



Modelagem e simulação da transmissão de arboviroses e Covid-19 baseado em aprendizado de máquina para a predição de dengue, zika, chikungunya e Covid-19

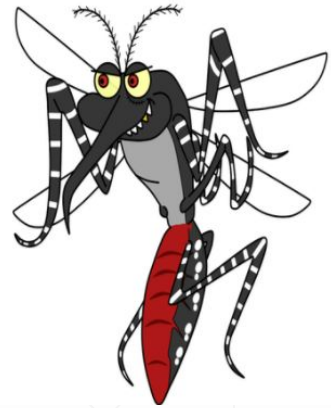


Ms. Ana Clara Gomes da Silva

Prof. Dr. Wellington Pinheiro dos Santos

Introdução

- A prevenção e controle das arboviroses têm sido um grande desafio de saúde pública para muitos países;
- Estima-se que 390 milhões de casos de dengue ocorram a cada ano;
- A partir do ano de 2015, outros arbovírus passaram a interagir com o vetor transmissor do mosquito da dengue (chikungunya e zika);
- A interação dos vírus com o *Aedes aegypti* juntamente com a rápida difusão das doenças levantam novos desafios em relação ao monitoramento e controle do vetor;

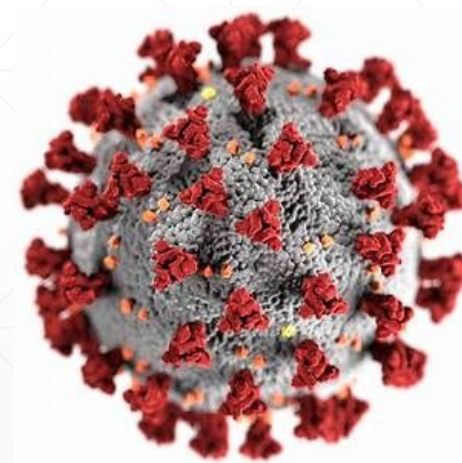


Introdução

- A transmissão das arboviroses é um processo complexo que envolve a interação de múltiplos agentes em um espaço bastante heterogêneo:
 - Populações humanas;
 - Mosquito e vírus;
 - Fatores ambientais.
- Condições climáticas locais interferem:
 - Eclosão dos ovos;
 - O tempo de vida dos mosquitos;
 - Dispersão dos mosquitos.

Introdução

- Dezembro de 2019, surto de uma pneumonia atípica foi atribuído ao novo coronavírus;
- Grande mobilidade facilitou a rápida transmissão do vírus;
- 30/01/2020 declarada Emergência de Saúde Pública de Interesse Internacional (PHEIC) e em março de 2020 declarada pandemia pela OMS;
- Mais de 760 milhões de casos confirmados e quase 8 milhões de mortes em todo o mundo*.



Introdução

- Exames mais comuns para detectar a infecção por Covid-19:
 - Testes rápidos;
 - RT-PCR (transcrição reversa seguida de reação em cadeia da polimerase).
- Não existe tratamento específico e a vacinação começou só no início de 2021;
- Estratégias de isolamento social e diminuição da capacidade máxima de pessoas em estabelecimentos e até fechamento de locais;
- Importância da predição de casos para medidas de curto e médio prazo.

Motivação e Justificativa

- É importante entender a dinâmica das arboviroses e Covid-19 em escala local:

Epidemiologia
Digital

Geoprocessamento

Mineração de
Dados

Aprendizagem de
Máquina

- Controle e simulação da disseminação das doenças com abordagem temporal e espaço-temporal.

Métodos de predição de arboviroses e Covid-19

- Modelagem matemática (sistema de EDOs, EDPs);
- Modelagem estatística (modelos autorregressivos e lineares generalizados);
- Modelagem usando aprendizado de máquina (regressão linear, random forest, support vector machine, multilayer perceptron, etc)

Proposta

- Softwares:



Weka
Versão 3.8



QuantumGIS
Versão 2.18.25



Python
Versão 3.7.4



R Studio
Versão 3.6.3

STICKERS

Proposta

Base de Dados Arboviroses

- Plataforma de Dados Abertos da Prefeitura do Recife: dados de localização de *Aedes aegypti* e distribuição das arboviroses e de sintomas;
- Agência Pernambucana de Águas e Clima APAC: informações de distribuição de temperatura e umidade;
- Sistema de Informações Geográficas SIGHPE: informações pluviométricas mensais;
- Dados coletados de 2013 a 2016, e monta as bases de dados de 2014 a 2016.

Proposta

Predição espaço-temporal de arboviroses

Sistema de predição da distribuição espaço-temporal do mosquito *Aedes aegypti* no Recife usando mapas de distribuição de criadouros e variáveis climáticas como chuva, temperatura e velocidade dos ventos

- Coleta de dados de criadouros e das variáveis climáticas dos anos de 2014, 2015 e 2016;
- Criação dos shapefile (arquivo .shp) e geração de mapas de distribuição espacial (método IDW) com o software QGIS;
- Agrupamento dos mapas construídos com um script em Python;
- Criação dos conjuntos de predição do mapa de distribuição de criadouros;
- Visualização dos mapas de predição da distribuição de criadouros usando o QGIS.



Fonte: Google Maps

Proposta

Base de dados de Covid-19 - abordagem temporal

- Portal Brasil.io: dados de casos confirmados e óbitos de cada estado brasileiro.

Base de dados de Covid-19 - abordagem espaço-temporal

- Portal Brasil.io: dados de casos confirmados e óbitos de cada município do Brasil.

Proposta

Predição temporal de Covid-19

Sistema para predição de casos e óbitos de covid-19 para o Brasil e cada uma das suas unidades federativas.

- No portal Brasil.io, o software COVID-SGIS coleta os dados de casos confirmados e óbitos de cada estado do Brasil;
 - Dados reunidos em um arquivo .csv
 - Criação dos conjuntos de dados de treinamento para os modelos criados de acumulados de casos confirmados e mortes de Covid-19;
 - Criação dos modelos ARIMA;
 - Avaliação se a série é estacionária ou não;
 - Estimação dos parâmetros p e q usando ACF e PACF;
 - Avaliação dos modelos usando os critérios AIC e BIC.
-

Proposta

Predição espaço-temporal de Covid-19

Sistema para predição da distribuição de casos e óbitos de Covid-19 para cada município do Brasil.

- No portal Brasil.io são coletadas as informações de casos confirmados e óbitos de cada município do Brasil;
 - Os bandos de dados são enviados para o módulo de interpolação para criação dos conjuntos de treino e neste módulo os números de casos são distribuídos em uma grade não homogênea (com latitude e longitude de cada município);
 - Geração da grade homogênea usando o método de ponderação de distância inversa (IDW) e estimada a distribuição de casos confirmados e óbitos;
 - Geração dos mapas de predição (mapas de 3 dias consecutivos para prever o mapa de distribuição do dia seguinte).
-

Resultados

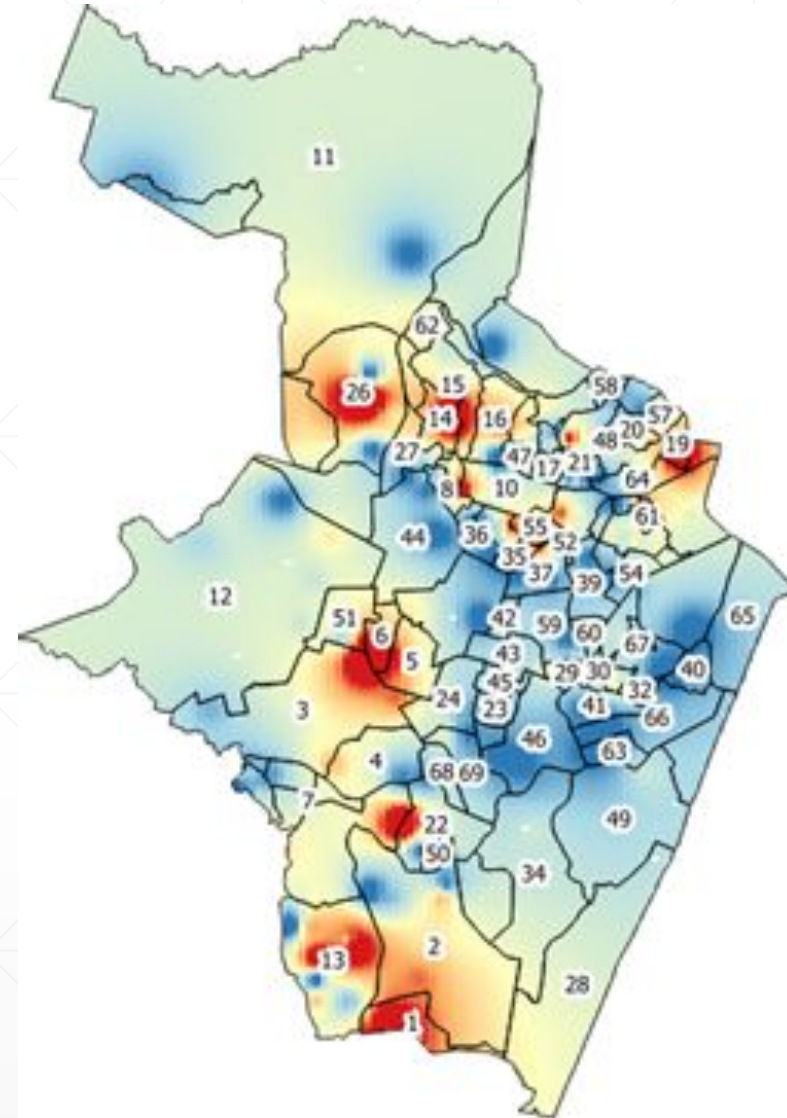
Predição espaço-temporal de arboviroses

Regression method	Configuration	RRSE(%)		R		Training time (s)	
		average	standard deviation	average	standard deviation	average	standard deviation
Linear Regression	-	51.09	13.86	0.8488	0.05	0.09	0.03
Random Forest	10 trees	25.47	5.07	0.9687	0.01	0.31	0.06
	20 trees	23.54	4.93	0.9745	0.01	0.62	0.11
	30 trees	22.87	4.91	0.9764	0.01	0.93	0.16
	40 trees	22.53	4.91	0.9773	0.01	1.27	0.20
Support Vector Regressor	polynomial kernel, p=1	53.34	14.50	0.8343	0.06	18.72	6.23
	polynomial kernel, p=2	30.30	9.98	0.9488	0.03	191.17	65.32
	polynomial kernel, p=3	14.60	6.41	0.9875	0.01	3121.20	1083.96
	RBF kernel	63.84	7.44	0.8002	0.06	19.49	4.48
MLP, one hidden layer	10 neurons	23.29	9.44	0.9750	0.02	62.68	7.95
	20 neurons	22.63	8.44	0.9770	0.01	74.50	9.50
	30 neurons	20.61	7.61	0.9809	0.01	86.76	10.49

Resultados

Predição espaço-temporal de arboviroses

- mapa de predição da distribuição de criadouros de *Aedes aegypti* no Recife para o 6º bimestre (nov/dez) de 2014. Gerado usando o regressor SVM com kernel polinomial $p=3$.



Resultados

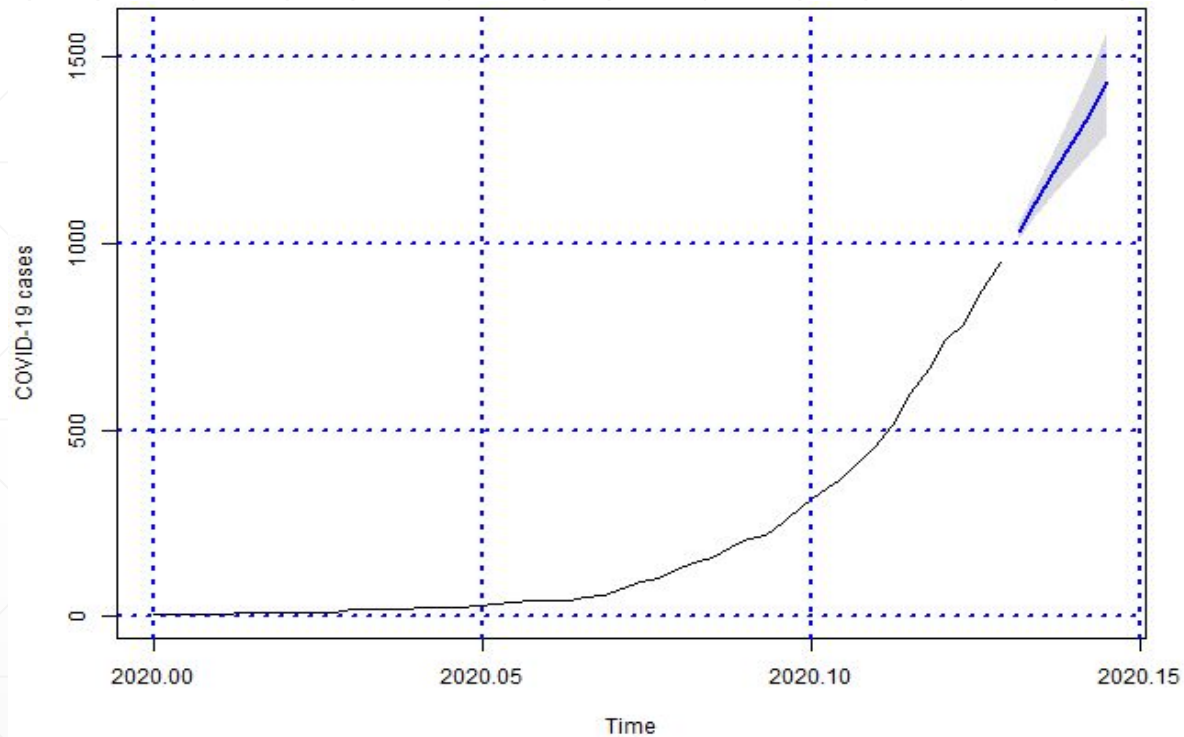
Predição temporal de Covid-19

- Métricas de avaliação dos modelos: coeficientes de correlação de Pearson, Spearman e Kendall e o erro quadrático relativo RMSE(%)

state	model	pearson	spearman	kendall	rmse	rmse (%)	MAPE	MPE	MAE	ME	number of observations
Acre	ARIMA(2,2,0)	0.99890	0.99515	0.96645	9.30	4.78	7.50	1.48	6.34	1.56	50
Alagoas	ARIMA(0,2,1)	0.99873	0.99371	0.96337	22.21	5.13	6.93	2.36	10.83	2.50	59
Amazonas	ARIMA(0,2,1)	0.99928	0.99899	0.99194	87.65	4.10	6.56	3.77	54.04	20.92	54
Amapá	ARIMA(0,2,1)	0.99647	0.99826	0.98327	43.04	8.78	10.59	3.77	21.64	8.31	47
Bahia	ARIMA(2,2,1)	0.99971	0.99892	0.98795	28.59	2.56	7.78	1.30	18.40	7.36	61
Ceará	ARIMA(0,2,1)	0.99408	1.00000	1.00000	351.5	11.31	5.69	4.65	122.61	75.44	51
Distrito Federal	ARIMA(0,2,1)	0.99922	0.99937	0.99151	20.70	4.02	7.00	3.56	14.29	3.87	60
Espírito Santo	ARIMA(0,2,1)	0.99874	0.99870	0.98565	51.50	5.09	7.82	3.70	26.9	8.60	62
Goiás	ARIMA(0,2,1)	0.99884	0.99971	0.99596	13.55	4.91	5.50	2.62	8.54	2.67	55
Maranhão	ARIMA(0,2,0)	0.99908	0.99867	0.98841	63.78	4.42	10.00	1.20	41.77	10.57	47
Minas Gerais	ARIMA(1,2,2)	0.99945	0.99982	0.99766	24.16	3.46	6.88	3.40	15.64	4.57	59
Mato Grosso do Sul	ARIMA(0,2,1)	0.99860	0.99966	0.99600	4.88	5.40	5.06	2.37	3.36	0.78	53
Mato Grosso	ARIMA(0,2,1)	0.99828	0.99841	0.98654	6.76	6.03	7.75	3.54	4.96	1.58	47
Pará	ARIMA(0,2,1)	0.99868	0.99941	0.99191	71.56	5.52	9.5	4.44	41.76	19.46	49
Paraíba	ARIMA(3,2,0)	0.99931	0.99586	0.97577	15.11	3.91	13.92	3.37	8.52	3.42	55
Pernambuco	ARIMA(0,2,1)	0.99949	0.99955	0.99360	89.79	3.28	7.58	3.08	48.25	19.18	55
Piauí	ARIMA(1,2,0)	0.99928	0.99813	0.98266	10.28	4.00	7.2	1.12	6.41	2.55	48
Paraná	ARIMA(0,2,2)	0.99942	0.99949	0.99392	18.18	3.47	4.96	2.90	12.96	2.67	55
Rio de Janeiro	ARIMA(0,2,2)	0.99960	0.99918	0.99258	105.29	2.96	8.78	2.44	67.51	27.06	62
Rio Grande do Norte	ARIMA(0,2,1)	0.99741	0.99727	0.97657	34.52	7.37	8.24	4.36	18.31	6.99	55
Rondônia	ARIMA(0,2,0)	0.99819	0.99485	0.96633	14.40	6.20	10.61	-0.09	7.98	2.19	47
Roraima	ARIMA(0,2,0)	0.97919	0.99321	0.94769	43.73	21.08	12.08	1.56	19.02	6.33	46
Rio Grande do Sul	ARIMA(2,2,2)	0.99882	0.99940	0.99278	27.95	5.10	7.56	2.71	17.73	6.35	57
Santa Catarina	ARIMA(0,2,1)	0.99254	0.99955	0.99562	96.88	12.41	6.72	4.25	38.95	17.18	55
Sergipe	ARIMA(3,2,0)	0.99598	0.99631	0.97041	18.97	9.11	9.84	1.39	7.93	3.22	53
São Paulo	ARIMA(0,2,2)	0.99917	0.99914	0.98688	411.73	4.14	9.96	4.26	243.34	74.83	71
Tocantins	ARIMA(0,2,2)	0.99785	0.99308	0.95766	5.52	6.76	10.02	2.40	3.07	1.20	49
Brazil	ARIMA(0,2,1)	0.99978	0.99718	0.97963	688.24	2.20	15.65	-0.06	401.29	142.02	71

Resultados

Predição temporal de Covid-19



Predição dos casos acumulados de Covid-19 para o estado do Piauí para 06/05/2020 a 11/05/2020 Fonte: LIMA et al, 2020

- Predição ARIMA: foram criados modelos para os 26 estados brasileiros e o distrito federal;

Resultados

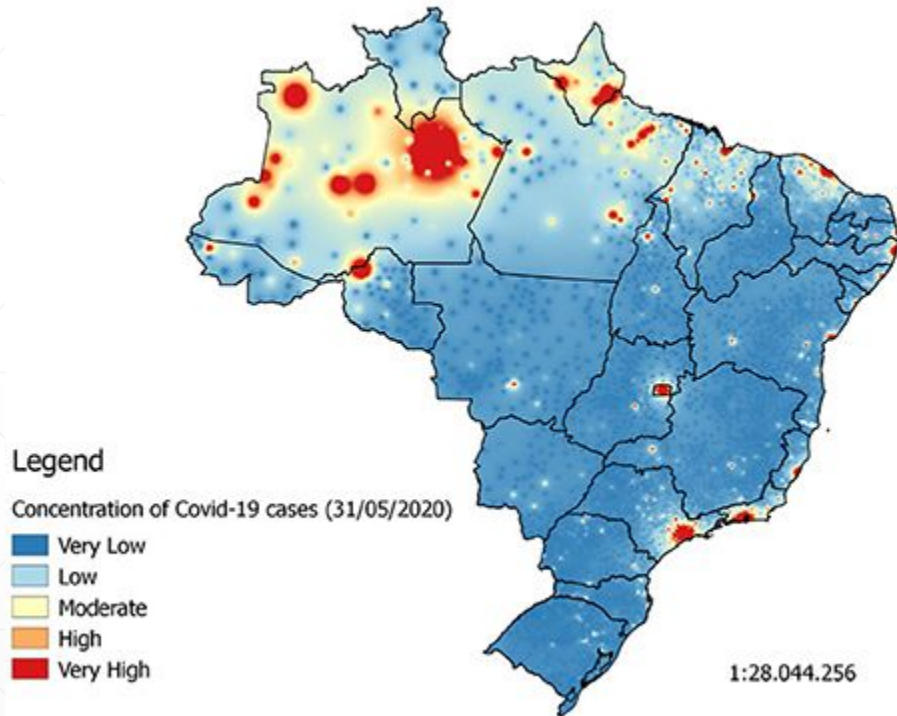
Predição espaço-temporal de Covid-19

- Métricas de avaliação dos modelo: o coeficiente de correlação e o Erro Quadrático Relativo (RMSE percentual);
 - Para predição: janela de 3 dias para a predição da distribuição espacial para prever a do dia seguinte;
-

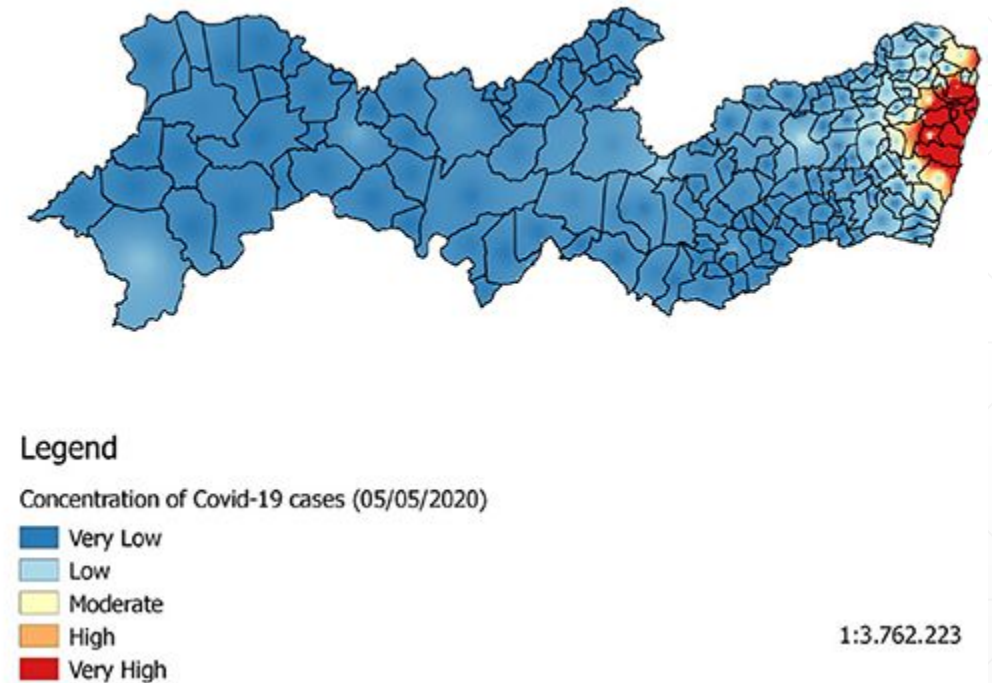
Resultados

Predição espaço-temporal de Covid-19

Distribution map of the number of accumulated cases of Covid-19 (31/05/2020)



Distribution map of the number of accumulated cases of Covid-19 (05/05/2020)



À direita está o mapa de distribuição de casos acumulados da Covid-19 (mapa de calor) no Estado de Pernambuco em 5 de maio de 2020. À esquerda está o mapa de distribuição de casos acumulados da Covid-19 no Brasil em 31 de maio de 2020. Fonte: da Silva et al, 2021

Resultados

Predição espaço-temporal de Covid-19

Regression method	Configuration	Correlation coefficient		RMSE (%)		Training time (s)	
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Linear regression	–	0.9826	0.143	11.42	22.94	0.003	0.006
	10 trees	0.9737	0.072	22.35	24.08	0.06	0.01
	20 trees	0.9752	0.070	21.59	23.03	0.20	0.04
	30 trees	0.9756	0.069	21.32	22.68	0.29	0.06
Random forest	40 trees	0.9759	0.068	21.17	22.33	0.22	0.04
	20 neurons	0.9948	0.006	11.29	7.12	13.91	0.19
MLP	30 neurons	0.9867	0.011	17.9	4.12	6.48	0.30
	Polynomial kernel, $p = 1$	0.9670	0.109	18.39	24.08	3.82	2.14
	Polynomial kernel, $p = 2$	0.9818	0.006	31.70	10.62	7.99	10.57
SVR	Polynomial kernel, $p = 3$	0.9623	0.004	45.15	17.35	7.55	9.45
	RBF kernel	0.5341	0.375	87.19	14.94	39.97	56.17

For each regressor, we calculated the correlation coefficient, the relative square error [RMSE (%)], and the training time.

Resultados do desempenho dos algoritmos de regressão linear, do perceptron multicamadas (MLP) e do regressor vetorial de suporte (SVR) para o conjunto de dados do Brasil. Fonte: da Silva et al, 2021

Resultados

Predição espaço-temporal de Covid-19

Regression method	Configuration	Correlation coefficient		RMSE (%)		Training time(s)	
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Linear regression	–	0.9991	0.006	1.92	3.81	0.02	0.01
	10 trees	0.9983	0.003	6.14	4.14	0.27	0.05
	20 trees	0.9985	0.003	5.78	4.03	0.52	0.09
	30 trees	0.9985	0.003	5.65	4.01	0.78	0.11
Random forest	40 trees	0.9985	0.003	5.57	3.99	1.21	0.14
	20 neurons	0.9990	0.005	4.07	4.95	45.82	1.56
MLP	30 neurons	0.9991	0.005	3.81	4.95	62.69	3.60
	Polynomial kernel, $p = 1$	0.9988	0.006	3.28	4.05	18.61	22.15
	Polynomial kernel, $p = 2$	0.9964	0.011	6.48	5.85	82.79	34.22
	Polynomial kernel, $p = 3$	0.9989	0.001	4.48	1.99	89.64	8.34
SVR	RBF kernel	0.9387	0.064	76.86	13.82	112.78	122.81

For each regressor, we calculated the correlation coefficient, the relative square error [RMSE (%)], and the training time.

Resultados do desempenho dos algoritmos de regressão linear, do perceptron multicamadas (MLP) e do regressor vetorial de suporte (SVR) para o conjunto de dados de Pernambuco. Fonte: da Silva et al, 2021

Resultados

Predição espaço-temporal de Covid-19

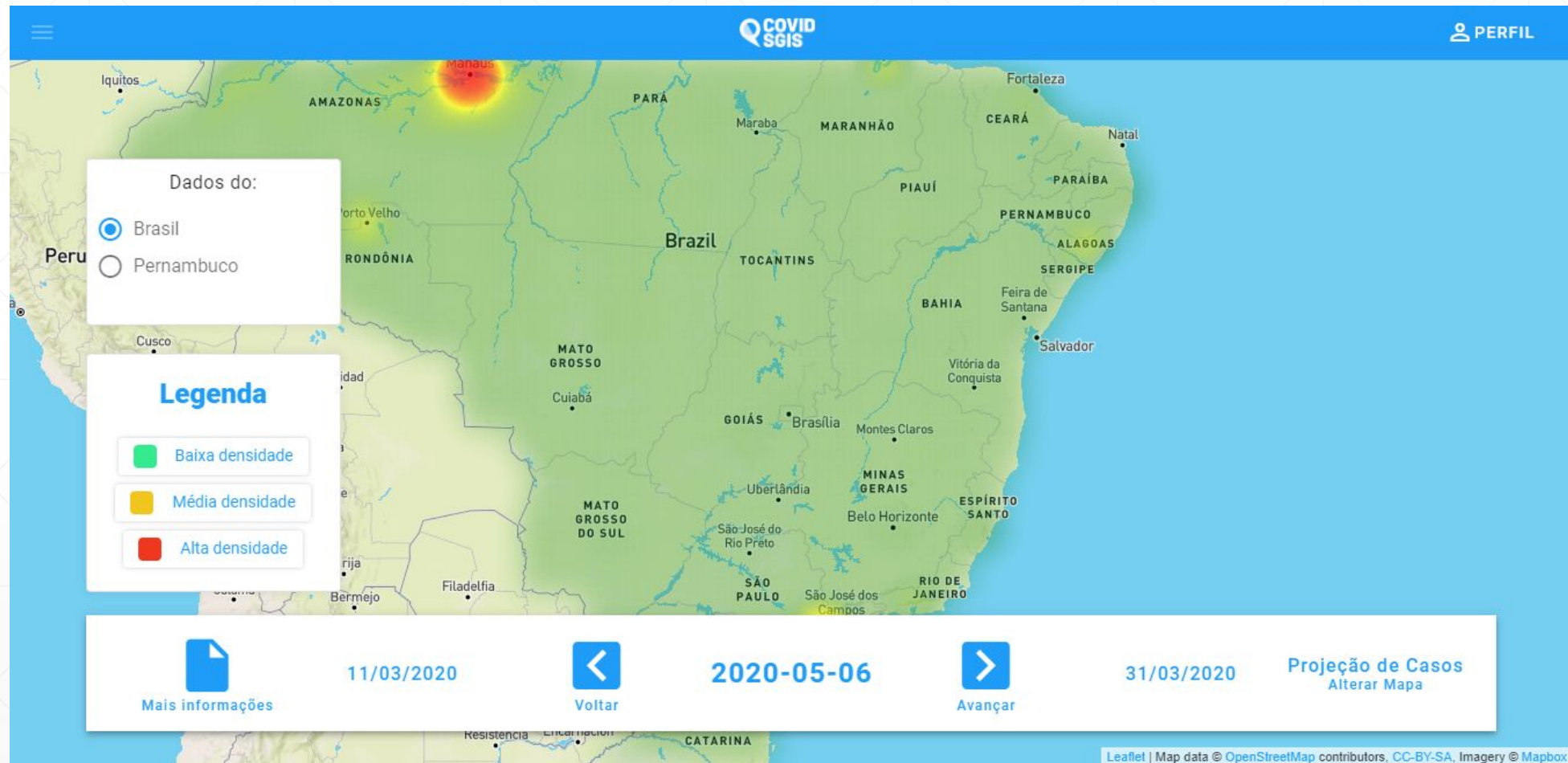
	Regression method	Prediction date	Correlation coefficient	RMSE (%)
PERNAMBUCO	Linear regression	25/05/2020	0.999994	1.32
		26/05/2020	0.999990	3.27
		27/05/2020	0.999970	3.84
BRAZIL	Linear regression	25/05/2020	0.998447	5.68
		26/05/2020	0.998840	5.03
		27/05/2020	0.978811	20.86
PERNAMBUCO	SVR, kernel = RBF	25/05/2020	0.949280	86.55
		26/05/2020	0.869430	99.33
		27/05/2020	0.511389	101.2
BRAZIL	SVR, kernel = RBF	25/05/2020	0.993423	42.15
		26/05/2020	0.951116	71.23
		27/05/2020	0.751036	87.91

We forecasted the distribution of accumulated cases of Covid-19 between May 25 and 27, 2020.

Resultados da validação dos modelos criados por meio de regressão linear e SVR, kernel RBF, para Brasil e Pernambuco. Fonte: da Silva et al, 2021

Resultados

Aplicação web: COVID-SGIS



Resultados

Aplicação web: COVID-SGIS

Dashboard do Brasil

Óbitos Totais

17834

Números totais somados

Casos Confirmados

267922

Último valor registrado

Confirmados Relativo

156.84

Valores para cada 100.000 habitantes

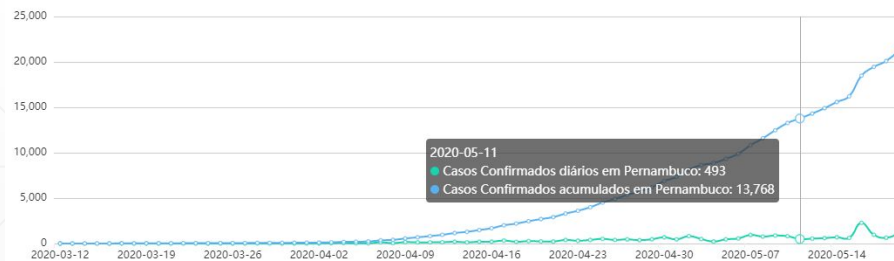
Taxa de Letalidade no Brasil



Números percentuais baseados no número total de casos confirmados, de acordo com a razão n° Óbitos ÷ n° Casos Confirmados

Casos Confirmados Diários e Acumulados

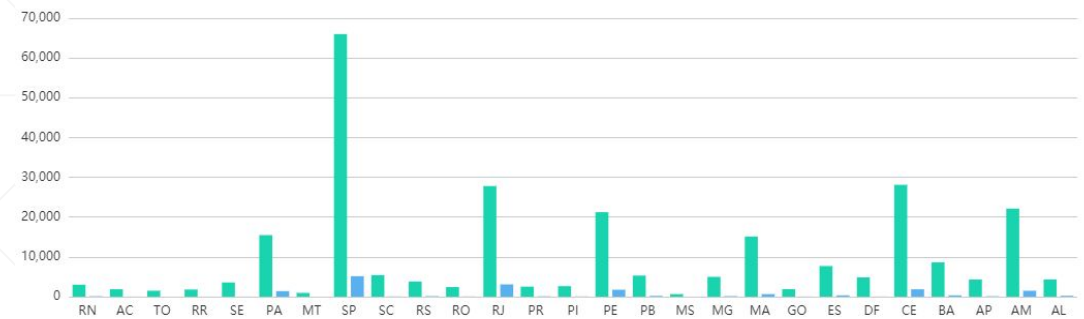
○ Casos Confirmados diários em Pernambuco ○ Casos Confirmados acumulados em Pernambuco



Selecione o país: Seleccione o estado: Seleccione a cidade: [FILTRAR](#)

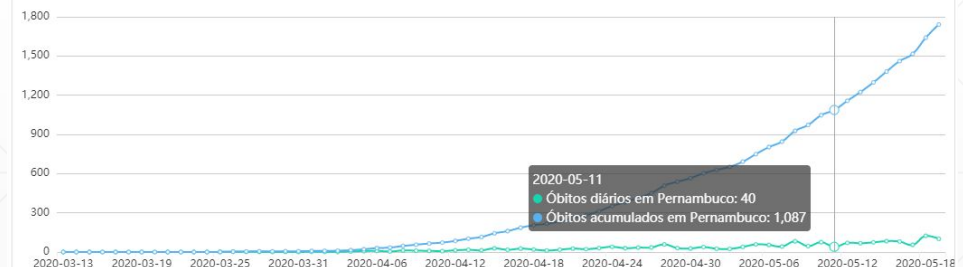
Distribuição de Casos Confirmados e Óbitos no Brasil

■ Confirmados ■ Óbitos



Número de Óbitos Diários e Acumulados

○ Óbitos diários em Pernambuco ○ Óbitos acumulados em Pernambuco



Selecione o país: Seleccione o estado: Seleccione a cidade: [FILTRAR](#)

Conclusões

- Sistemas robustos;
 - Estas abordagens contém informações importantes para dar suporte aos agentes de saúde e epidemiologistas no planejamento de ações a longo e curto prazo, para evitar grandes surtos de arboviroses;
 - A aplicação web COVID-SGIS tem uma interface amigável onde o usuários podem personalizar painéis e visualizar informações sobre Covid-19 no Brasil;
 - A abordagem temporal de Covid-19 proporciona o entendimento de como a doença evolui com o passar do tempo;
 - A abordagem espaço-temporal de Covid-19 torna possível um melhor entendimento sobre a dispersão da doença e como ocorre a evolução desta dispersão;
 - Ao juntar as duas abordagens, o COVID-SGIS traz uma avaliação ampla das regiões;
-

Obrigada!

