



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO
LABORATÓRIO DE REDES DE COMPUTADORES

Análise de Poluentes do Ar baseada em Sensores Virtuais

Orientando: **Gabriel Oliveira Campos**
Orientador: **Prof. Dr. Leandro Aparecido Villas**

Campinas, 20 de julho de 2021

Agenda



- 1. Introdução**
- 2. Conceitos Importantes**
- 3. Trabalhos Relacionados**
- 4. Proposta**
- 5. Resultados**
- 6. Cronograma**

Agenda



1. Introdução
2. Conceitos Importantes
3. Trabalhos Relacionados
4. Proposta
5. Resultados
6. Cronograma

Introdução

Contextualização



Figura: Poluição no México



Figura: Queimadas no Brasil

Introdução

Motivação



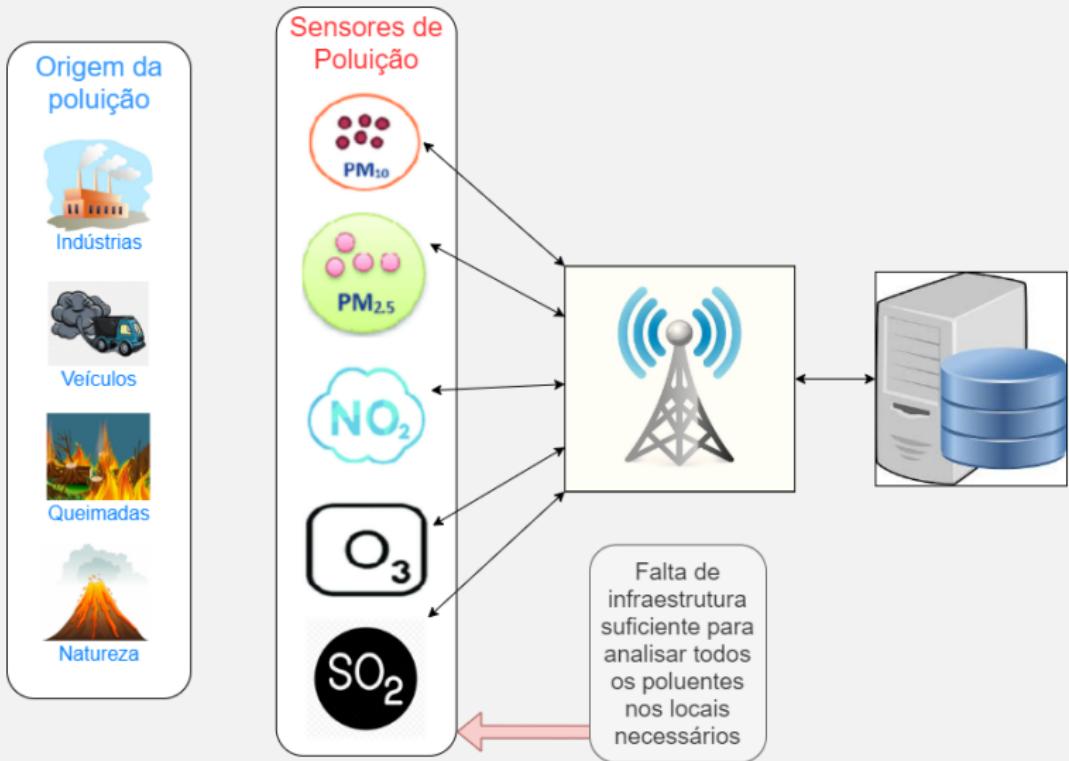
4

- Estimativa de 4.2 milhões de mortes prematuras atribuídas à poluição do ar no mundo anualmente
- Custos com doenças respiratórias ultrapassou R\$1.3 Bilhão em 2018 no Brasil
- É estimado que os gastos entre 2008 e 2019 chegaram a 14 bilhões com doenças respiratórias no Brasil
- Aproximadamente R\$ 500 milhões foram gastos no tratamento da asma em 2019 no Brasil



Introdução

Análises



Agenda



1. Introdução
- 2. Conceitos Importantes**
3. Trabalhos Relacionados
4. Proposta
5. Resultados
6. Cronograma

Conceitos Importantes

Poluentes



- Um poluente atmosférico é qualquer substância presente no ar e que, pela sua concentração, possa torná-lo impróprio ou nocivo à saúde

Poluentes	Causa
SO ₂	Vulcões Processos Industriais Queima de combustível
NO ₂	Processos Industriais Incêndios Veículos Automotores
CO	Queima de materiais ricos em petróleo
O ₃	Queima de combustíveis fósseis Volatilização de combustíveis Agricultura
MP _{2,5} MP ₁₀	Queima de combustíveis Incêndios Agricultura

Conceitos Importantes

Índice da Qualidade do Ar



US EPA Air Quality Index Standard

Qualidade	Índice	MP ₁₀ 24hrs	MP _{2.5} 24hrs	O ₃ 8hrs	CO 8hrs	NO ₂ 1hr	SO ₂ 24hrs
N1 - Boa	0 - 50	0 - 50	0 - 30	0 - 50	0 - 1.0	0 - 40	0 - 40
N2 - Moderada	51 - 100	>51 - 100	>31 - 60	>51 - 100	>1.1 - 2.0	>41 - 80	>41 - 80
N3 - Ruim	101 - 200	>101 - 250	>61 - 90	>101 - 168	>2.1 - 10	>81 - 180	>81 - 380
N4 - Muito Ruim	201 - 300	>251 - 350	>91 - 120	>169 - 208	>10 - 17	>181 - 280	>381 - 800
N5 - Péssimo	301 - 400	>351 - 430	>121 - 250	>209 - 748	>18 - 34	>281 - 400	>800 - 1800
N6 - Perigoso	>401	>430	>250	>748	>34	>400	>1800

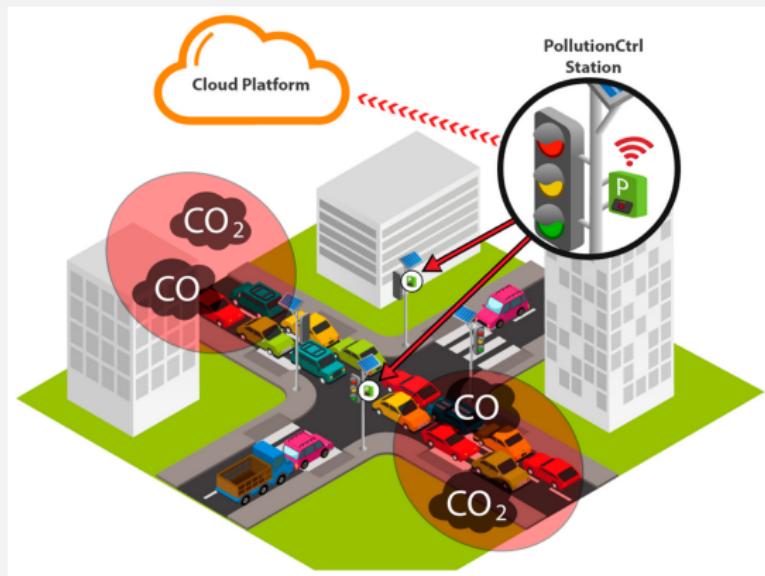
Conceitos Importantes

Sensores Físicos



9

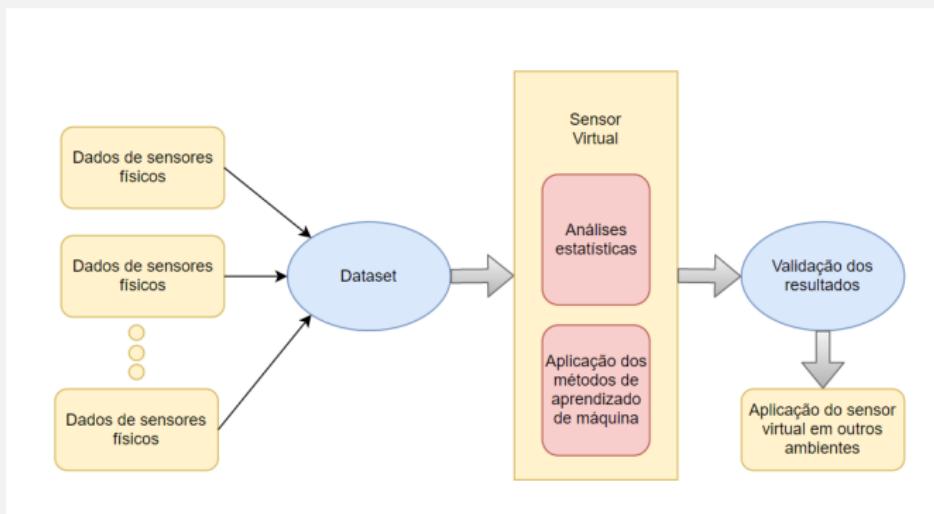
- Um sensor é um dispositivo capaz de monitorar e detectar alterações em um ambiente



Conceitos Importantes

Sensores Virtuais

- Um sensor virtual é um tipo de software que, dada a informação disponível, processa o que um sensor físico faria. Ele aprende a interpretar as relações entre as diferentes variáveis e observa as leituras dos diferentes instrumentos



Agenda



1. Introdução
2. Conceitos Importantes
- 3. Trabalhos Relacionados**
4. Proposta
5. Resultados
6. Cronograma

Trabalhos Relacionados



Artigo	Caso de utilização	Dados obtidos em
Zaidan et al. [1], 2020	<i>Low-Cost-Sensor</i> de poluição atmosférica	Helsinki
Campolina et al. [2], 2017	Sensores veiculares	Veículos e Celulares
Nguyen et al .[3], 2018	Sensores de IoT	Dispositivos IoT
Oehmcke et al. [4], 2018	Sensores marinhos	Mar norte entre as ilhas Langeoog e Spiekeroog
Modelo proposto	Sensores de poluição atmosférica	Diversas cidades

Agenda



1. Introdução
2. Conceitos Importantes
3. Trabalhos Relacionados
- 4. Proposta**
5. Resultados
6. Cronograma

■ Desafios aos se utilizar sensores físicos

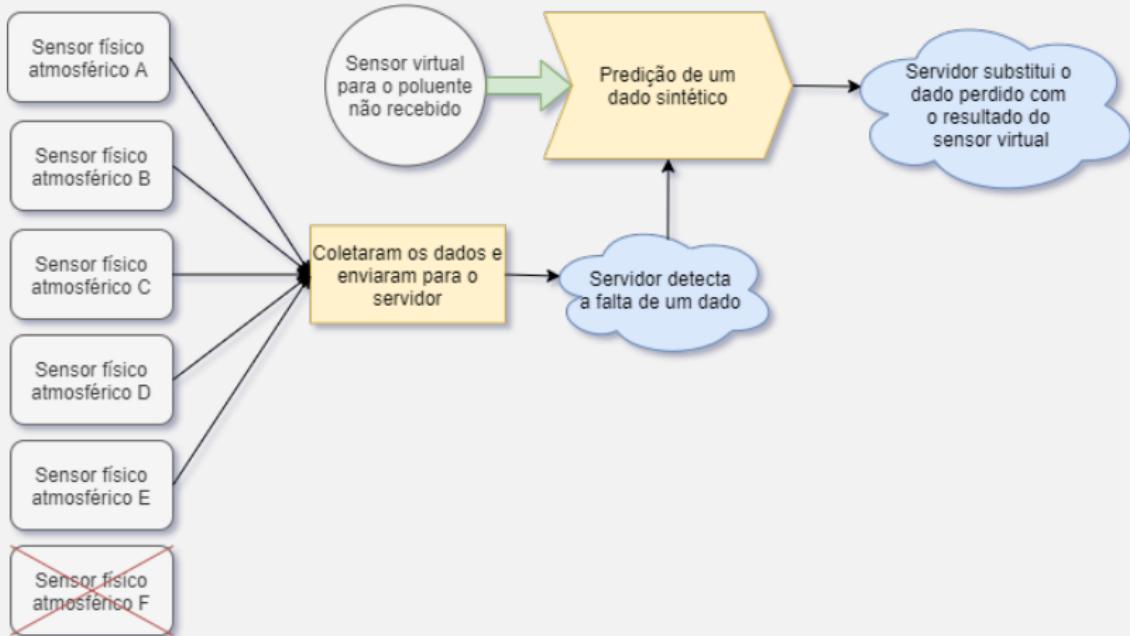
- **Falta de Infraestrutura**
- **Inviabilidade Financeira**

■ Proposta

- Elaboração de Sensores virtuais para o CO, MP₁₀ e MP_{2,5}
- Análise de modelos gerais com modelos de cidades mais específicas
- Aplicação de métodos de aprendizado de máquina e aprendizado profundo
- Análise dos melhores modelos de sensoriamento virtual
- Observação dos fatores mais impactantes nos modelos propostos

Proposta

Sensores Virtuais



Agenda



1. Introdução
2. Conceitos Importantes
3. Trabalhos Relacionados
4. Proposta
- 5. Resultados**
6. Cronograma

Resultados

Base de dados



Resultados

Pré-processamento



- Remoção dos dados faltantes
- 5 K-Fold para os métodos de aprendizado de máquina (80% para treino e 20% para teste)
- 10 buscas por hiper-parâmetros em cada fold
- Divisão de 70% para treino, 15% para validação e 15% para teste nos modelos de aprendizado profundo
- Remoção de outliers da base de dados
 - Z-Score (desvio padrão abaixo da média)
 - Interquartile Range (mantém tudo dentro do terceiro quartil e o primeiro quartil)
- Acréscimo de informações climáticas e informações sobre estações do ano na base de dados

Resultados

Quantidade de dados



Quantidade de dados de cada cidade presente na base:

	Base Completa	Base sem outliers
Amsterdam	495	380
Beijing	2227	910
Busan	1955	1429
Copenhagen	597	566
São Paulo	793	786
Las Vegas	623	559
Londres	2303	2097
Shanghai	2295	1900
Tokyo	760	753
Total	12048	9380

Resultados

Correlação dos dados



	CO	MP10	MP25
Amsterdam	-0.159	0.372	0.251
Beijing	-0.172	0.446	0.33
Busan	0.174	0.511	0.4
Copenhagen	-0.076	0.265	0.133
São Paulo	0.1827	0.538	0.397
Las Vegas	-0.048	0.331	0.189
Londres	-0.151	0.299	0.235
Shanghai	-0.012	0.317	0.35
Tokyo	0.055	0.251	0.236

Resultados

Métodos de aprendizado de máquina avaliados



- Regressão Linear
- Árvore de Regressão
- k-NN
- Vetores de Suporte para Regressão
- Vetores de Suporte para Regressão (Kernel RBF)
- Florestas Randômicas
- *Gradient Boosting Machine*
- Perceptron Multicamadas

Resultados

Métodos de aprendizado profundo avaliados



22

- *Boosted Trees*
- *Deep Learning Belief Networks*
- *Deep Neural Network (Linear Combined Regressor)*
- *Gated Recurrent Units (GRU)*
- *Long-Short Term Memory (LSTM)*
- *Linear Estimator*
- *Sequential model (4 camadas)*

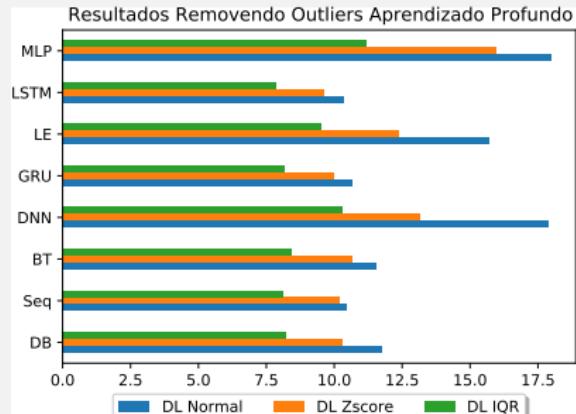
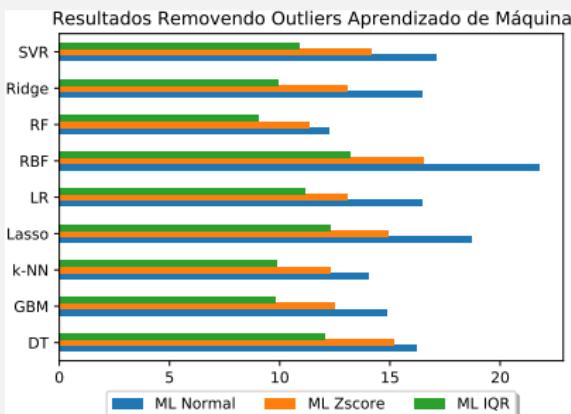
Resultados

Melhora dos resultados removendo outliers



23

Resultados utilizando os métodos de remoção de outliers



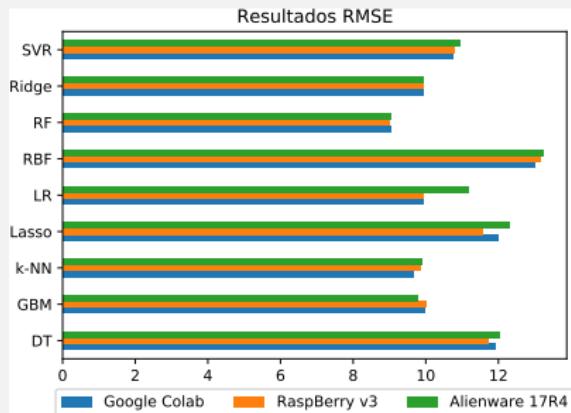
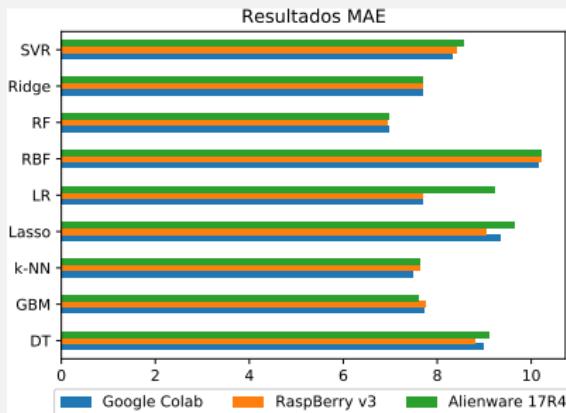
Resultados CO

Resultados Aprendizado de Máquina



24

Resultados usando MAE e RMSE:

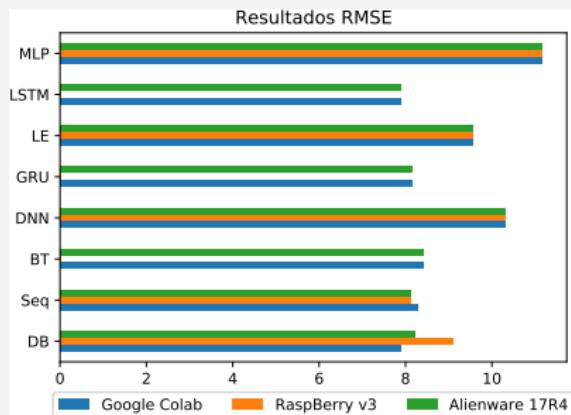
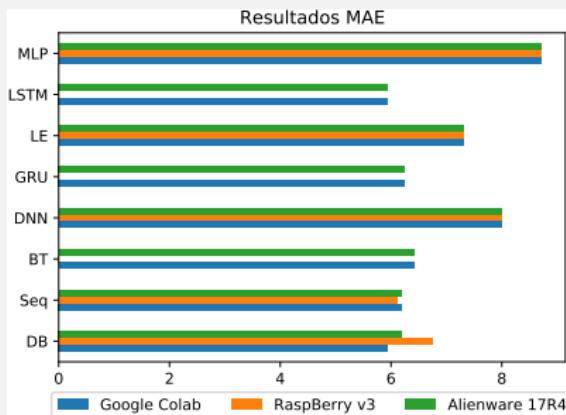


Resultados CO

Resultados Aprendizado de Profundo



Resultados usando MAE e RMSE



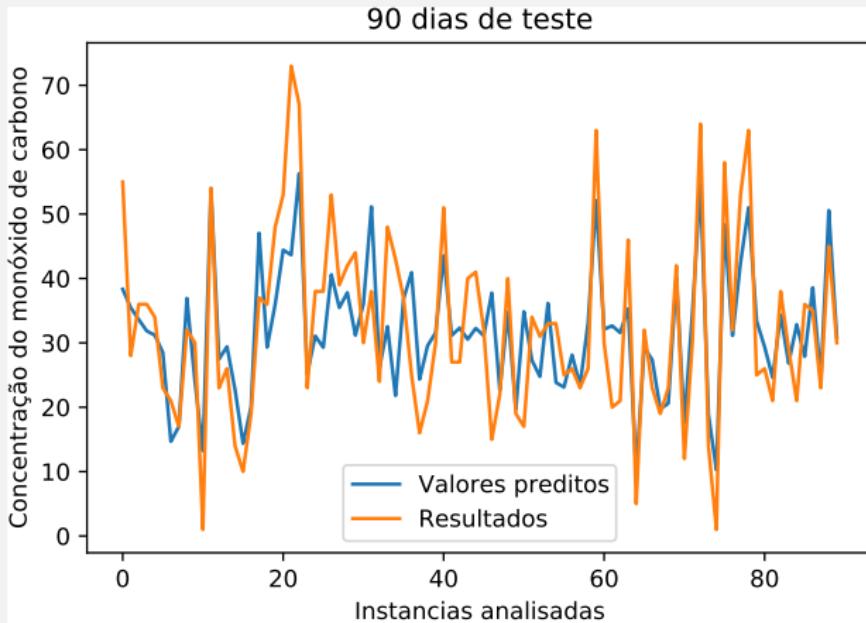
Resultados CO

Resultados do modelo generalizado



26

Últimos 90 dias do modelo generalizado utilizando
BoostedTrees



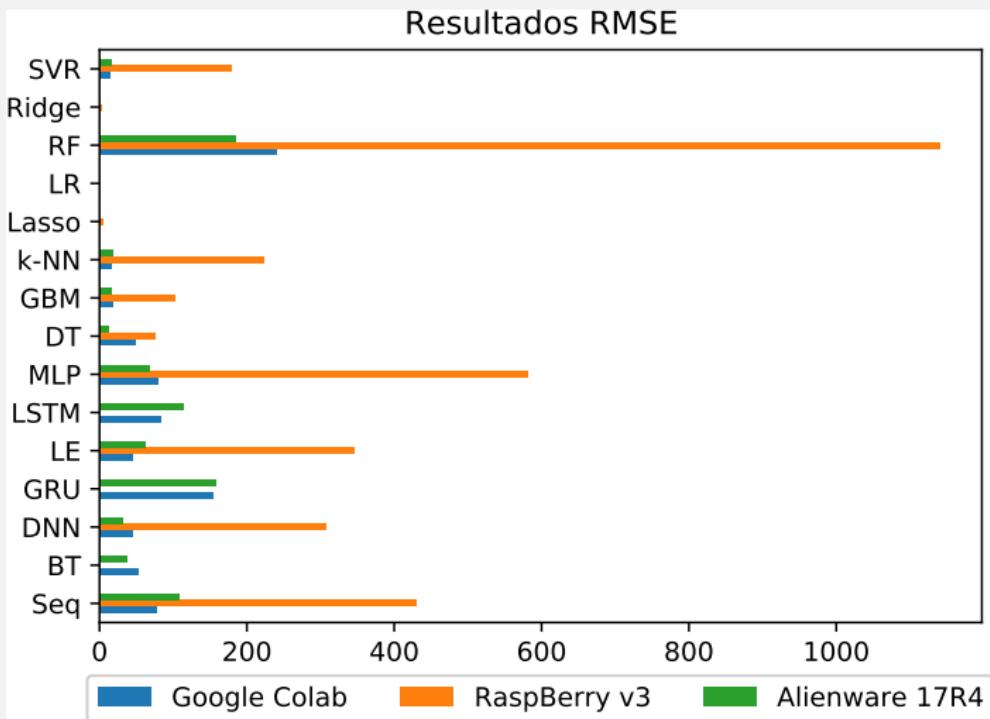
Resultados CO

Tempo de processamento



27

Tempo de processamento (em segundos):



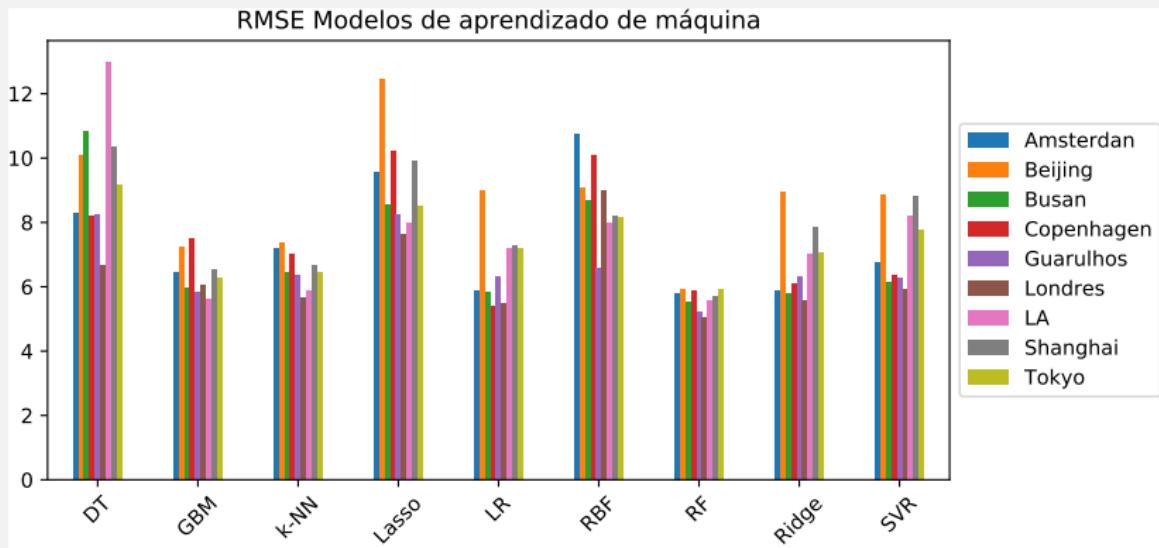
Resultados CO

Modelos de cidades específicas



28

RMSE de cada cidade em cada modelo de aprendizado de máquina analisado

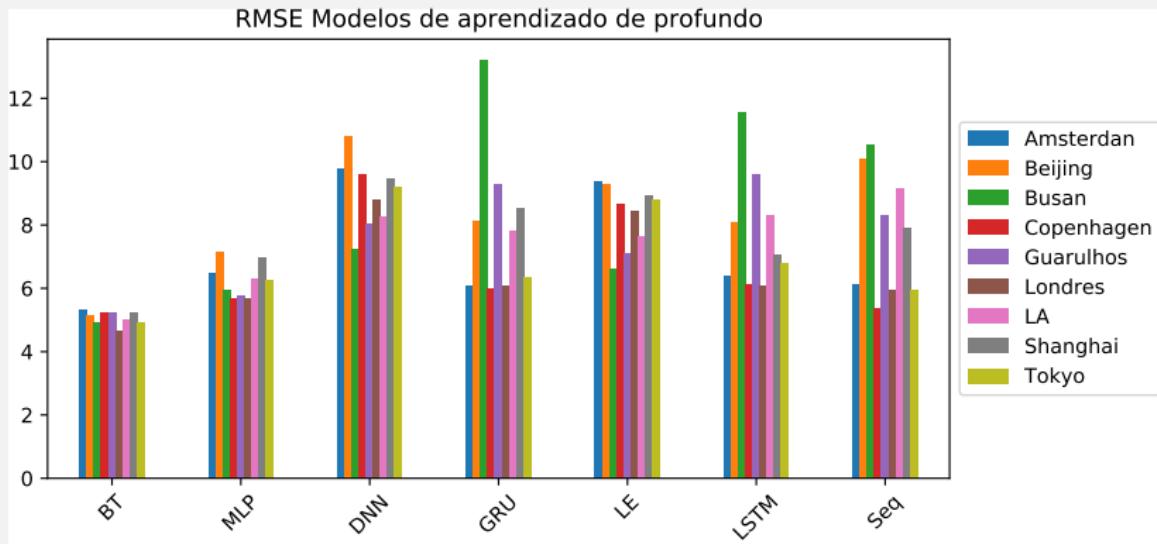


Resultados CO

Modelos de cidades específicas



RMSE de cada cidade em cada modelo de aprendizado profundo analisado

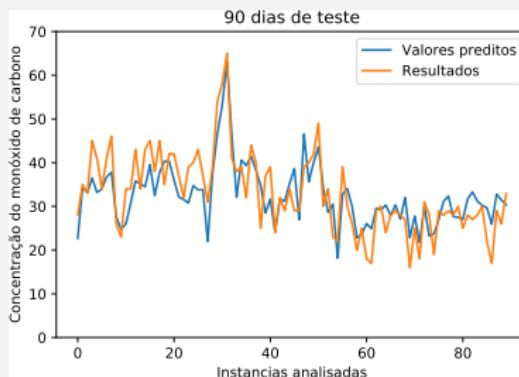
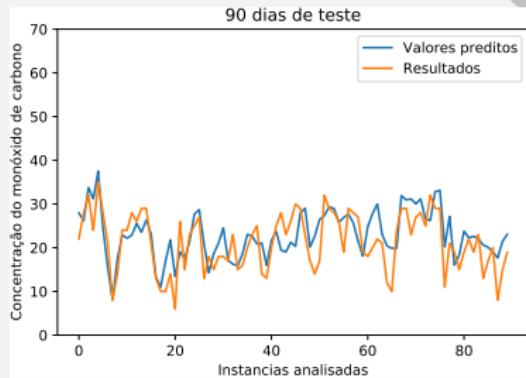
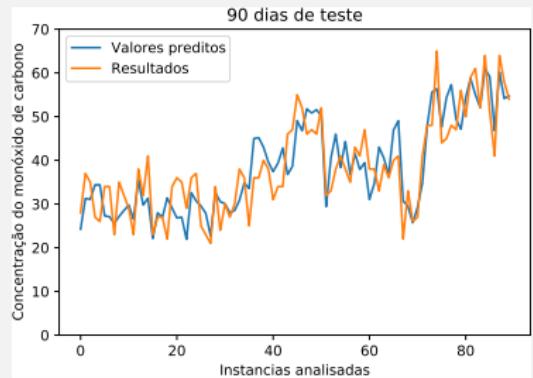


Resultados CO

Modelos reais de Busan, Londres e Tokyo

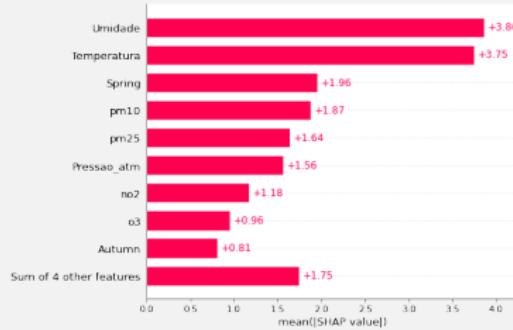
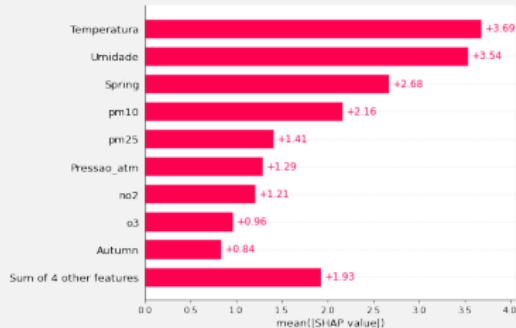
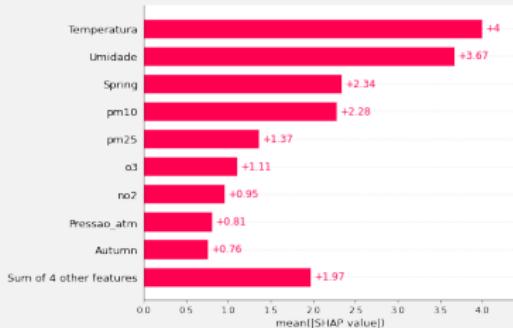


30



Shap Values CO

Modelos reais de Busan, Londres e Tokyo



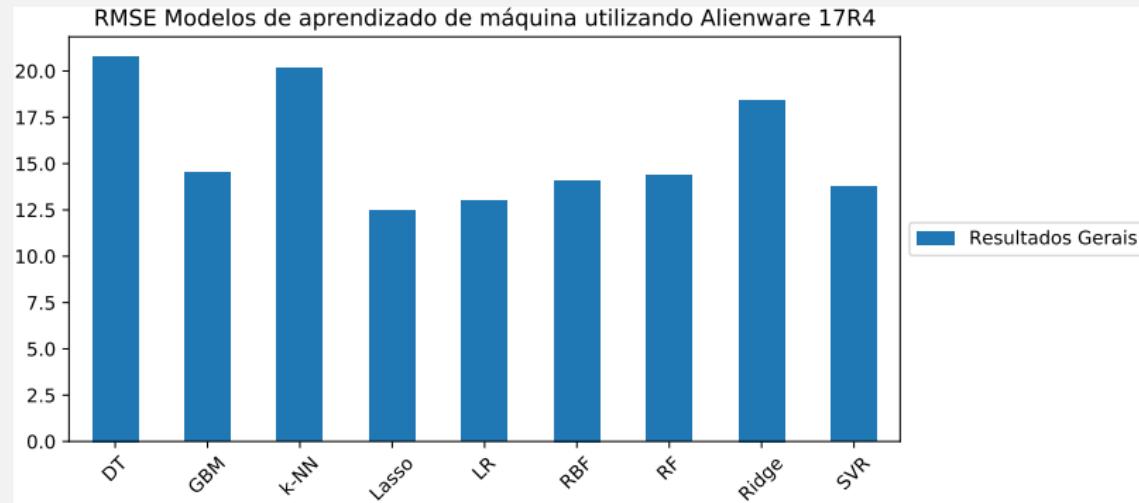
Resultados MP_{2.5}

Resultados Aprendizado de Máquina



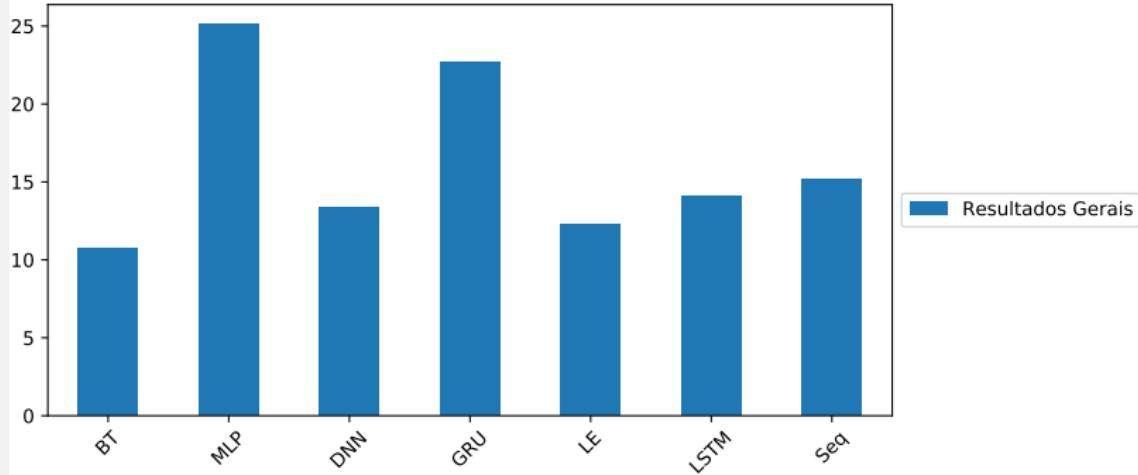
32

Resultados usando RMSE



Resultados usando RMSE

RMSE Modelos de aprendizado de profundo utilizando Alienware 17R4



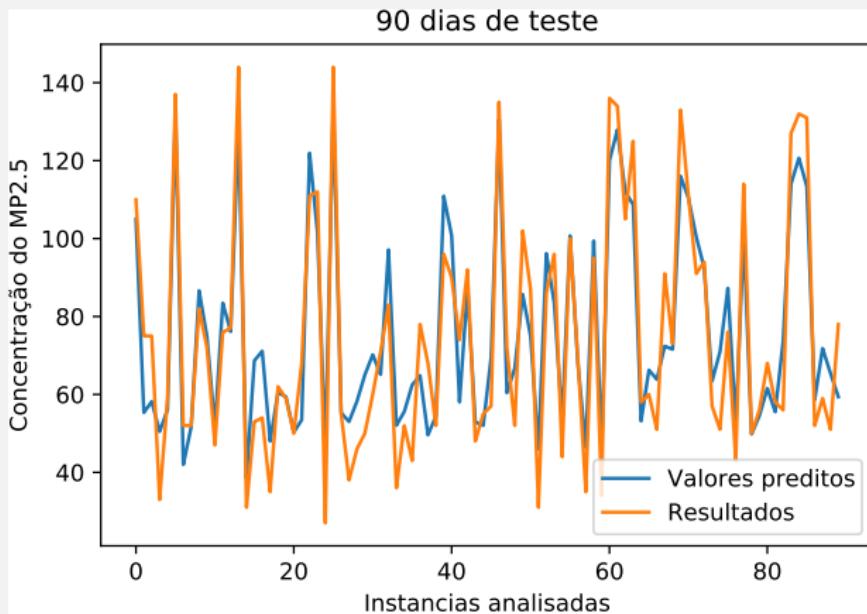
Resultados MP_{2.5}

Resultados do modelo generalizado



34

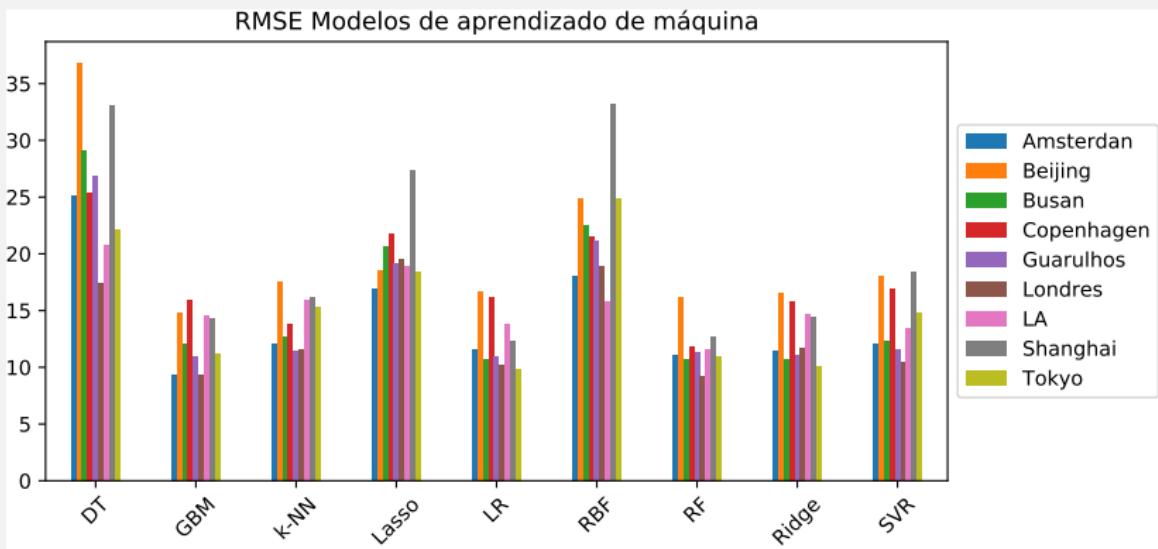
Últimos 90 dias do modelo generalizado utilizando
BoostedTrees



Resultados MP_{2.5}

Modelos de cidades específicas

RMSE de cada cidade em cada modelo de aprendizado de máquina analisado



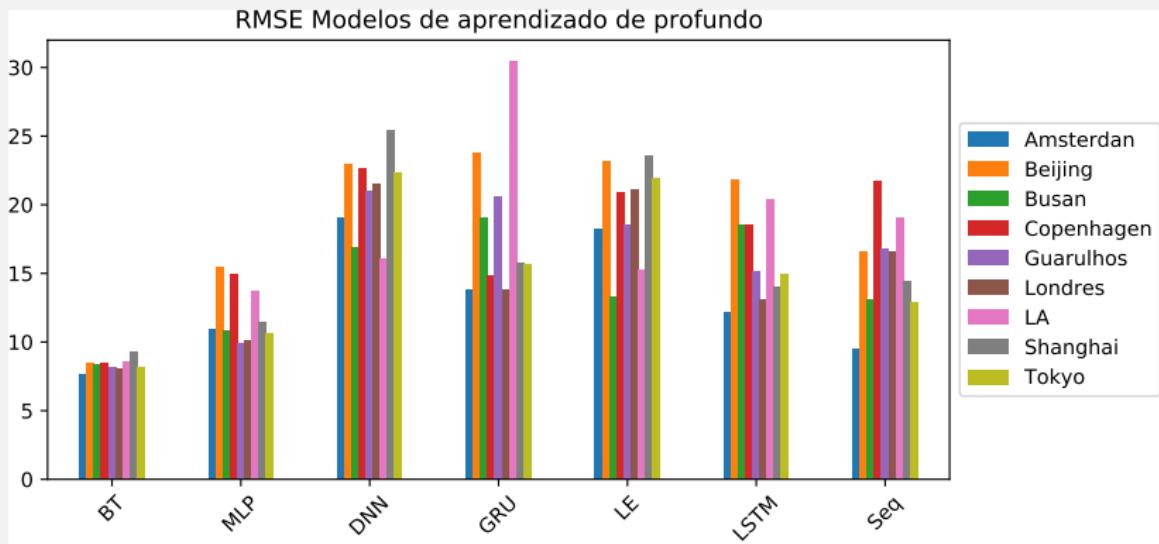
Resultados MP_{2.5}

Modelos de cidades específicas



36

RMSE de cada cidade em cada modelo de aprendizado profundo analisado

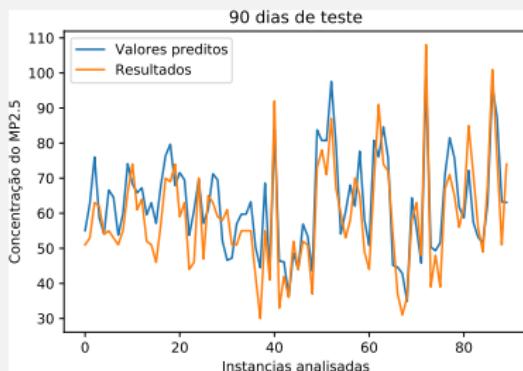
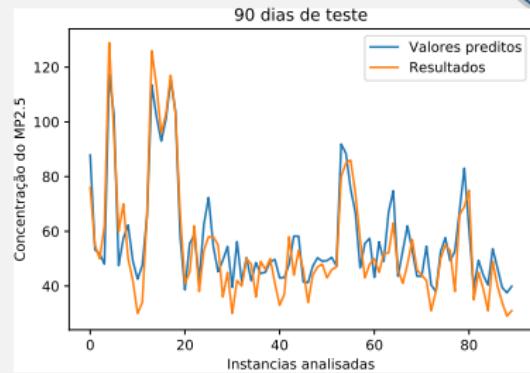
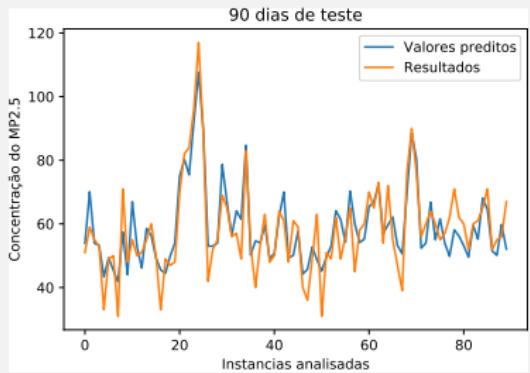


Resultados MP_{2.5}

Modelos reais de Amsterdam, São Paulo e Londres



37

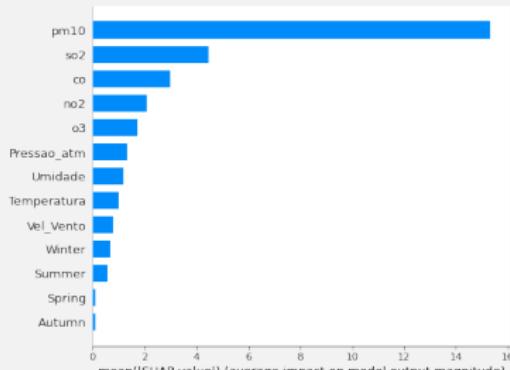
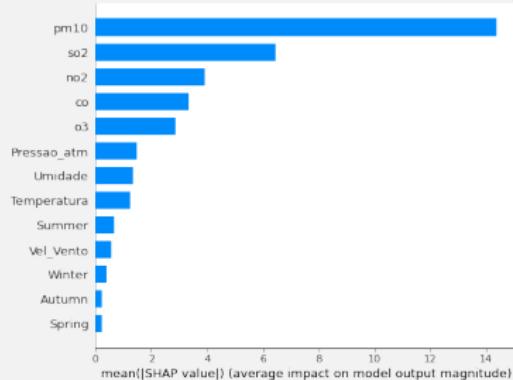
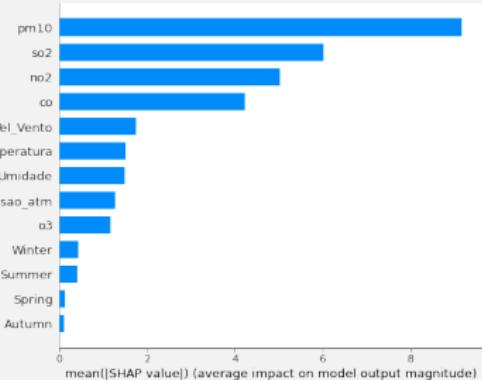


Shap Values MP_{2.5}

Modelos reais de Amsterdam, São Paulo e Londres



38



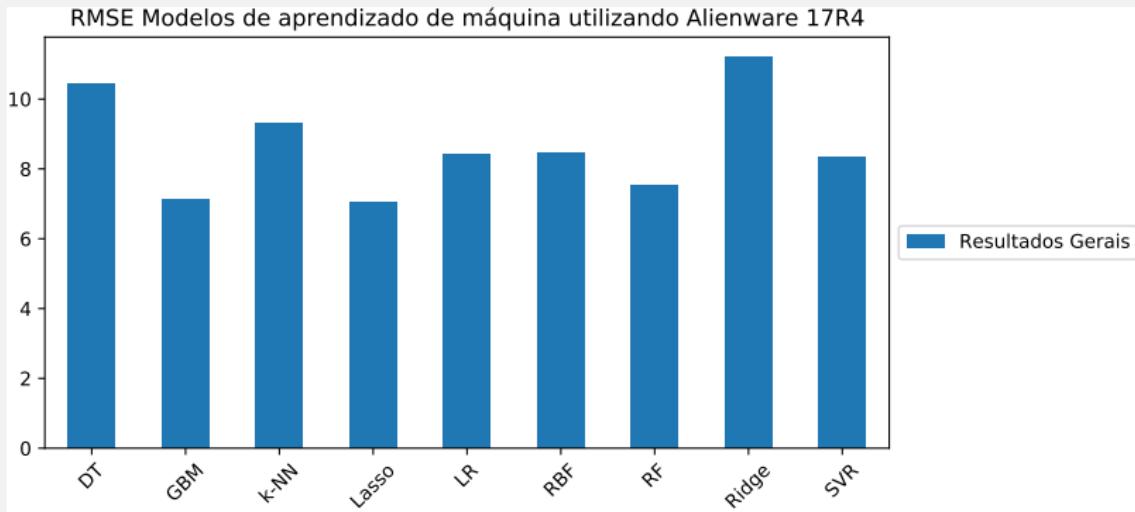
Resultados MP₁₀

Resultados Aprendizado de Máquina



39

Resultados usando RMSE



Resultados MP₁₀

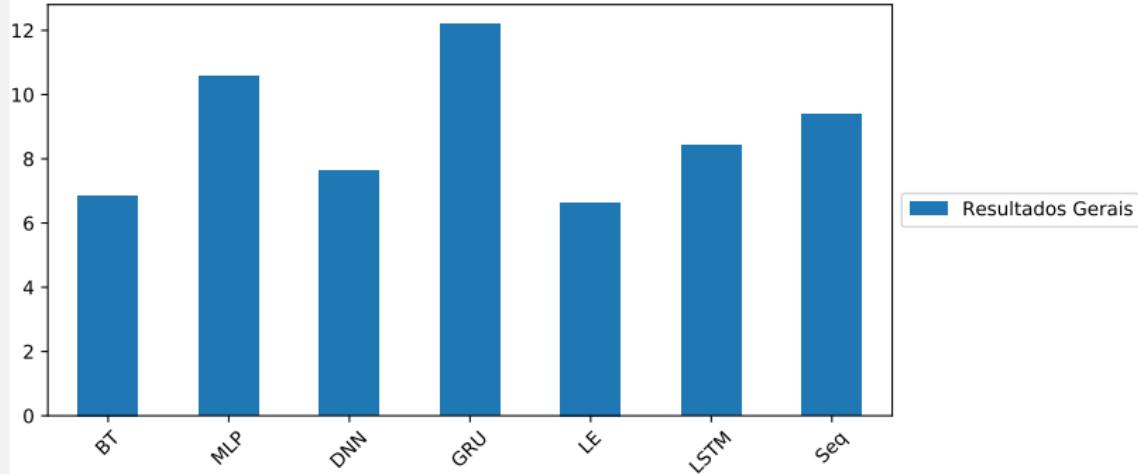
Resultados Aprendizado de Profundo



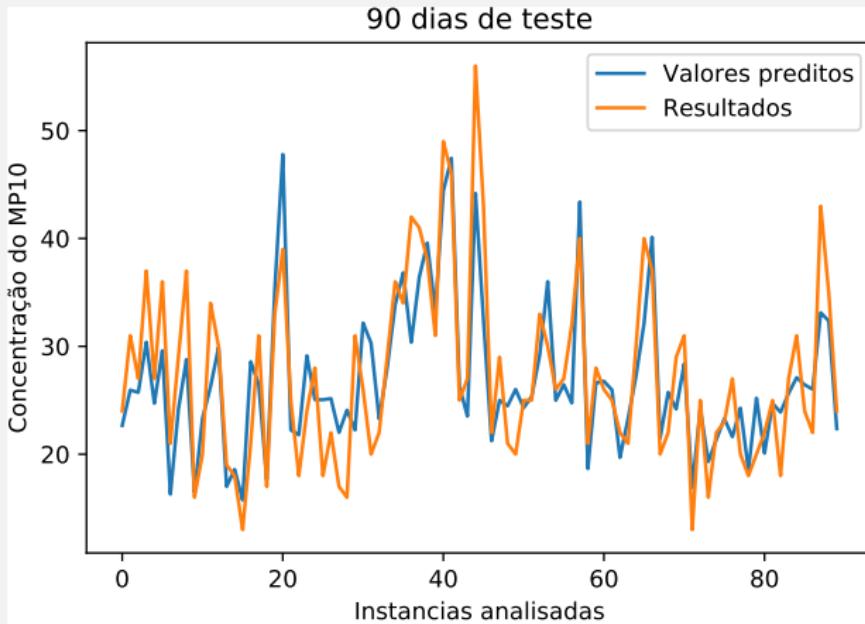
40

Resultados usando RMSE

RMSE Modelos de aprendizado de profundo utilizando Alienware 17R4



Últimos 90 dias do modelo generalizado utilizando BoostedTrees



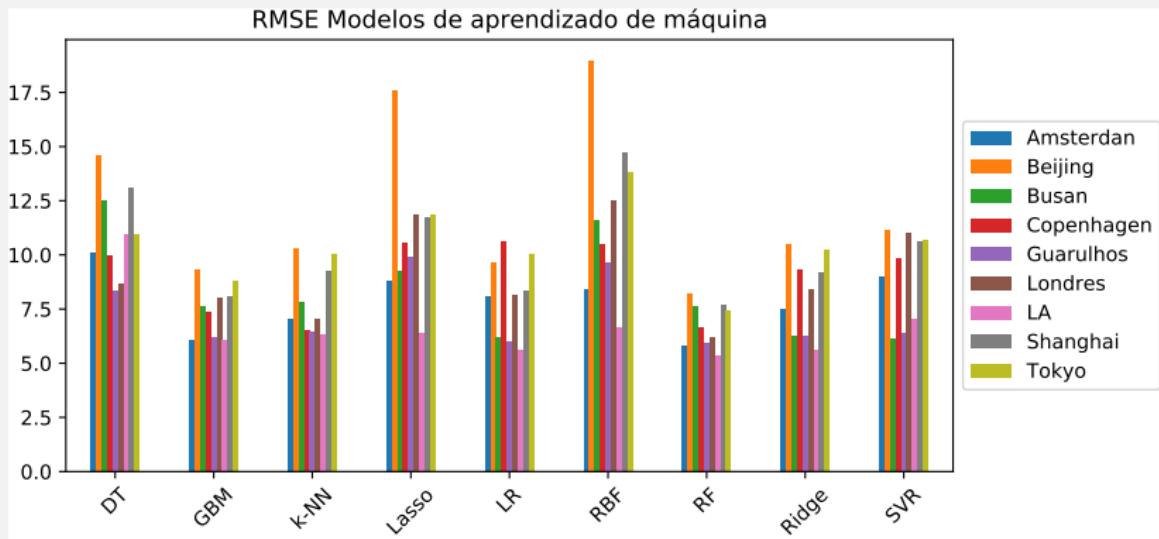
Resultados MP₁₀

Modelos de cidades específicas



42

RMSE de cada cidade em cada modelo de aprendizado de máquina analisado



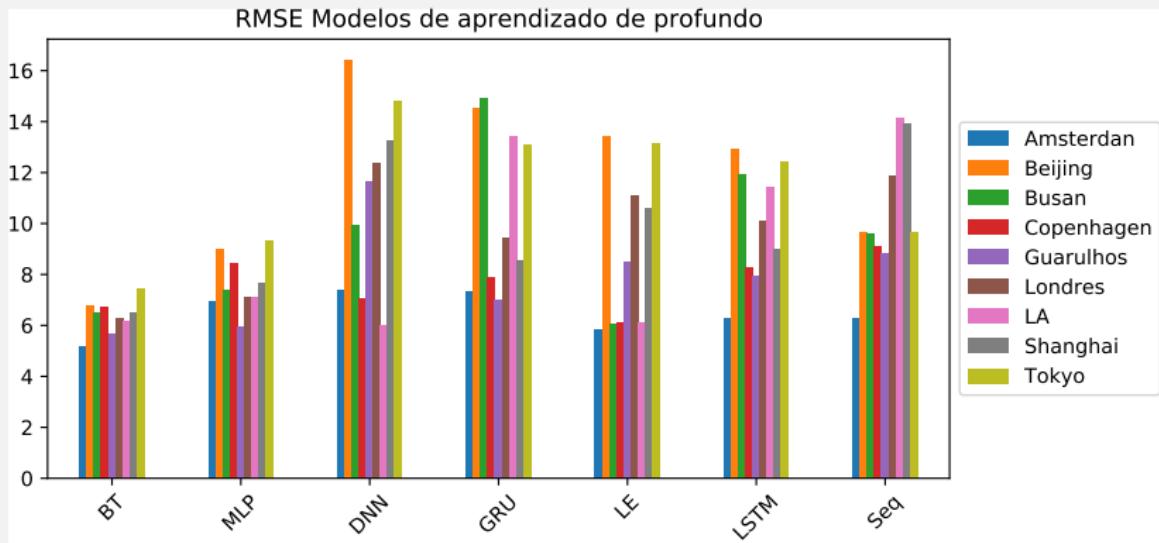
Resultados MP₁₀

Modelos de cidades específicas



43

RMSE de cada cidade em cada modelo de aprendizado profundo analisado

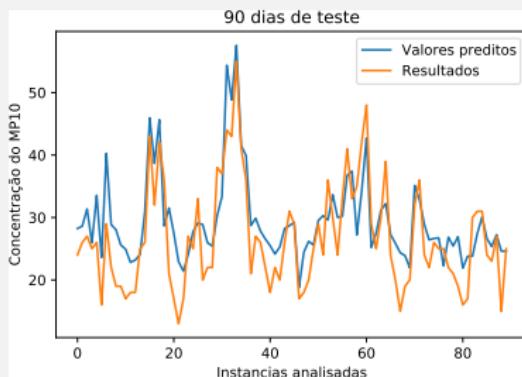
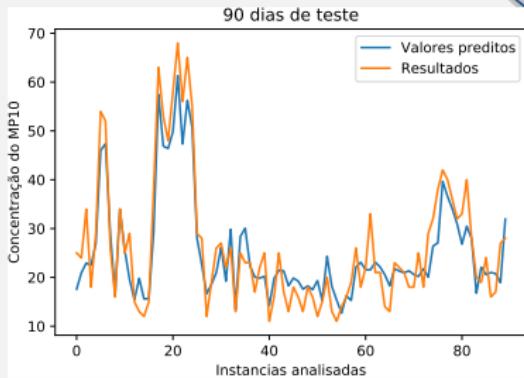
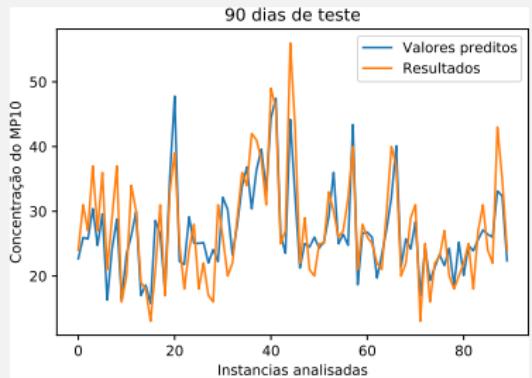


Resultados MP₁₀

Modelos reais de Amsterdam, São Paulo e Los Angeles



44

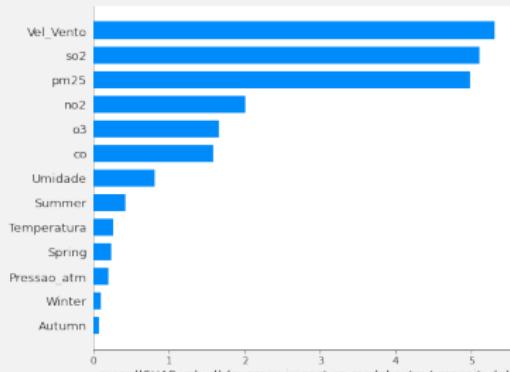
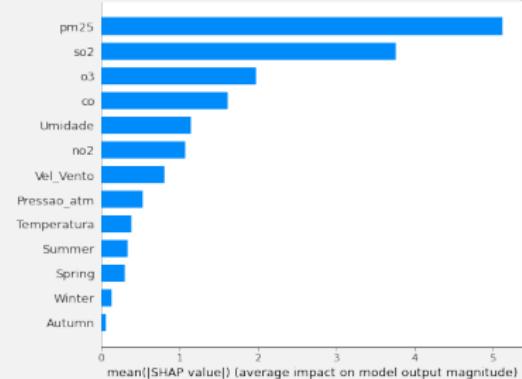
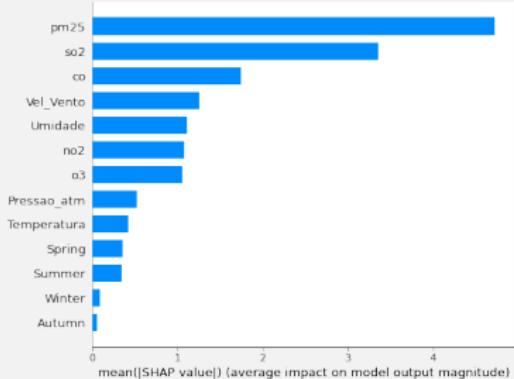


Shap Values MP₁₀

Modelos reais de Amsterdam, São Paulo e Los Angeles



45



Agenda



1. Introdução

2. Conceitos Importantes

3. Trabalhos Relacionados

4. Proposta

5. Resultados

6. Cronograma

- Análise de outras janelas de tempo dos poluentes

Referências I



- [1] M. A. Zaidan, N. H. Motlagh, P. L. Fung, D. Lu, H. Timonen, J. Kuula, J. V. Niemi, S. Tarkoma, T. Petäjä, M. Kulmala *et al.*, “Intelligent calibration and virtual sensing for integrated low-cost air quality sensors,” *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 638–13 652, 2020.
- [2] A. B. Campolina, P. H. L. Rettore, M. D. V. Machado, and A. A. Loureiro, “On the design of vehicular virtual sensors,” in *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2017, pp. 134–141.
- [3] C. Nguyen and D. Hoang, “Software-defined virtual sensors for provisioning iot services on demand,” in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 2020, pp. 796–802.

Referências II



- [4] S. Oehmcke, O. Zielinski, and O. Kramer, "Input quality aware convolutional lstm networks for virtual marine sensors," *Neurocomputing*, vol. 275, pp. 2603–2615, 2018.



Duvidas?