EVENTOS NAS CÉLULAS



ETAPAS PRINCIPAIS

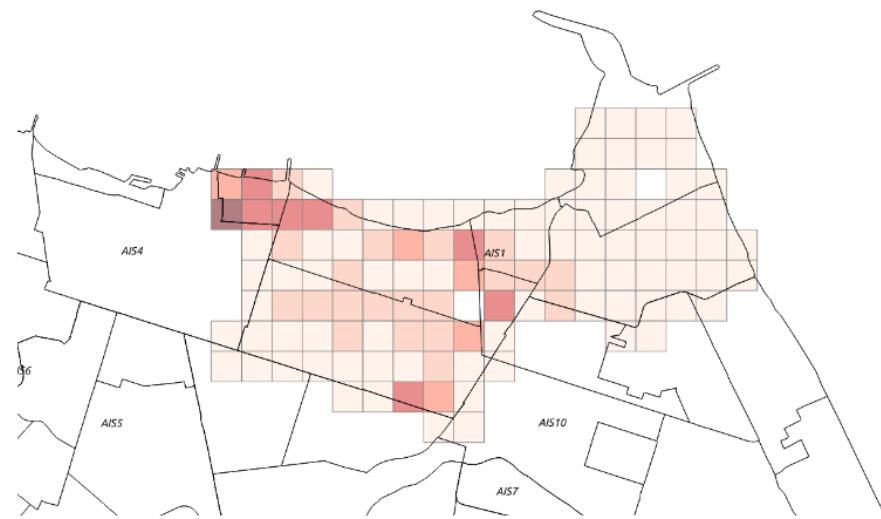
1

2

3

4

ESCOLHER AS CÉLULAS COM OS MAIORES ESCORES



ETAPAS PRINCIPAIS

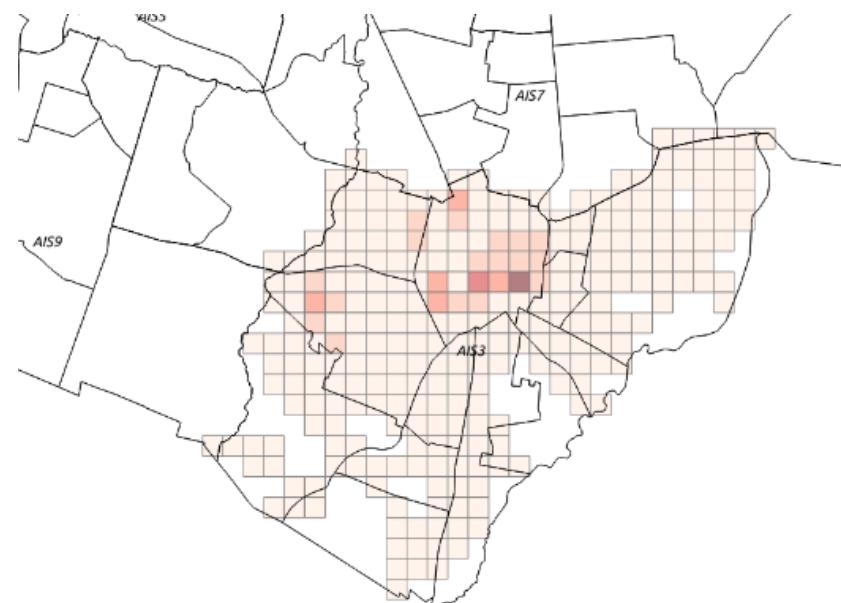
1

2

3

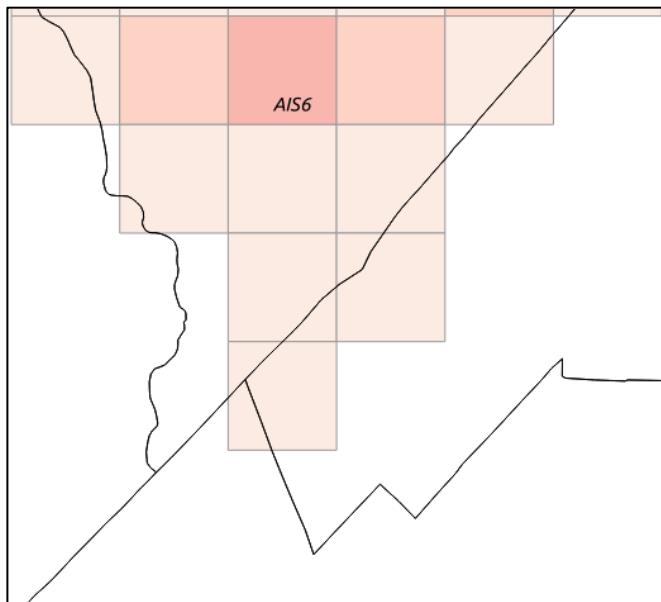
4

ESCOLHER AS CÉLULAS COM OS MAIORES ESCORES



PROBLEMAS

- Células entre áreas de estudo diferentes



- Células entre barreiras geográficas



Técnica Usando Arruamento

Projetar os eventos em nós de ruas e criar clusters de geometria dinâmica.



ETAPAS PRINCIPAIS

1

2

3

4

MAPEAMENTO DE EVENTO SOBRE AS INTERSECÇÕES DAS RUAS



ETAPAS PRINCIPAIS

1

2

3

4

PROJETAR OS EVENTOS COM O KDE NOS NÓS



35

ETAPAS PRINCIPAIS

1



2



3

CLUSTERIZAR OS NÓS LEVANDO EM CONSIDERAÇÃO A DISTÂNCIA E O ESCORE



ETAPAS PRINCIPAIS

1

2

3

4

SELECIONAR OS CLUSTERS COM MAIORES ESCORES



ETAPAS PRINCIPAIS

1

2

3

4

DEFINIR A MATRIZ DE DISTÂNCIA ENTRE OS NÓS

Longitude

Latitude

Escore

0	[-38.56507476937193,	-3.741666758246722,	5.05183878764381]
1	[-38.56640544309628,	-3.7419670166514893,	4.330221128474968]
2	[-38.56397394606413,	-3.7414585145143833,	5.315014101977351]
3	[-38.56288578874363,	-3.7412323893882125,	5.332718979800232]
4	[-38.56179036710688,	-3.7410135285782857,	4.80180142366419]
5	[-38.546134089219585,	-3.7471136915783663,	0.827287582118793]

$$distance(a, b) = \alpha * \exp(\gamma * (euclidean(a, b))) + |a.prob - b.prob|$$

ETAPAS PRINCIPAIS

1

2

3

4

DEFINIR O NÚMERO DE CLUSTERS

$$n = \frac{\text{Área da AIS}}{\text{Área da célula da GRID}}$$

ETAPAS PRINCIPAIS

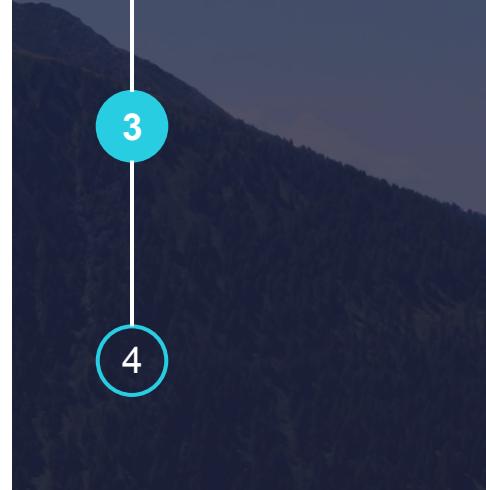
1



2

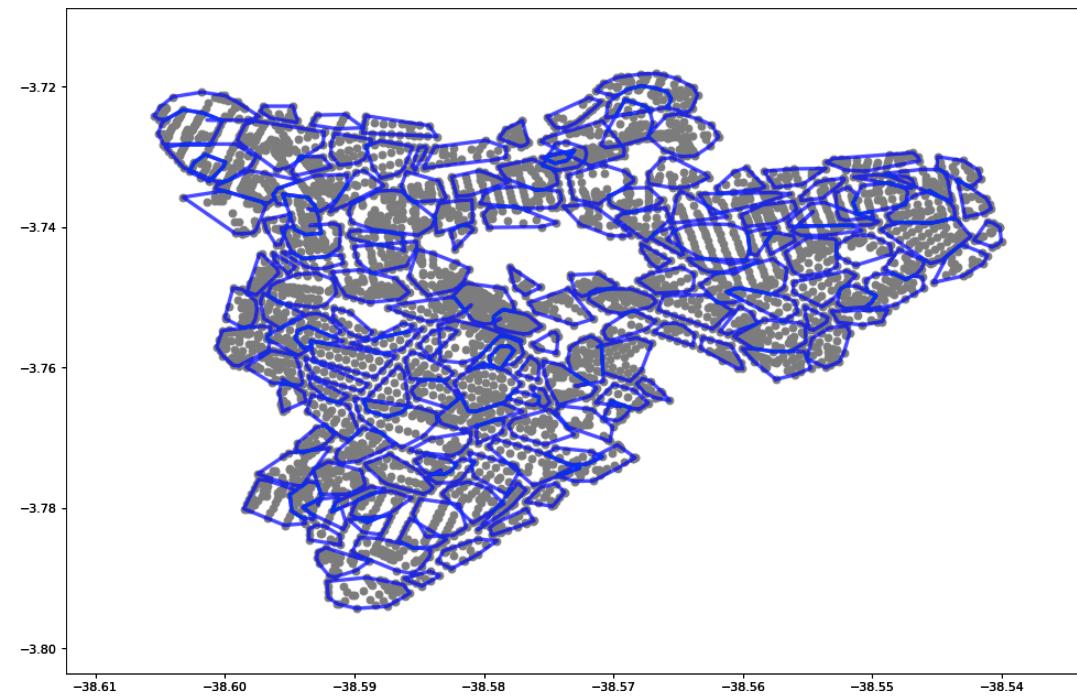


3



4

CLUSTERIZAR E CRIAR A GEOMETRIA DAS CÉLULAS



ETAPAS PRINCIPAIS

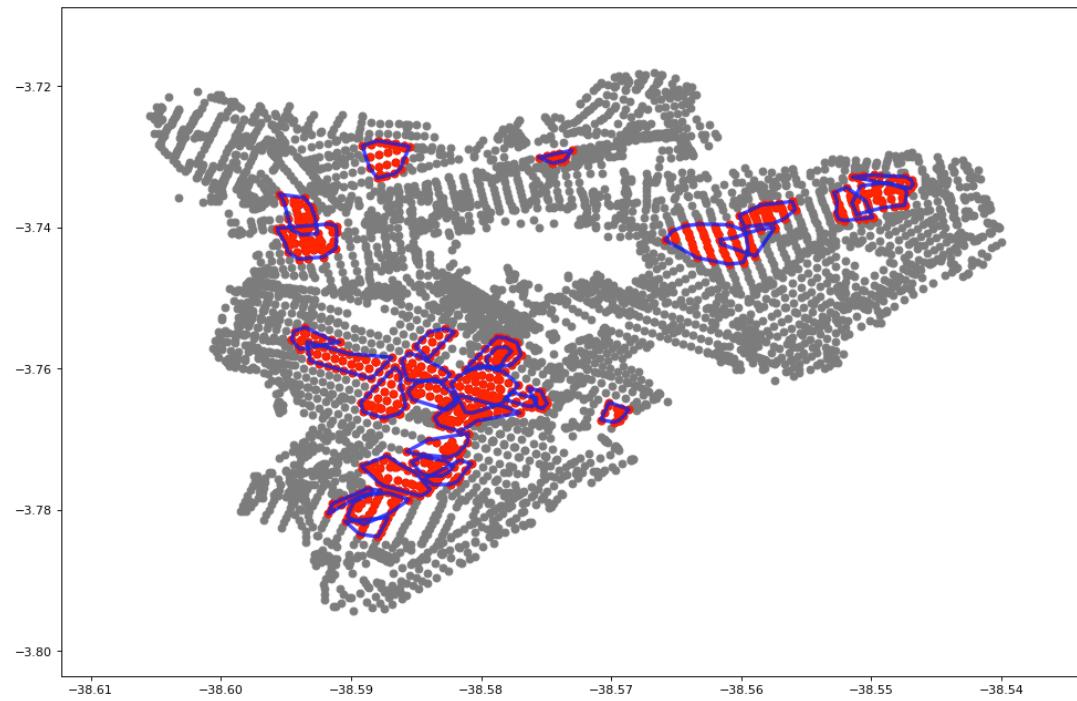
1

2

3

4

SELECIONAR AS CÉLULAS QUE POSSUEM OS MAIORES ESCORES



RESULTADOS

Resultado do modelo de grid (baseline):

AIS6	AIS8	AIS4	AIS5	AIS10	AIS1	AIS7	AIS3	AIS9	AIS2	mean	std
2.463252	2.500429	1.808779	2.731763	3.083716	2.268861	5.423122	8.063044	4.077971	2.431437	3.485237	1.822275

Resultado do modelo de clusters de nós:

AIS6	AIS8	AIS4	AIS5	AIS10	AIS1	AIS7	AIS3	AIS9	AIS2	mean	std
3.199995	3.903972	2.669808	3.774036	5.356855	3.490556	9.369791	12.763474	5.658429	3.614298	5.380121	3.066842

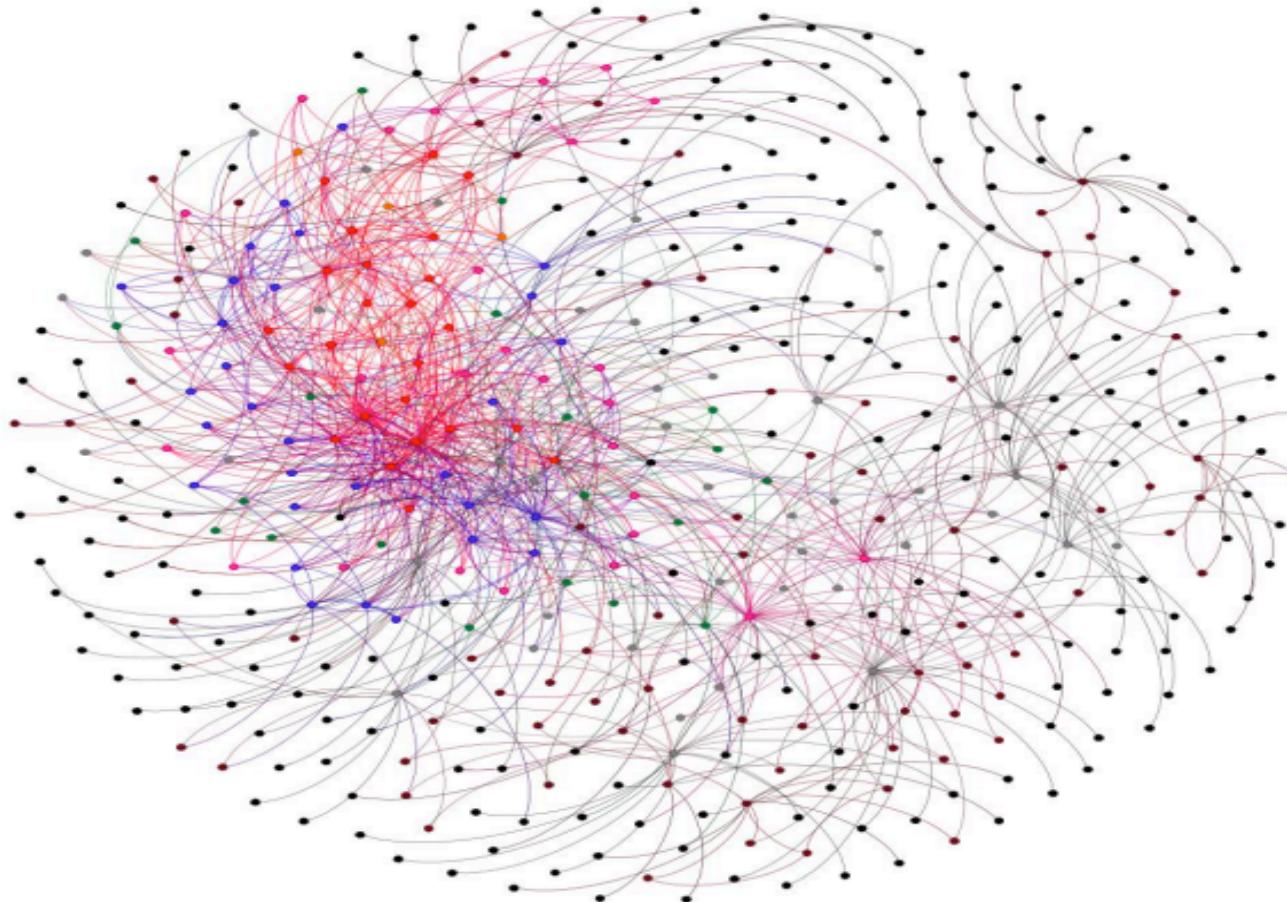
***Houve uma melhoria no PAI médio em mais de 50%**
considerando algumas aproximações realizadas na etapa de clusterização

PROBLEMA:

Detecção de Alvos Influentes



Desenvolver Técnicas para Analisar Redes Delitivas e identificar nós influentes



- $K_s = 1$
- $K_s = 2$
- $K_s = 3$
- $K_s = 4$
- $K_s = 5$
- $K_s = 6$
- $K_s = 7$
- $K_s = 8$

ANÁLISE DE REDES DELITIVAS

7

MODELOS DE DIFUSÃO

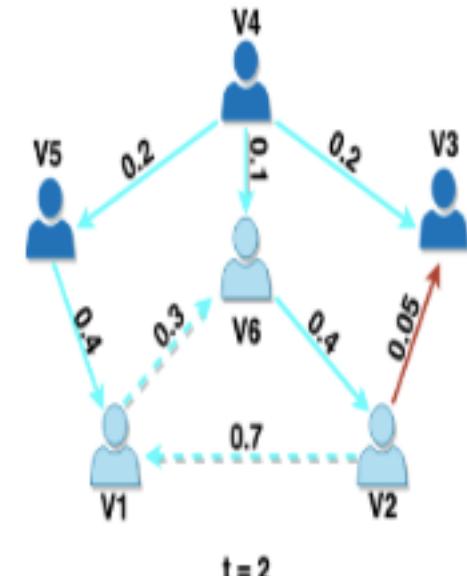
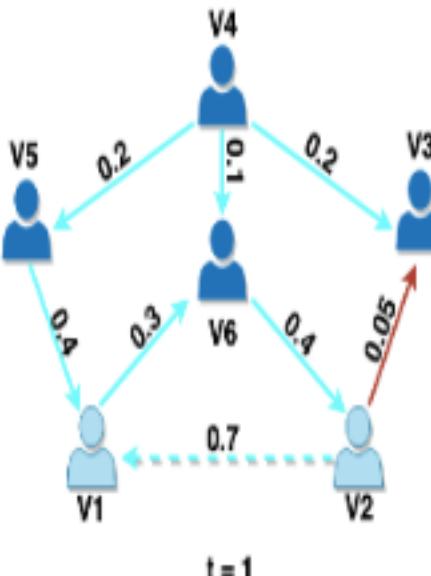
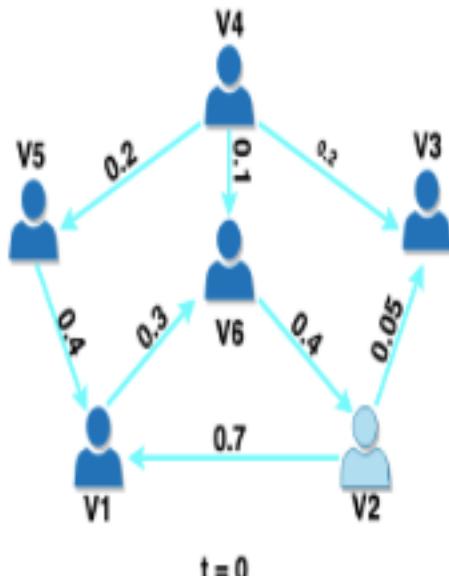
- Um modelo de difusão determina como a informação será propagada pela rede

Um modelo de difusão estocástico para um grafo $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ determina o processo randomizado de geração de conjuntos ativos \mathbf{S}_t a partir de um conjunto de *seeds* iniciais \mathbf{S}_0 para todo $t \geq 1$.

8

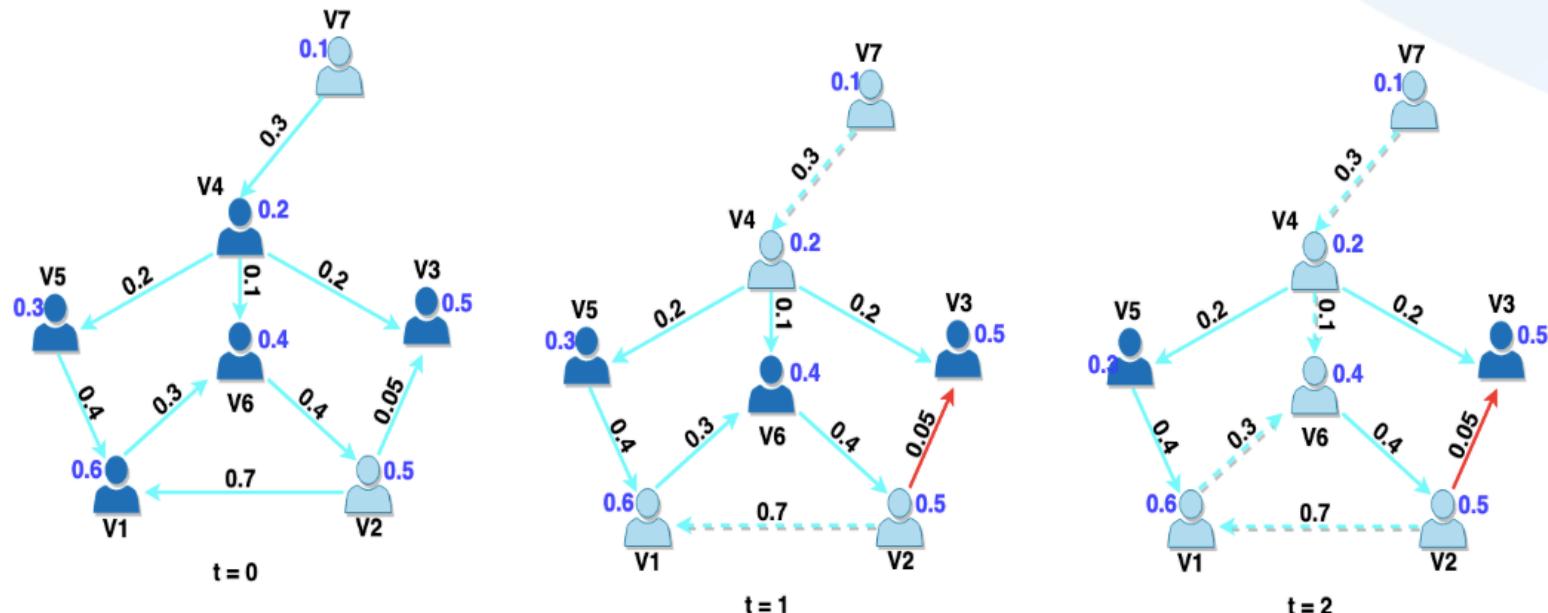
MODELOS DE DIFUSÃO

INDEPENDENT CASCADE (IC)



MODELOS DE DIFUSÃO

LINEAR THRESHOLD (LT)



INFLUENCE MAXIMIZATION

- Identificar o conjunto de *seed nodes* que **maximiza** a propagação de informação pela rede;

Dado um grafo $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, um modelo de difusão estocástico em \mathbf{G} , e um *budget* k , encontre um *seed set* $S_0 \subseteq \mathbf{V}$ com $|S_0| \leq k$, tal que a propagação de influência de S_0 , $\sigma(S_0)$, seja maximizada sob o modelo de difusão. Ou seja, computar $S^* \subseteq \mathbf{V}$ tal que:

$$S^* = \operatorname{argmax} \sigma(S_0).$$

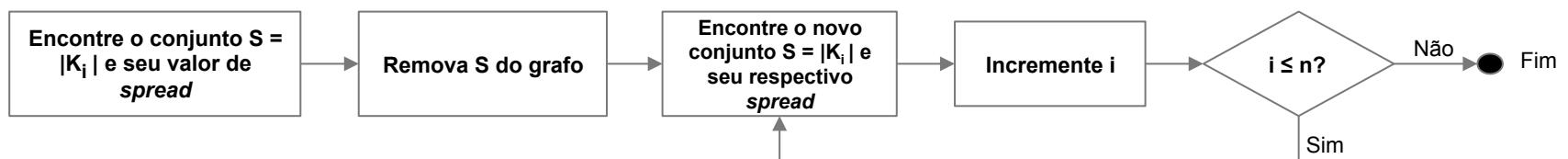
$$S_0 \subseteq \mathbf{V}, |S_0| = k$$

INFLUENCE MAXIMIZATION

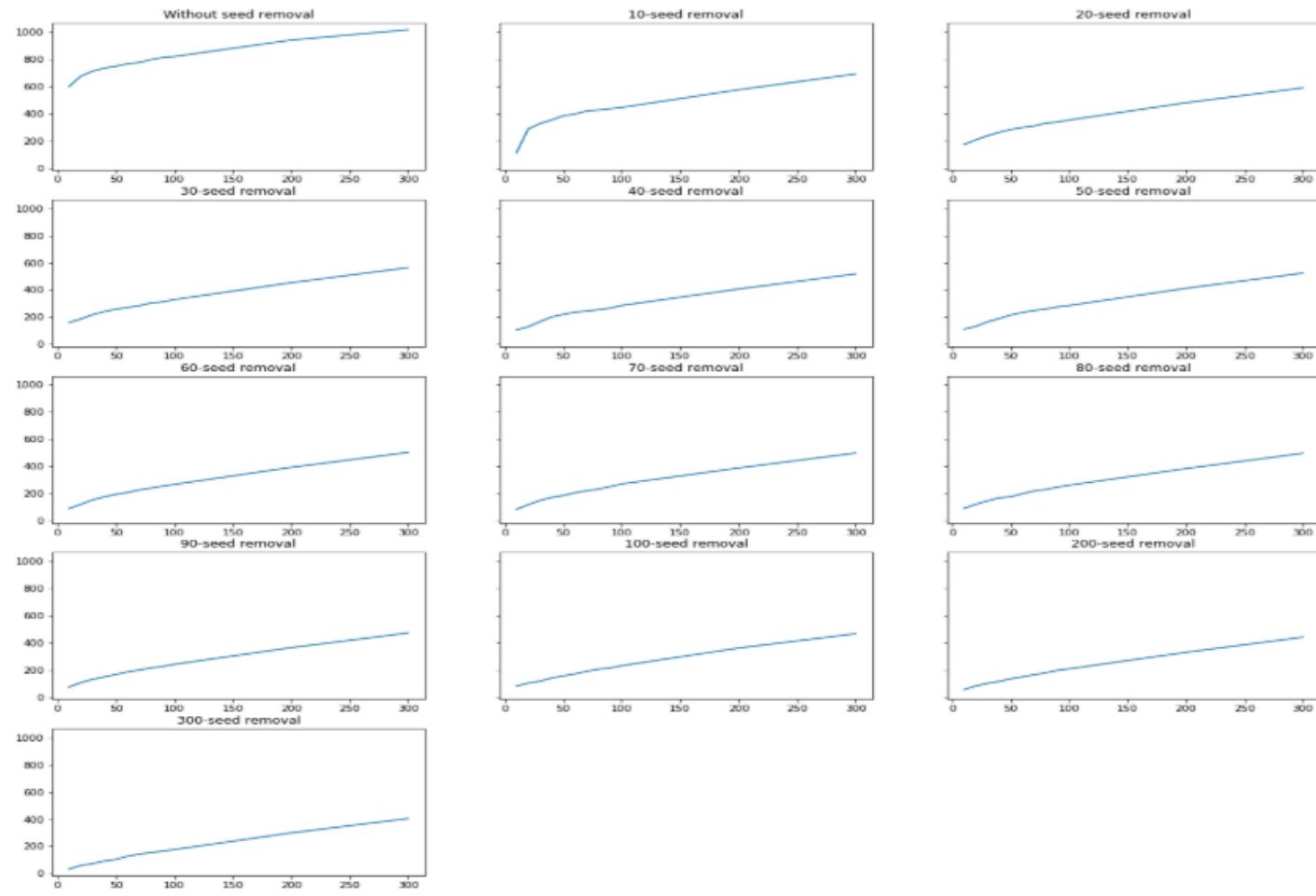
- Duas tarefas computacionais:
 - Computação da influência: calcular a propagação de influência $\sigma(S_0)$ a partir de um seed set S_0 ;
 - Encontrar o *seed set* que maximiza a influência no modelo.
- Ambas as tarefas são **NP-Difícil**

ENCONTRANDO OS K ALVOS MAIS INFLUENTES

- Intuição:
 - Remoção de conjuntos de diferentes tamanhos a fim de avaliar o impacto no valor esperado do *spreading*.
 - Seja um grafo de relações $G = (V, E)$, $K = \{K_1, K_2, \dots, K_n\}$ um conjunto de valores de *budget* utilizados para definir o conjunto de *seeds* $S \subseteq V$:



Spreading variation after seeds removal





Reconhecimento de Entidades (NER) em Textos

Human NERD

TEXT REVIEW

RESET CHANGES

SAVE

NEW TEXTS

A noticiante informa que na data e hora supracitadas o marido dela RAIMUNDO NONATO DE SOUSA retornava do trabalho conduzindo pela rodovia BR 402, altura do KM 239, no sentido Amontada à Itapipoca, a motocicleta Honda Bros, cor preta, não sabendo a placa, quando foi colhido por um caminhão até então desconhecido, o qual se evadiu do local; Que RAIMUNDO foi socorrido pelo SAMU para o hospital São Camilo de Itapipoca, vindo a óbito em virtude das lesões; Que até o momento a única informação que a noticiante tem é a de que a pessoa conhecida por WILSON teria visto o ocorrido e foi o primeiro a chegar no local; E, nada mais disse;

Remove

23/05/2019

PARA EXPEDIÇÃO DE GUIA
CADAVÉRICA; RABECÃO FORA
ACIONADO HÁ 13:15...

Remove

23/05/2019

Advertido das penas de falso
testemunho disse que era casada...

Remove

23/05/2019

REVIEWED TEXTS

Afirma que hoje, por volta das
08h30min, o seu cunhado...

Remove

23/05/2019

A noticiante informa que na data e hora supracitadas o marido dela RAIMUNDO NONATO DE SOUSA retornava do trabalho conduzindo pela rodovia BR 402, altura do KM 239, no sentido Amontada à Itapipoca, a motocicleta Honda Bros, cor preta, não sabendo a placa, quando foi colhido por um caminhão até então desconhecido, o qual se evadiu do local; Que RAIMUNDO foi socorrido pelo SAMU para o hospital São Camilo de Itapipoca, vindo a óbito em virtude das lesões; Que até o momento a única informação que a noticiante tem é a de que a pessoa conhecida por WILSON teria visto o ocorrido e foi o primeiro a chegar no local; E, nada mais disse;

LOC MISC PER ORG ADF AB HOM FEM INTER_POL ACHAD_CAD UNID_PRISIONAL ACID_TRANSITO LEG_DEFESA MORADOR_RUA DUPLO_HOM
LATROC LEG_DEF_TER MORTE_HOSP MORREU_DEPOIS EXECUCAO CONFLITO_PM CONFLITO_PC CONFLITO_BM FOLGA TRIPLO_HOM MORTE_DEPOIS
DUBLO_HOM ACHADO_CADÁVER SAPURIL PE LES_COPORAL ADF LES_CORPORAL HOM_DOLOSO

6 CLASSIFICATIONS

Selection	Class	Action
RAIMUNDO NONATO DE SOUSA	PER	<input type="button"/>
BR 402,	LOC	<input type="button"/>
KM 239,	LOC	<input type="button"/>
colhido por um caminhão	ACID_TRANSITO	<input type="button"/>
hospital São Camilo de Itapipoca, vindo a óbito em	MORTE_HOSP	<input type="button"/>
WILSON	PER	<input type="button"/>

Modelo NER

- ▷ Treinamento realizado com aproximadamente 4,000 BOs classificados manualmente (2010 a 2018);
- ▷ BOs referentes a CVLI que possuam identificação das vítimas;
- ▷ Modelo que combina redes multilayer perceptron + camadas convolucionais (Deep Learning).

CLASSES:

**PESSOA, LOCALIZAÇÃO, ADF, ACID_TRANSITO, EXECUCAO,
LATROCINIO, ...**

TEXTO 1:

Afirma que hoje, por volta das 08h30min, o seu cunhado, **FRANCISCO TIAGO MOTA DOS SANTOS**, se encontrava comprando peças de motocicleta na Oficina do CIÊ, localizada na **Avenida Perimetral Sul, centro de Nova Olinda/Ce**, momento em que foi surpreendido por disparos de **arma de fogo**, efetuados pelo garupeiro de uma motocicleta HONDA BROZ; QUE foi almejado por mais de seis disparos, tendo morte no local do fato; QUE não sabe informar quem seriam os autores, nem o motivo da morte; HÁ dois anos a vítima tentou contra a vida de um rapaz na **cidade de Nova Olinda**; QUE não sabe informar o nome deste; QUE após o crime se apresentou com advogado e ficou respondendo em liberdade. E NADA MAIS DISSSE

```
1 text = '''Afirma que hoje, por volta das 08h30min, o seu cunhado, FRANCISCO TIAGO MOTA DOS SANTOS,  
2 se encontrava comprando peças de motocicleta na Oficina do CIÊ, localizada na Avenida Perimetral Sul,  
3 centro de Nova Olinda/Ce, momento em que foi surpreendido por disparos de arma de fogo, efetuados pelo  
4 garupeiro de uma motocicleta HONDA BROZ; QUE foi almejado por mais de seis disparos, tendo morte no local  
5 do fato; QUE não sabe informar quem seriam os autores, nem o motivo da morte; HÁ dois anos a vítima  
6 tentou contra a vida de um rapaz na cidade de Nova Olinda; QUE não sabe informar o nome deste; QUE  
7 após o crime se apresentou com advogado e ficou respondendo em liberdade. E NADA MAIS DISSE'''  
8 parsed_cpi = model(text)  
9 for num, entity in enumerate(parsed_cpi.ents):  
10     print('Entity {}:{}{}'.format(num + 1), entity, '-', entity.label_)

Entity 1: FRANCISCO TIAGO MOTA DOS SANTOS - PER
Entity 2: Avenida Perimetral Sul - LOC
Entity 3: centro de Nova Olinda/Ce - LOC
Entity 4: arma de fogo - ADF
Entity 5: cidade de Nova Olinda - LOC
```

CLASSES:

**PESSOA, LOCALIZAÇÃO, ADF, ACID_TRANSITO, MORTE_HOSP,
EXECUCAO, LATROCINIO, ...**

TEXTO 2:

A noticiante informa que na data e hora supracitadas o marido dela **RAIMUNDO NONATO DE**

SOUSA retornava do trabalho conduzindo pela rodovia **BR 402**, altura do **KM 239**, no sentido Amontada à Itapipoca, a motocicleta Honda Bros, cor preta, não sabendo a placa, quando foi **colhido por um caminhão** até então desconhecido, o qual se evadiu do local; Que RAIMUNDO foi socorrido pelo SAMU para o **hospital São Camilo de Itapipoca, vindo a óbito** em virtude das lesões; Que até o momento a única informação que a noticiante tem é a de que a pessoa conhecida por **WILSON** teria visto o ocorrido e foi o primeiro a chegar no local; E, nada mais disse;

```
1 text = '''A noticiante informa que na data e hora supracitadas o marido dela RAIMUNDO NONATO DE  
2 SOUSA retornava do trabalho conduzindo pela rodovia BR 402, altura do KM 239, no sentido Amontada à  
3 Itapipoca, a motocicleta Honda Bros, cor preta, não sabendo a placa, quando foi colhido por um caminhão  
4 até então desconhecido, o qual se evadiu do local; Que RAIMUNDO foi socorrido pelo SAMU para o hospital  
5 São Camilo de Itapipoca, vindo a óbito em virtude das lesões; Que até o momento a única informação  
6 que a noticiante tem é a de que a pessoa conhecida por WILSON teria visto o ocorrido e foi o primeiro  
7 a chegar no local; E, nada mais disse'''  
8 parsed_cpi = model(text)  
9 for num, entity in enumerate(parsed_cpi.ents):  
10     print('Entity {}:{}{}'.format(num + 1), entity, '-', entity.label_)
```

```
Entity 1: RAIMUNDO NONATO DE SOUSA - PER  
Entity 2: BR 402 - LOC  
Entity 3: KM 239 - LOC  
Entity 4: colhido por um caminhão - ACID_TRANSITO  
Entity 5: hospital São Camilo de Itapipoca, vindo a óbito - MORTE_HOSP  
Entity 6: WILSON - PER
```

CLASSES:

PESSOA, LOCALIZAÇÃO, ADF, ACID_TRANSITO, MORTE_HOSP,
EXECUCAO, LATROCINIO, ...

TEXTO 4:

AFIRMA O DECLARANTE QUE É FILHO DE **RAIMUNDO NONATO ARRUDA MARTINS** O QUAL FOI VÍTIMA DE **ASSALTO**, TENDO SIDO ATINGIDO COM TRÊS DISPAROS DE **ARMA DE FOGO**; QUE SEU GENITOR FOI SOCORRIDO PARA O **FRROTINHA DA PARANGABA ONDE VEIO A ÓBITO.**

```
1 text = '''AFIRMA O DECLARANTE QUE É FILHO DE RAIMUNDO NONATO ARRUDA MARTINS O QUAL FOI VÍTIMA DE ASSALTO,  
2 TENDO SIDO ATINGIDO COM TRÊS DISPAROS DE ARMA DE FOGO; QUE SEU GENITOR FOI SOCORRIDO PARA O FRROTINHA DA  
3 PARANGABA ONDE VEIO A ÓBITO.'''  
4 parsed_cpi = model(text)  
5 for num, entity in enumerate(parsed_cpi.ents):  
6     print('Entity {}:{}{}'.format(num + 1), entity, '-', entity.label_)
```

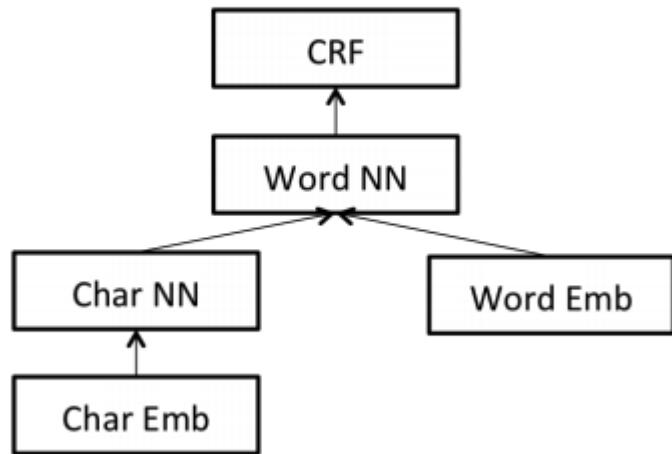
```
Entity 1: RAIMUNDO NONATO ARRUDA MARTINS - PER  
Entity 2: ASSALTO - LATROC  
Entity 3: ARMA DE FOGO - ADF  
Entity 4: FRROTINHA DA PARANGABA ONDE VEIO A ÓBITO - MORTE_HOSP
```

VÍTIMA VERSUS QUALQUER PESSOA (QUE NÃO VÍTIMA)

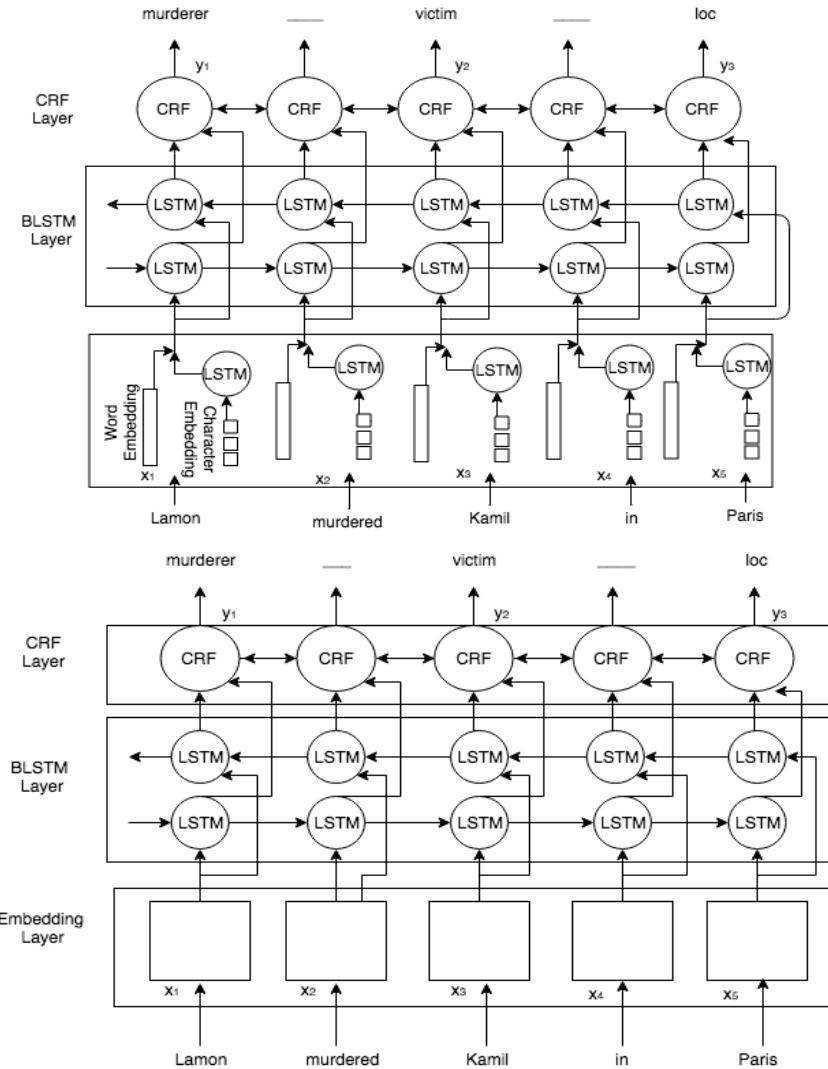
“Afirma o noticiante que é policial militar; QUE, hoje, às 13:00 h, encontrava-se de serviço com os policiais militares SGT DE QUEIROS e SD DE MESQUITA+, quando foram acionados para uma ocorrência de homicídio no local acima informado; QUE, quando chegaram ao local já tinha outra viatura no local do crime, a do Sub Tenente LAERCIO; QUE, a **vítima** já se encontrava sem vida com perfurações de arma de fogo e foi identificada como sendo **JOSE ERNILDO PINHEIRO MARINHO**; QUE, na casa da **vítima** foi apreendido uma pasta pequena com dois cartões do bolsa-família de **FRANCISCA CELIANA SILVA MARQUES** e **LUCINETE MONTEIRO LIMA**, um cartão do Banco Bradesco de **LALESCA E BARBOSA MENDES** com cartão-chave e mais dois cartões em nome de **RENE LUIZ HOLANDA OLIVEIRA**, sendo um da Caixa Econômica Federal e um cartão cidadão; QUE, uma sobrinha da **vítima** de nome **LÍVIA NARA NUNES LIRA**, residente à Rua Santa Edwirgens, n ° 602, Mutirão, estava dentro da casa onde a vítima foi morta.”

Modelos Experimentados

- Modelos Deep Learning
- BILSTM-CRF
- Char-BILSTM-CRF



(a) Base model: both of Char NN and Word NN can be implemented as CNNs or RNNs.



Experimentos

- ▷ Treinamento realizado com 1,399 textos e teste com 350 textos classificados manualmente
- ▷ MLDM'19

Network	hidden units	VICTIM	PERSON	OTHERS
BRNN-CRF	8	0,57	0,42	0,99
	16	0,60	0	0,99
	32	0,59	0	0,99
	64	0,60	0	0,99
	128	0,62	0,19	0,99
BLSTM-CRF	8	0,71	0,57	0,99
	16	0,71	0,57	0,99
	32	0,79	0,60	0,99
	64	0,79	0,60	0,99
	128	0,77	0,61	0,99
Char-BLSTM-CRF	8	0,80	0,56	0,99
	16	0,67	0,63	0,99
	32	0,67	0,47	0,99
	64	0,85	0,65	0,99
	128	0,76	0,49	0,99

Table 2. Precision results for VICTIM, PERSON and OTHERS classes.

(Hyper-) parameter	Range	Chosen Value
dropout-rate	[0,1 - 0,5], step 0,2	0,5
number of epochs	[30 - 50], step 10	30
batch size	[16 - 128], step ×2	32
hidden units	[8 - 128], step ×2	[8 - 128], step ×2
optimizer	[sgd, rmsprop, adam]	rmsprop

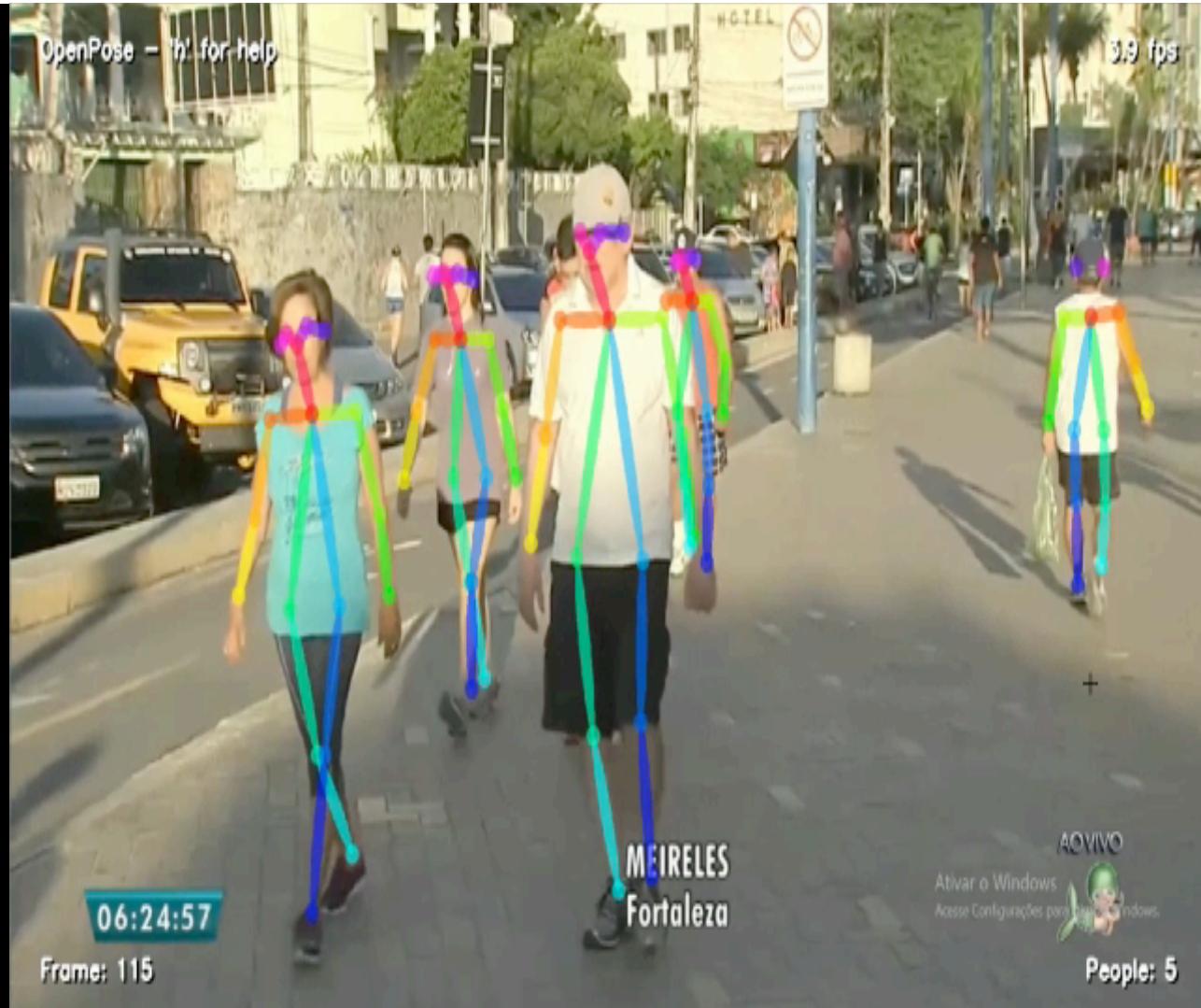
Table 1. (Hyper-) parameters for all experiments.

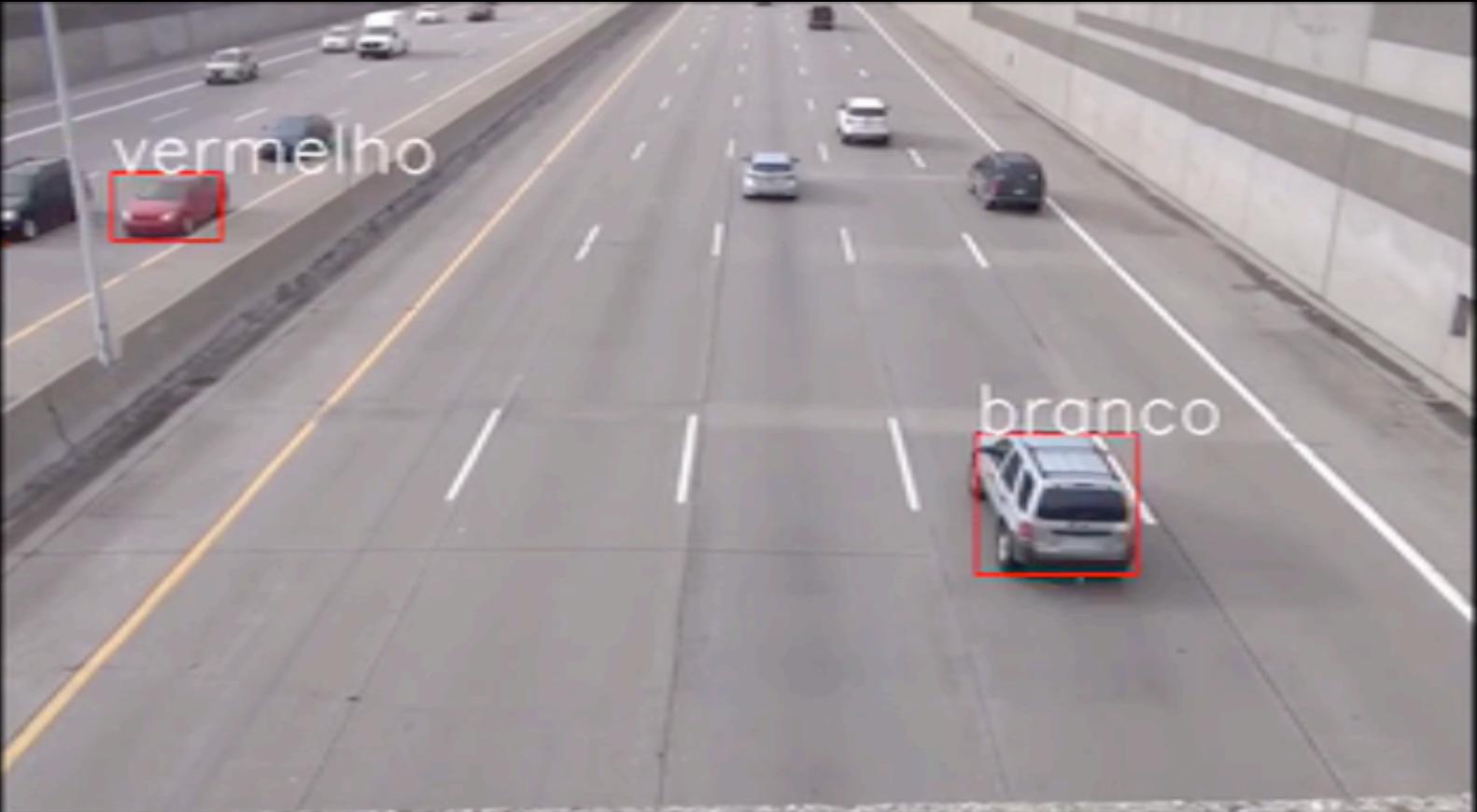
Network	hidden units	VICTIM	PERSON	OTHERS
BRNN-CRF	8	0,86	0,01	0,99
	16	0,80	0	0,99
	32	0,81	0	0,99
	64	0,80	0	0,99
	128	0,70	0,02	0,98
BLSTM-CRF	8	0,82	0,43	0,99
	16	0,81	0,33	0,99
	32	0,76	0,47	0,99
	64	0,79	0,61	0,99
	128	0,77	0,45	0,99
Char-BLSTM-CRF	8	0,80	0,58	0,99
	16	0,83	0,30	0,99
	32	0,78	0,37	0,99
	64	0,73	0,64	0,99
	128	0,82	0,54	0,99

Table 3. Recall results for VICTIM, PERSON and OTHERS classes.



Outras pesquisas





vermelho

branco



INSIGHT

Data Science Laboratory
Federal University of Ceará





Muito
Obrigado!

Contato

 jose.macedo@dc.ufc.br