

# Criação do índice de vulnerabilidade e prioridade de vacinação das cidades brasileiras em relação ao COVID-19

Grupo Dice-Science

Gabriel Rauta Buiar  
Engenharia de Computação  
UTFPR  
Curitiba, Brasil  
gabrielbuiar@gmail.com

Luan Carlos Klein  
Engenharia de Computação  
UTFPR  
Curitiba, Brasil  
luanklein@alunos.utfpr.edu.br

Otávio Thomas Bertucini  
Sistemas de Informação  
UTFPR  
Curitiba, Brasil  
otaviobertucini@gmail.com

**Resumo**—Due to the COVID-19 pandemic, vaccination of the population is of fundamental importance and of essential value. However, it is a fact that, at first, not enough vaccinations will be available to the population. Thus, there is a need to define what the vaccination priorities will be and which cities should be prioritized in this process. The present work has two research questions: Which factors influence the vulnerability of Brazilian cities in relation to COVID-19 and which cities present greater vulnerability in terms of the impacts on health related to the new coronavirus?. After several analyzes and researches, it was possible to verify that factors such as hospital occupation, mortality rates and demographic indexes have huge influence over the vulnerability of a city. In addition, it was also possible to determine an index for each Brazilian city, giving different weights for each chosen factor.

## I. INTRODUÇÃO

Diante da pandemia do COVID-19, a vacinação da população é de fundamental importância e de valor imprescindível. Entretanto, é fato de que no primeiro momento, não haverá vacinas suficiente para toda a população. Com isso, é necessário definir quais serão as prioridades de vacinação e quais cidades devem ser priorizadas nesse processo. Para tal efetividade, é importante que seja levado em conta a vulnerabilidade das cidades em geral. O presente trabalho tem como perguntas de pesquisa: Quais fatores influenciam na vulnerabilidade das cidades brasileiras em relação ao COVID-19 e quais cidades apresentam maior vulnerabilidade em termos dos impactos na saúde referente ao novo coronavírus? Após diversas análises e pesquisas, foi possível constatar que fatores como ocupação hospitalar, mortalidade e dados demográficos são influenciadores diretos. Além disso, também foi possível definir um índice para cada cidade brasileira, dando pesos diferentes a cada fator definido.

## II. TRABALHOS RELACIONADOS

Por se tratar de um tema de alta relevância, diversos estudos e pesquisas foram realizadas visando entender a vulnerabilidade de cidades. Ressalta-se aqui, em especial, o Índice de

Vulnerabilidade Municipal ao alastramento do Coronavírus (IVC), calculado pela Fundação Perseu Abramo. Esse índice é focado em identificar o quão vulnerável uma cidade estaria em relação ao alastramento do COVID-19, e foi desenvolvido no mês de Abril de 2020.

## III. PROCESSAMENTO DE DADOS

O primeiro processamento realizado foi o ajuste na taxa de mortalidade de acordo com a idade dos óbitos, como sugerido pela Organização Mundial da Saúde (OMS). Para tal, foram coletados os dados de mortes fornecidos pelo SIVEP-Gripe (mortes) e pelo Censo (dados demográficos). Calculou-se o número total de mortes em cada faixa etária (de 5 em 5 anos) para cada município, multiplicado pela constante definida pela OMS e dividido pela população de cada faixa.

$$mortalidade(faixa) = \frac{mortes(faixa) * cte(faixa)}{pop(faixa)}$$

Para a mortalidade do município foi somado as mortalidade de todas as faixas.

$$mortalidade\_municip = \sum_{i=0}^{16} mortalidade(i)$$

Após, foram coletados os dados referentes ao COVID-19 (por exemplo, casos), provenientes do *Brasil.io*. Primeiramente foi feito tratamento desses dados, como remoção de linhas com dados faltantes e a filtragem para utilizar apenas dados de municípios, e não de estados e regiões. Também foram coletados dados hospitalares, provenientes da plataforma *Elastic.search*, que contém os dados da disponibilidade hospitalar dos municípios bem como sua ocupação no atual momento. Como diversas cidades têm mais um hospital, os dados tiveram

que ser agregados por cidade (por meio da função *GroupBy*). Após isso, foi feito o ajuste de acordo com a população correspondente. Isso foi necessário pois grandes centros apresentam uma grande população e altos números de leitos e internações. Para a ocupação hospitalar, os dados de internação foram substituídos pelo percentual de ocupação. Através desse ajuste, os dados podem ser comparados diretamente entre si.

Feito isso, vários dados devem ser levados em conta na análise. É possível dividir em duas partes os dados hospitalares: Oferta e Ocupação. Dentro da oferta, encontram-se o número de respiradores e os 4 tipos leitos: UTI, CLI, UTI para SRAG e CLI para SRAG. Já na Ocupação, estão a ocupação de cada leito citado anteriormente. Visto que há uma grande quantidade de dados em cada grupo, foi utilizado a técnica de PCA (*Principal Component Analysis*) visando a diminuição da dimensão dos dados. Ao realizar o procedimento, o grupo da Oferta (hospitais) foi reduzido a uma componente (que atingia mais de 90% da variância), enquanto que o grupo de Ocupação foi reduzida a duas componentes, representando cerca de 80% da incerteza.

O próximo *dataset* a ser utilizado foi o IVC, calculado pela Fundação Perseu Abramo. Também foram lidos os dados de mortalidade gerados anteriormente e os códigos do IBGE (uma vez que até certo período o código utilizado tinha 6 dígitos, e que passou a ser 7 recentemente). Todos os dados presentes foram agregados em um único DataFrame, deixando o mesmo pronto para a aplicação do cálculo do índice.

Após a união de todos os dados, foi feita a normalização deles para que todos fiquem entre -1 e 1. Para tal foi utilizando a função *MaxAbsScaler*. Essa normalização de todos os dados é importante visto que muitos deles se encontram em escalas diferentes.

Em seguida, foi realizado a divisão de cidades em grupos, utilizando como critério a taxa de mortalidade, o IVC, a oferta dos hospitais, e os dois dados referentes a ocupação (esses 3 últimos obtidos via PCA). Para a realização da clusterização, foi utilizado o método *K-means*. Primeiro identificou-se qual a melhor quantidade de grupos a ser utilizada. Para tal, foi calculado o erro quadrático das distâncias de 1 até 15 grupos, como mostrador na Figura 1. Ao analisá-la, o grupo definiu utilizar 4 grupos como o valor mais adequado, visto que a utilização de 2 *clusters* (que apresenta a menor queda para o anterior) foi considerado pouco e de 3 para 4 foi o que apresentou uma grande queda.

Após a definição, foi realizado de maneira efetiva o *K-means* utilizando 4 grupos para a clusterização. Na Figura 2 está presente o resultado após a procedimento. Entretanto, ele é um gráfico que apresenta apenas 2 das 5 variáveis utilizadas, não sendo possível visualizar todas as dimensões do mesmo.

Uma vez dividido os grupos, fez-se necessário entender quais membros desse grupo eram considerados *outliers*. A apuração desses dados foi feita por meio do algoritmo *Local Outlier Factor*, que é um método de detecção de anormalidades não supervisionado que computa a diferença de densidade local de um determinado ponto de dados em relação aos seus vizinhos mais próximos. São considerados *outliers* as

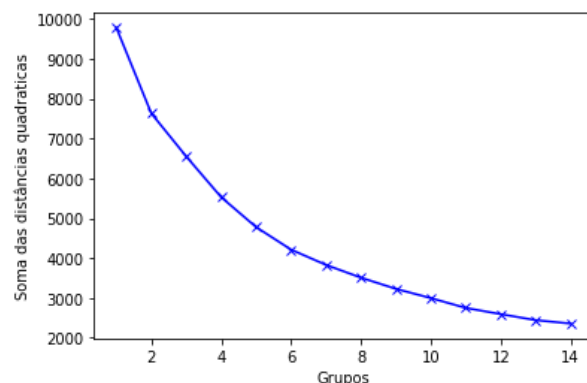


Figura 1. Erro quadrático para cada quantidade de grupos no K-means.

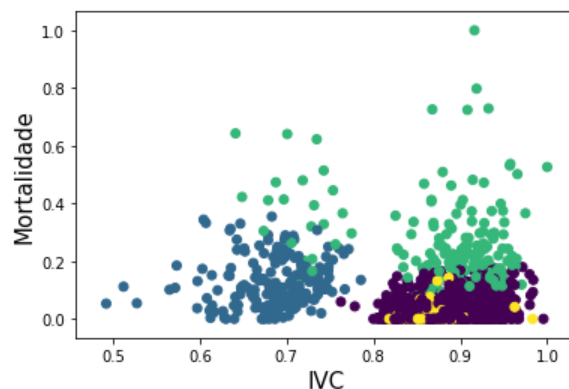


Figura 2. Resultado da aplicação do K-means nas cidades.

amostras que apresentam uma densidade significativamente menor que seus vizinhos. Uma vez que os *outliers* foram encontrados, foi possível identificar determinados comportamentos nos dados (por exemplo, a alta vulnerabilidade de cidades demasiadamente populosas) e posteriormente tratar esses dados de maneira diferente no cálculo do índice.

Para o cálculo final, foi utilizado a fórmula:

$$\begin{aligned} \text{Índice Final} = & (\text{Taxa de mortalidade} \times 0.3 \\ & - \text{IVC} \times 0.1 - \text{Oferta Hospitais} \times 0.1 \\ & + (\text{Ocupação1} + \text{Ocupação2}) \times 0.05 \\ & + \text{População Idosa Percentual} \times 0.1 \\ & + \text{População Total} \times 0.25) \times \text{Grupo.Outlier}(1) \end{aligned}$$

Para a escolha dos pesos, foi levado em conta a importância de cada fator para o aumento da vulnerabilidade daquela determinada cidade. Tendo em vista que o foco do projeto é a vulnerabilidade das cidades cujo principal fator agravante é a mortalidade, esta foi considerada como elemento de maior relevância. Logo, o peso escolhido para esse foi de 0.3, se configurando como o elemento mais expressivo do índice.

O IVC foi escolhido para compor o cálculo pelo fato de agregar informações demográficas e socio-econômicas dos

municípios brasileiros. Entende-se que essas informações são importantes pois a vulnerabilidade de uma população está estreitamente ligada com a sua condições de vida. Deduz-se que uma cidade com baixo IDH se inclina a ser mais vulnerável ao COVID-19. O peso escolhido foi de 0.1 para esse fator.

Outro aspecto que foi considerado no cálculo do índice foi a estrutura hospitalar do município, que leva em conta o número do leitos clínicos normais e específicos para COVID-19 e a quantidade de leitos de UTI. Julga-se que esse fator seja relevante pois por meio dele consegue-se inferir a capacidade do município a responder aos casos de COVID-19 e quanto maior o número dessas instalações, melhor é a resposta do município em relação a doença. O peso definido para essa categoria foi 0.1. Da mesma forma, foram consideradas as taxas de ocupação desses leitos, tendo 0.05 como peso, uma vez que os dois índices gerados pelo PCA foram somados.

Também foi levado em consideração a porcentagem de população idosa do município para a composição do índice, uma vez que o risco de morte por COVID-19 aumenta significativamente em pessoa com idade elevada. Por consequência, foi definido que o peso para esse fator fosse 0.1.

Outro fator importante para a composição do índice é o tamanho da população do município, uma vez que a mortalidade e taxa de contaminação aumenta deveras em cidades com muitas pessoas. Cidades como São Paulo e Rio de Janeiro (as duas maiores cidades brasileiras) foram imensamente afetadas pelo COVID-19. Tendo em vista os fatos supracitados, o peso escolhido para esse fator foi de 2.5, ligeiramente menor que o de mortalidade pois levamos em conta que o foco do índice de vulnerabilidade gerado seja utilizado na vacinação visando reduzir a mortalidade.

De maneira geral, é possível afirmar que o índice é composto por dois grandes fatores (mortalidade e população) e quatro fatores importantes porém com menor relevância (número de leitos, ocupação, população idosa e IVC).

Além desses fatores, o índice ainda é composto por um fator de multiplicação que varia de acordo com o grupo em que o município se encontra dentro do agrupamento por vulnerabilidade (calculado pelo K-Means) e se esse município é um outlier ou não dentro desse grupo. Ou seja, se um município se encontra no Grupo 1 de vulnerabilidade (alto risco) e ainda é um outlier dentro desse grupo, o índice é multiplicado por 1.3, de forma que o valor aumente. Se por outro lado o município pertencer ao grupo de risco 4 (baixo risco) e índice é multiplicado por 0.75. Os fatores de multiplicação para cada grupo pode ser visto na Tabela I.

#### IV. RESULTADOS

Através da análise exploratória dos dados coletados e de pesquisas e leituras de artigos relacionados ao assunto, foi possível responder a primeira pergunta de pesquisa: "Quais fatores influenciam na vulnerabilidade das cidades brasileiras em relação ao COVID-19?". Foi considerado que o principal fator era a taxa de mortalidade de cada cidade. Realizando a análise da correlação com a mortalidade, foi encontrado que

Tabela I  
TABELA DE GRUPOS E SEUS VALORES

Tipo	Valores		
	<i>Outlier Negativo</i>	<i>Normal</i>	<i>Outlier Positivo</i>
Nível 1	1.20	1.25	1.30
Nível 2	1.05	1.10	1.15
Nível 3	0.90	0.95	1.00
Nível 4	0.70	0.75	0.80

<sup>a</sup>Tabela de valores de ajuste de acordo com o grupo.

indicadores como o IVC e a ocupação hospitalar tem uma correlação relevante (-0.197 e 0.227 respectivamente).

Outros fatores encontrados foram:

- **Oferta de hospitais:** Pois consiste na capacidade de contenção da crise de saúde;
- **Ocupação hospitalar:** Pois representa a situação atual dos hospitais;
- **População idosa:** Pois consiste em um dos principais grupos de risco, e deve ser o foco da vacinação quando o objetivo é evitar mortes, como citado no artigo do MIT<sup>1</sup>;
- **População total:** Pois além de ser o fator que pode indicar maior contágio (pensando no número de pessoas), também ajuda a representar fatores como locomoção e localização.

Já para a segunda pergunta, "Quais cidades apresentam maior vulnerabilidade em termos dos impactos na saúde referente ao coronavírus?", foi desenvolvido o cálculo do índice, como indicado pela Equação 1, e aplicado para cada uma das cidades brasileiras correspondentes. Após, elas foram ordenadas do maior para o menor índice. Na Figura 3 é possível observar as 10 cidades com o maior índice de vulnerabilidade (e consequentemente, uma maior prioridade na vacinação). A que apresentou o maior índice foi a cidade de Rio Claro, no Rio de Janeiro. Destaca-se também a presença de duas capitais na lista, que são São Paulo e Rio de Janeiro.

	UF	Nome do município	Vulnerabilidade Geral
791	RJ	Rio Claro	0.298157
1334	SP	Pariquera-Açu	0.294387
1410	SP	São Paulo	0.272736
619	PE	Trindade	0.217289
670	PR	Campo Largo	0.198541
792	RJ	Rio de Janeiro	0.187121
669	PR	Campina Grande do Sul	0.178680
1458	TO	Augustinópolis	0.177420
181	CE	Sobral	0.176774
514	PB	Santa Rita	0.169902

Figura 3. Top 10 com os maiores índices de vulnerabilidade.

<sup>1</sup>Disponível em: <https://www.technologyreview.com/2020/11/20/1012313/who-should-get-a-covid-19-vaccine-first/>

Por outro lado, observa-se na Figura 4 as 10 cidades com os menores índices, tendo a cidade de Chapadão do Céu como o menor índice. Vale observar o fato de que a maioria absoluta das cidades presentes nas com 10 menores vulnerabilidades são da região centro-oeste.

	UF	Nome do município	Vulnerabilidade Geral
446	MT	Colniza	-0.075790
37	PA	Santana do Araguaia	-0.077214
737	PR	São José dos Pinhais	-0.078344
1908	MS	Nova Alvorada do Sul	-0.078405
371	GO	Santo Antônio do Descoberto	-0.079316
1916	MS	Ribas do Rio Pardo	-0.079481
430	MT	Alto Taquari	-0.080525
1861	MS	Água Clara	-0.082440
468	MT	Nova Maringá	-0.084469
258	GO	Chapadão do Céu	-0.091174

Figura 4. Top 10 com os menores índices de vulnerabilidade.

Esse índice gerado apresentou uma correlação de 0.378 com o IVC. Através da Figura 5 é possível ver o histograma da frequência dos índices. É possível perceber que apesar de conterem valores e faixas distintas, apresentam histogramas semelhantes.

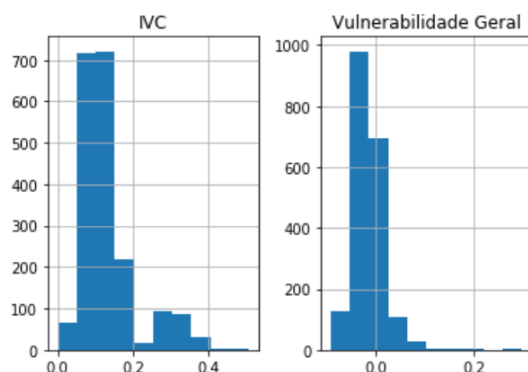


Figura 5. Comparação das distribuições de vulnerabilidade do IVC e do atual trabalho.

## V. LIMITAÇÕES/TRABALHOS FUTUROS

### A. Limitações

Diante da complexidade do desenvolvimento de um índice que corresponda a realidade dos fatos, inúmeros fatores são limitadores, fazendo com que o trabalho desenvolvido contenha inúmeros pontos que devem ser ressaltados. Um dos principais fatos é a discrepância dos dados demográficos atuais com os disponíveis, uma vez que o último Censo realizado foi no ano de 2010. Essa década de diferença faz com que dados desatualizados fossem utilizados para a análise.

Uma limitação adicional é a consideração de cidades de maneira individual, ignorando a região metropolitana em que

se encontra. Por exemplo, São José dos Pinhais está contido na mesma região metropolitana que Campo Largo (Curitiba). Entretanto, pode-se observar que estão em posições opostas no índice de vulnerabilidade gerado (vide f Figuras 3 e 4).

A utilização do índice para vida real seria questionável devida a alta complexidade para a geração do mesmo, tendo em vista o uso do PCA seguido de uma clusterização e do uso do LOF. Tais técnicas não permitem que esse índice seja facilmente replicado.

Outra limitação do presente trabalho é a simplificação dos indicadores utilizados para a construção do índice. Nele foram consideradas apenas cidades com unidade hospitalar (menos de 3000 das 5000 cidades brasileiras têm pelo menos um hospital). Além disso, ao considerar uma unidade hospitalar, foi considerado que ela é responsável apenas pela população da cidade que tem sede o mesmo, desconsiderando o fato de que inúmeras cidades utilizam os hospitais de cidades vizinhas.

Outras simplificações são as variáveis atribuídas ao cálculo do índice final, que são limitadas e não corresponde a todos os pontos necessários para se analisar a vulnerabilidade de maneira realmente efetiva. Dados como mobilidade foram "substituídos" pelo tamanho da população. Além disso, os pesos de cada indicador também são imperfeitos, sendo necessária uma revisão profunda de especialistas da área de medicina.

### B. Trabalhos Futuros

Ante as limitações citadas anteriormente, os principais trabalhos futuros consistem na correção de algumas limitações. A primeira delas seria acrescentar ao índice variáveis que expressem mais informações relevantes sobre cada cidade e dentre elas pode-se citar dados de localização e locomoção. Além disso, seria interessante realizar a revisão e correção dos pesos de cada indicador na fórmula, uma vez que o método escolhido para a definição dos mesmo é imperfeito.

Outro ponto de melhoria seria a simplificação do cálculo do índice, substituindo os algoritmos complexos utilizados por pares mais simples, tais como a regressão para a identificação de *outliers*.