

Predição da velocidade média em vias de Recife monitoradas por sensores

Thiago Machado de Azevedo



Quem sou eu?



- **Desenvolvedor da Thoughtworks**
- **Entusiasta em Machine Learning e Inteligência Artificial**

Motivação



Qual foi a motivação?

Estudos trazem que pessoas em média perdem 1 mês e meio por ano no trânsito na cidade de São Paulo. Em Recife, esse valor é de 27 dias no ano.



Perda de tempo no trânsito



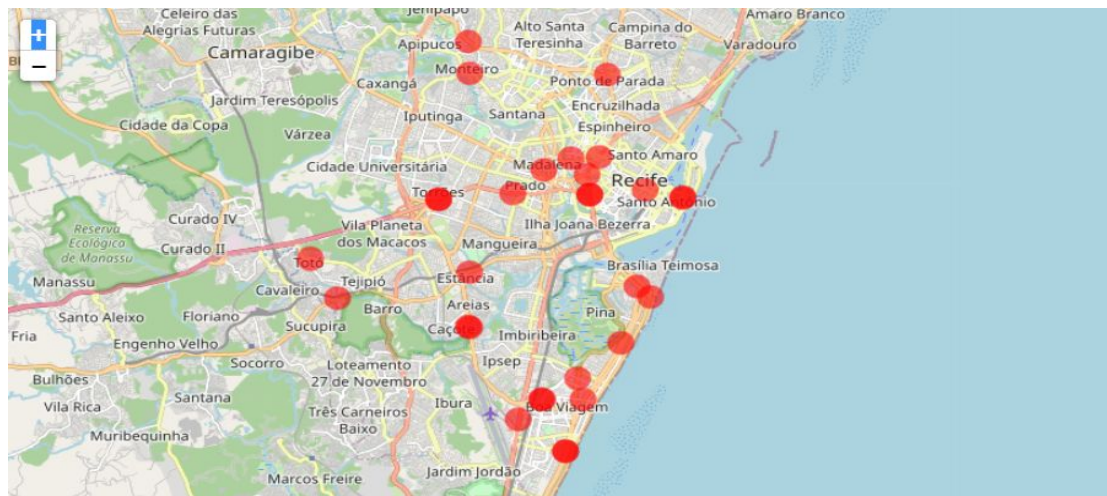
Previsão de tráfego com base na Velocidade de carros



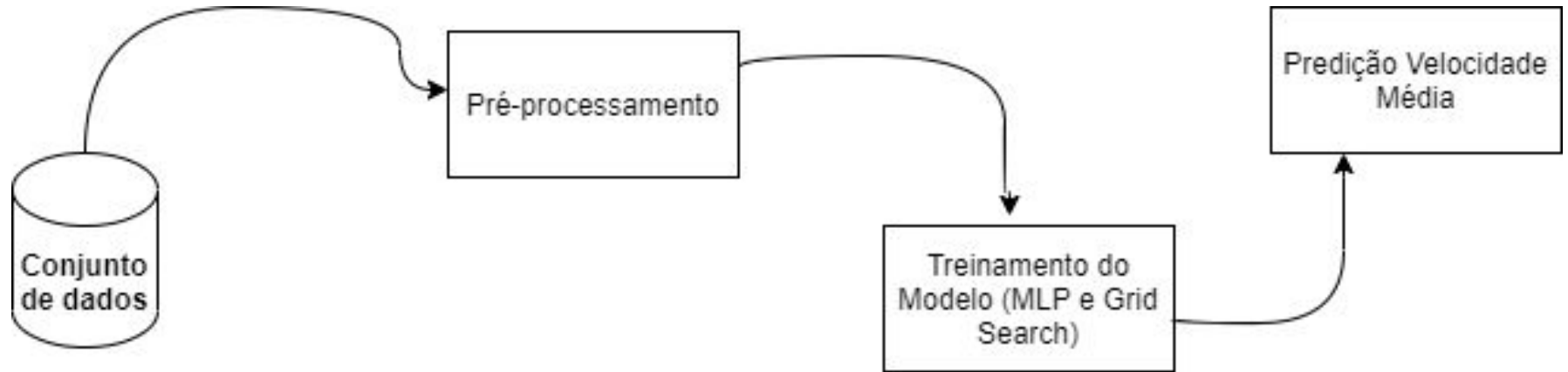
Sensores de trânsito

Objetivo

Desenvolver um modelo preditivo a partir dos dados abertos da cidade de Recife, baseado em Redes Neurais utilizando a velocidade média dos carros que passaram pelos sensores.



Predição da velocidade média nas vias de Recife



Dados



Base de dados

The screenshot shows the 'Dados Recife' website interface. At the top, there is a navigation bar with the city logo and menu items: 'A PREFEITURA', 'SERVIÇOS', 'SECRETARIA E ÓRGÃOS', and 'CONHEÇA O RECIFE'. Below this is a secondary navigation bar with 'Conjuntos de dados', 'Organizações', 'Grupos', 'Visualizações', 'Consultas Livres', 'Contato', and 'Sobre'. A search bar is located on the right side of this bar.

The main content area displays a data set titled 'Velocidade das Vias - Quantitativo por Velocidade Média - 2018'. On the left side of this data set, there is a sidebar with the following information:

- Seguidores: 0
- Organização: CTTU (Autarquia de Trânsito e Transporte Urbano do Recife)

The main content area includes a description of the data set: 'Este conjunto de dados apresenta os quantitativos de veículos por velocidade média no intervalo de 15 minutos. Os dados apresentados descrevem quantos veículos por equipamento passaram dentro de cada faixa de intervalo de tempo e de velocidade. Ex: Entre 08:00 e 08:15, no equipamento 022, passaram entre a velocidade de 0 - 10km/h um total de 60 veículos, entre 10 km/h e 20km/h um total de 67 veículos...'. It also provides a link to the data location: 'http://dados.recife.pe.gov.br/dataset/equipamentos-de-monitoramento-e-fiscalizacao-de-transito'.

Below the description, there is a list of datasets for the years 2016 through 2020, each with a corresponding URL. At the bottom of the data set page, there is a section titled 'Dados e recursos'.

Base de dados
abertos de Recife

9 Milhões de dados

Sensores de coleta
de velocidade

hora	minuto	dia	mes	ano	diaDaSemana	rua	bairro	longitude	latitude	velocidade_via	velocidadeMedia	numeroVeiculos	equipamento
0	15	01	01	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-34.873669	-8.0634	30 Km/h	15.500000	2	FS002
0	30	01	01	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-34.873669	-8.0634	30 Km/h	15.500000	1	FS002
0	45	01	01	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-34.873669	-8.0634	30 Km/h	15.400000	5	FS002
0	59	01	01	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-34.873669	-8.0634	30 Km/h	10.250000	2	FS002
1	75	01	01	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-34.873669	-8.0634	30 Km/h	15.500000	4	FS002

Pré-processamento dos dados

Passo 1

mes	equipamento	faixa	data	hora	minutos_intervalo	qtd_0a10km	qtd_11a20km	...	qtd_91a100km	qtd_acimade100km
1	FS002REC	1	2016-01-01	0	0-15	0	2	...	0	0
1	FS002REC	1	2016-01-01	0	16-30	0	1	...	0	0
1	FS002REC	1	2016-01-01	0	31-45	1	3	...	0	0

Passo 2

mes	equipamento	faixa	hora	minutos_intervalo	velocidade_via	velocidadeMedia	numeroVeiculos	mes	ano	diaDaSemana
1	FS002REC	1	0	0-15	30 Km/h	15.5	2	1	2016	Friday
1	FS002REC	1	0	16-30	30 Km/h	15.5	1	1	2016	Friday
1	FS002REC	1	0	31-45	30 Km/h	15.4	5	1	2016	Friday

Passo 3

hora	minuto	dia	mes	ano	diaDaSemana	rua	bairro	longitude	latitude	velocidade_via	velocidadeMedia	numeroVeiculos	equipamento
0	15	1	1	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-348.736.694	-80.634.002	30 Km/h	15.5	2	FS002
0	30	1	1	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-348.736.694	-80.634.002	30 Km/h	15.5	1	FS002
0	45	1	1	2016	Friday	RUA MADRE DE DEUS, SEMAFORO 020.	Bairro do Recife	-348.736.694	-80.634.002	30 Km/h	15.4	5	FS002

Passo 4

seg	ter	quar	qui	sex	sab	dom	velocidadeMedia	hora	mes	isJan_isJuly
0	0	0	0	1	0	0	15.5	0	1	1
0	0	0	0	1	0	0	15.5	0	1	1
0	0	0	0	1	0	0	15.4	0	1	1

$$velocidadeMedia = \sum_{i=0}^{qtdCarros} \frac{Intervalo \times \left(\frac{intervaloMaior + intervaloMenor}{2} \right)}{quantidadeTotalCarros}$$

Normalização dos dados

$$velocidadeMediaNormalizada = \left(\frac{veloMediaValorColetado - minVeloMedia}{maxVeloMedia - minVeloMedia} \right)$$

$$horaNormalizada = \left(\frac{horaValorColetado - minHora}{maxHora - minHora} \right)$$

- Dias da Semana (Seg, Ter, Quar, Qui, Sex, Sab, Dom)
- Janeiro ou Julho
- Velocidade Média
- Hora

Passo 4

seg	ter	quar	qui	sex	sab	dom	velocidadeMedia	hora	mes	isJan_isJuly
0	0	0	0	1	0	0	15.5	0	1	1
0	0	0	0	1	0	0	15.5	0	1	1
0	0	0	0	1	0	0	15.4	0	1	1

Redes Neurais



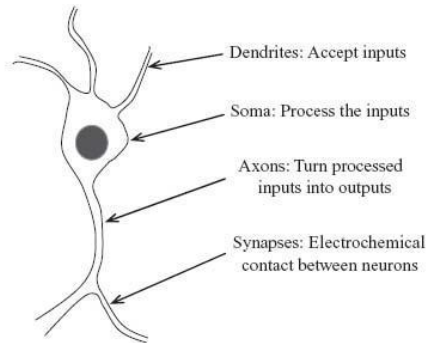
Aprendizagem de máquina aplicado a previsão de tráfego

- Sensores
 - Magnéticos
 - Micro-ondas
 - Infravermelhos
 - Áudio
- Técnicas de Inteligência Artificial
 - Séries Temporais
 - Filtro de Kalman
 - Redes Neurais

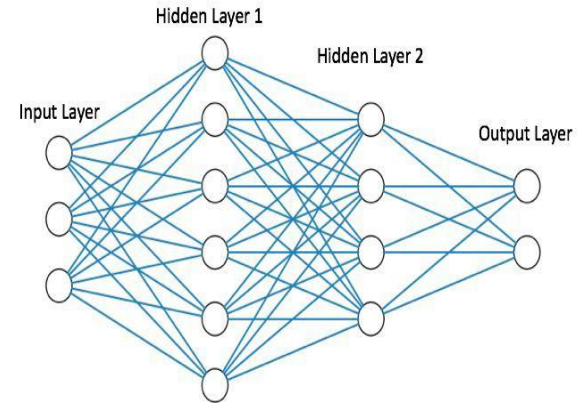
Sensor technology	Vehicle classification	Multiple detection zone
Inductive loop		
Magnetic sensor		
Video image processor	✓	✓
Microwave radar	✓	✓
Laser radar sensor	✓	✓
Active infrared	✓	✓
Passive infrared		
Audio sensor	✓	

Redes neurais Artificiais

Neurônio



Multilayer Perceptron (MLP)



Avaliação de modelos preditivos

- Métrica de Erro
 - Erro Médio Absoluto (MAE)
- Validação Cruzada K-folds (Grid-Search)

$$MAE = \frac{\sum_{i=0}^n |Y_i - X_i|}{n}$$

Treinamento do Modelo



Treinamento do Modelo

Estratégias para predição da velocidade

- Gerar um modelo de todos os dados
- Gerar modelos por sensores

- *Inputs* da Rede Neural
 - Dias Semana (Seg, Ter, Quar, Qui, Sex, Sab, Dom)
 - Hora
 - Mês
 - Janeiro ou julho (*isJan_isJuly*)
- *Output* da Rede Neural
 - Velocidade média

Treinamento do Modelo

- MLP
 - Neurônios escondidos: 5, 10, 20
 - Taxa de Aprendizado: 0.1 , 0.05
 - Número máximo de iterações: 1000, 10000
 - Algoritmo de retropropagação do erro
- Validação Cruzada
 - 10-Folds

```
param_grid = {
    'hidden_layer_sizes':[(5,),(10,),(20,)], # Feito # OK
    'activation':['logistic'], #Feito #OK
    'solver':['sgd'], # OK
    'momentum':[0.5],
    'learning_rate_init':[0.1,0.05], # Feito # OK
    'max_iter': [1000,10000], #Feito #OK
    'learning_rate':['constant'] #Valor default #Feito
}
```

```
print("Parte do treinamento")
df_x = np.asarray(equipamento[['seg','ter','quar','qui','sex','sab','dom','hora','isJan_isJuly']], dtype='float64')
df_y = np.asarray(equipamento['velocidadeMedia'], dtype='float64')
df_x = np.nan_to_num(df_x)
df_y = np.nan_to_num(df_y)
df_x = df_x.astype(np.float64)
df_y = df_y.astype(np.float64)
x_train, x_test, y_train, y_test = train_test_split(df_x,df_y, test_size=0.30)
nn = MLPRegressor()
gridSearch = GridSearchCV(nn,param_grid,cv=10,refit=False,scoring=['neg_mean_squared_error','neg_mean_squared_log_error',
'neg_median_absolute_error','neg_mean_absolute_error','neg_root_mean_squared_error'])
print("Fazendo o FIT com o DataFrame")
gridSearch = gridSearch.fit(x_train,y_train)
```

Resultados e Conclusão



Treinamento com todos os Equipamentos

Neurônios Escondidos	Taxa de Aprendizado	Número de Iterações	MAE
5	0.1	1000	5,736
5	0.1	10000	17,555
5	0.05	1000	5,722
5	0.05	10000	16,920
10	0.1	1000	5,781
10	0.1	10000	58,047
10	0.05	1000	57,048
10	0.05	10000	5,728
20	0.1	1000	17,973
20	0.1	10000	18,200
20	0.05	1000	5,759
20	0.05	10000	5,766

Treinamento por equipamento/sensor

Radar	Neurônios Escondidos	Taxa de Aprendizado	Número de Iterações	MAE
FS002	20	0,05	1000	4,821
FS003	5	0,05	10000	30,612
FS004	10	0,01	1000	5,026
FS005	5	0,05	10000	7,264
FS006	5	0,05	1000	5,548
FS007	20	0,05	1000	6,569
FS008	20	0,1	1000	8,745
FS009	5	0,05	1000	4,027
FS010	20	0,05	1000	6,357
FS011	5	0,1	1000	5,193
FS012	20	0,05	1000	5,176
FS013	5	0,05	10000	6,669
FS014	5	0,05	10000	5,366
FS015	5	0,05	1000	4,761
FS016	20	0,05	10000	8,564
FS017	5	0,1	1000	7,726
FS018	10	0,1	10000	7,523
FS019	5	0,1	1000	13,248
FS020	20	0,1	10000	4,510
FS021	5	0,1	10000	58,330
FS022	5	0,1	1000	7,243
FS023	5	0,05	1000	6,732
FS024	5	0,05	1000	7,512
FS025	5	0,05	10000	5,567
FS026	5	0,1	1000	41,410
FS027	5	0,05	1000	53,510
FS028	5	0,1	10000	5,802
FS029	5	0,1	10000	9,111
FS030	10	0,1	1000	30,548
FS033	20	0,1	10000	5,488
FS037	5	0,1	1000	4,449

Conclusão

- Predição da velocidade média a partir dos sensores
 - Média Por equipamento: MAE < 10 km/hr
- Desafios e Oportunidades
- Trabalhos Futuros

We look forward to working with you

Thiago Azevedo

Consultant Software Development

thiago.azevedo@thoughtworks.com

