



CIÊNCIA DE DADOS: FERRAMENTAS PARA O COMBATE À COVID-19

ARTIGO ORIGINAL

BRITO, Arlindo Matheus Santiago de ¹

BRITO, Arlindo Matheus Santiago de. **Ciência de dados: Ferramentas para o combate à Covid-19.** Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano 05, Ed. 06, Vol. 09, pp. 135-147. Junho de 2020. ISSN: 2448-0959, Link de acesso: <https://www.nucleodoconhecimento.com.br/tecnologia/ciencia-de-dados>

RESUMO

A Covid-19 é uma doença respiratória causada pelo vírus Sars-cov-2. Por se tratar de uma nova doença causada por um novo vírus, os cientistas passam pelo maior desafio deste século, na busca por uma cura, para tal enfermidade. A pandemia mudou totalmente o cotidiano das pessoas. Diferente de pandemias anteriores, o acesso às informações agora é cada vez mais simplificado, cada pessoa pode se informar e tomar para si os dados sobre a Covid-19. Entretanto, é necessário que esses dados e notícias sejam filtrados, pois além do avanço da doença, têm-se visto o avanço das notícias falsas e isso se torna muito perigoso, não apenas pelo momento crítico vivido, mas também o descrédito aos veículos de informação que trabalham com seriedade. Através dessas questões esse estudo objetiva indicar ferramentas para análise de dados que permitam avaliar o avanço da pandemia, além de apresentar uma ferramenta que reúne informações e orientações com embasamento científico para o combate ao Coronavírus. Os estudos realizados podem apresentar informações úteis para análise do comportamento da pandemia no Brasil e dados de previsão do avanço

¹ Licenciado em Ciências Biológicas, Mestrando em Ciências Ambientais do Campus Reitor Edgar Santos Universidade Federal do Oeste da Bahia (UFOB).



da curva de casos da doença, contribuindo para o planejamento de medidas para o combate ao avanço do Sars-cov-2 no país.

Palavras-chave: Coronavírus, pytho, glide, informações, tecnologia.

1. INTRODUÇÃO

Em 2019 o que parecia ser uma nova gripe na China, se confirmou como um dos maiores desafios do nosso tempo. Em um curto período a Covid-19 se espalhou por todos os continentes e em março de 2020 foi descrita pela Organização Mundial de Saúde como uma pandemia mundial (OMS/WHO, 2020). Em um momento de transformações tecnológicas e sociais, o Sars-cov-2 (Novo Coronavírus), foi capaz de expor grandes falhas e necessidades dos sistemas de saúde e questões socioeconômicas de cada país (FERNANDES, 2020). A grande velocidade de contaminações tem mudado o comportamento das pessoas pelo mundo. A principal recomendação da OMS, é o isolamento social, em muitos casos o fechamento quase por completo das atividades profissionais e acadêmicas, o chamado *Lockdown*[2].

Nos últimos anos o consumo de dados se intensificou muito além das expectativas (CASTELLS, 2010), com tantas informações que são transmitidas, compartilhadas e encaminhadas todos os dias, existe a necessidade de filtragens ainda maiores, para compreender se o que é publicado, é realmente verdadeiro. A pandemia do novo coronavírus e as medidas de isolamento social são acompanhadas de um número desmedido de informações sobre o tema. Com isso crescem também as notícias falsas, que nos últimos anos passaram a se tornar algo corriqueiro. Não é recente que a ciência passa por um processo de escassez, mas isso tem se intensificado cada vez mais, justamente pela crescente aceitação das chamadas *fake news*, esse processo em meio a uma Pandemia se torna algo perigoso e cruel.

Um estudo realizado pela Kaspersky®, empresa famosa no ramo da cibersegurança, mostra que cerca de 62% da população brasileira não consegue identificar notícias falsas (RODRIGUES, 2020). Esse número é preocupante, pois torna essas pessoas vulneráveis a crimes cibernéticos ou até mesmo ao próprio coronavírus. Com o intenso



e cada vez maior consumo de dados diariamente, ultrapassando a capacidade humana de analisar tais informações por conta própria, uma nova e importante área surgiu, a chamada Ciência de Dados, que através de linguagens de programação e softwares, consegue traduzir graficamente os dados/informações consumidos no mundo. Através da Ciência de Dados baseada em modelos matemáticos de previsão, o enfrentamento da pandemia se torna mais claro, pois ferramentas como mapas e gráficos de localidades mais acometidas pela doença podem facilitar o recebimento de tratamentos mais intensivos e eficientes (CHERIFA e PIRRACCHIO, 2019). Tendo em vista a necessidade de contribuições científicas para o enfrentamento da pandemia e ao avanço das notícias falsas, o objetivo desta pesquisa é apresentar algumas ferramentas para análise dados e uma central informações baseada em artigos científicos de grandes instituições acadêmicas do mundo. A informação científica é atualmente a principal arma contra essa pandemia.

2. FERRAMENTAS PARA ANÁLISE

2.1 OBTENÇÃO DOS DADOS

Os coronavírus pertencem a uma família de vírus (Coronaviridae) que causam doenças tanto em animais quanto em humanos, sendo que nos seres humanos causam infecções respiratórias, que variam desde um resfriado a síndromes respiratórias mais graves. O Sars-cov-2, conhecido como novo coronavírus foi descoberto no fim de 2019, os primeiros contaminados estiveram em contato pela vez no mercado de peixes em Hubei, província de Wuhan na China. Os coronavírus são RNA vírus de filamento único, isolados em humanos pela primeira vez em 1937. Mas apenas em 1965, o vírus foi descrito como coronavírus, em resultado do seu perfil microscópico, semelhando uma coroa (CSGICTV, 2020). De certo a pandemia faz com que haja questionamentos sobre os números de óbitos, casos confirmados e também recuperações da enfermidade. Esses números são disponibilizados e contabilizados pelas secretarias estaduais e municipais de saúde, além de uma contagem feita pelo Ministério da Saúde. Por uma questão de transparência por parte do poder público essas informações devem ser claras para a população.



Para o tratamento dos dados, utilizou-se (linguagem) Python que possui uma sintaxe simples e, ao mesmo tempo completa para a análise de dados, o que possibilita um alto índice de confiança e rapidez na execução das análises (GORELICK e OZSVALD, 2020). Agregando bibliotecas, a linguagem cria um ambiente de estudos científicos completo em um simples computador. Python possui uma comunidade bastante ativa, que contribui para a criação de um ecossistema bastante estável para a construção de projetos.

2.2 BIBLIOTECAS

As bibliotecas são pacotes, que possuem programações pré-carregadas que auxiliam o programador ao desenvolver um software, permitindo atribuir mais ferramentas ao código, através de poucas implementações. No caso da análise de dados, algumas bibliotecas são essenciais para a visualização de dados, as utilizadas nessa pesquisa foram:

2.2.1 PANDAS

O Pandas é uma biblioteca do Python utilizada para análise e manipulação de dados. A biblioteca trabalha com duas maneiras para apresentar os dados, Dataframe e Series.

- Dataframe: tem estrutura tabular, onde coleta os dados e os converte em tabela com as colunas sendo os campos dos dados e as linhas como registros de dados.
- Series: tem estrutura unidirecional acompanhada de um índice, ao contrário do Dataframe as series possuem apenas uma coluna que reúnem, uma série de dados em sequência com um índice que vai de 0 (zero) ao número total de registros



2.2.2 NUMPY

O NumPy é o pacote básico da linguagem Python com o qual é possível trabalhar arranjos, vetores e matrizes. Garante diversas funções e operações matemáticas, incluindo:

- Objeto *array* para a execução de arranjos multidimensionais;
- Objeto *matrix* para o cálculo com matrizes com N dimensões;
- Ferramentas matemáticas para álgebra linear;
- Ferramentas estatísticas.

2.2.3 MATPLOTLIB

Matplotlib é uma biblioteca para criação de gráficos e visualizações de dados, através do Matplotlib é possível gerar vários tipos de gráficos como:

- Gráficos de linhas;
- Gráficos de barras;
- Gráficos de dispersão;
- Gráficos de pizza;
- Gráficos em 3d;
- Gráficos personalizados;
- Funções matemáticas.

2.2.4 FOLIUM

Folium é uma biblioteca que produz visualizações de mapas a partir da biblioteca JavaScript Leaflet.js tornando mais simples a manipulação de dados e visualização em mapas. Através dessa biblioteca podem ser desenvolvidos:

- Mapas coropléticos;
- Mapas de calor;
- Mapas com animações de tempo.



2.2.5 PROPHET

O Prophet é um pacote para R e Python desenvolvido pelo Facebook. Ele executa um algoritmo de previsão de séries temporais, para encontrar padrões sazonais de uma série de entrada. Somado ao Matplotlib gera gráficos de previsão, com acurácia de 95% de certeza.

2.3 APLICATIVO PWA

Os *progressive web apps* (PWA), são aplicações que utilizam ferramentas nativas de celulares, mas são emulados a partir do navegador (REINKE, 2020). A plataforma utilizada foi o Glide App, uma ferramenta que não utiliza linguagem de programação para desenvolver aplicativos, mas apenas uma planilha do Google. O Glide permite a elaboração de aplicativos completos apenas com os dados contidos na planilha, cada célula da planilha representam um conteúdo que é apresentado no aplicativo.

Instituições por todo o mundo lançam inúmeros artigos científicos relevantes para o avanço do combate do novo coronavírus, reunir essas informações em um único lugar, permite que o público geral tenha um arcabouço de informações reais muito maiores, em relação a pandemia.

3. RESULTADOS

A matemática por ser uma ciência exata, permite que outras disciplinas utilizem seus métodos para enfrentar problemas e encontrar soluções lógicas. Não diferente disso a programação e mais especificamente a ciência de dados, tem como estrutura modelos matemáticos de previsão e estatística. Por ser um conhecimento interdisciplinar, a ciência de dados pode contribuir em qualquer situação.

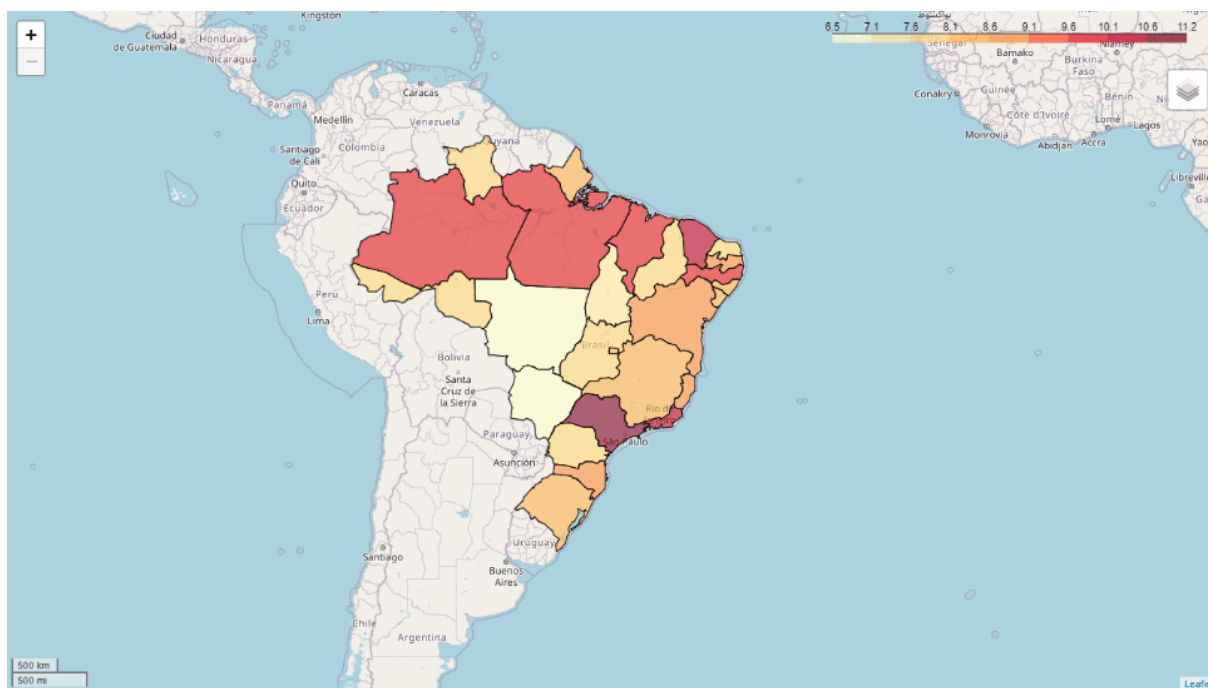
A Covid-19 é uma doença que tem alto índice de contaminações e com clara rapidez, vendo a necessidade de comparações, previsões e visualizações dos dados, partindo dos dados disponíveis no Brasil a pesquisa obteve os seguintes resultados.

3.1 MAPA EPIDEMIOLÓGICO

Os dados da epidemia são disponibilizados pelos governos Estaduais e Federal, tais informações atualizadas diariamente, são importantes para diagnosticar o desenvolvimento da doença nos estados (BRASIL, 2020). O mapa epidemiológico é uma maneira visual de explicar esses dados, como é possível observar na Figura 1.

Os estados de São Paulo, Rio de Janeiro, Amazonas, Ceará e Pará, possuem números de contaminações muito expressivo, enquanto que os estados de Mato Grosso, Mato Grosso do Sul e Tocantins possuem menores taxas de contaminações. Segundo a OMS, quanto antes medidas de distanciamento forem tomadas, além do uso de álcool em gel e máscaras, o número de contaminações é reduzido.

Figura 1 - Mapa Epidemiológico Covid-19



Fonte de dados: Secretarias de Saúde

A formula por trás da execução e organização das cores é uma função logarítmica, básica:

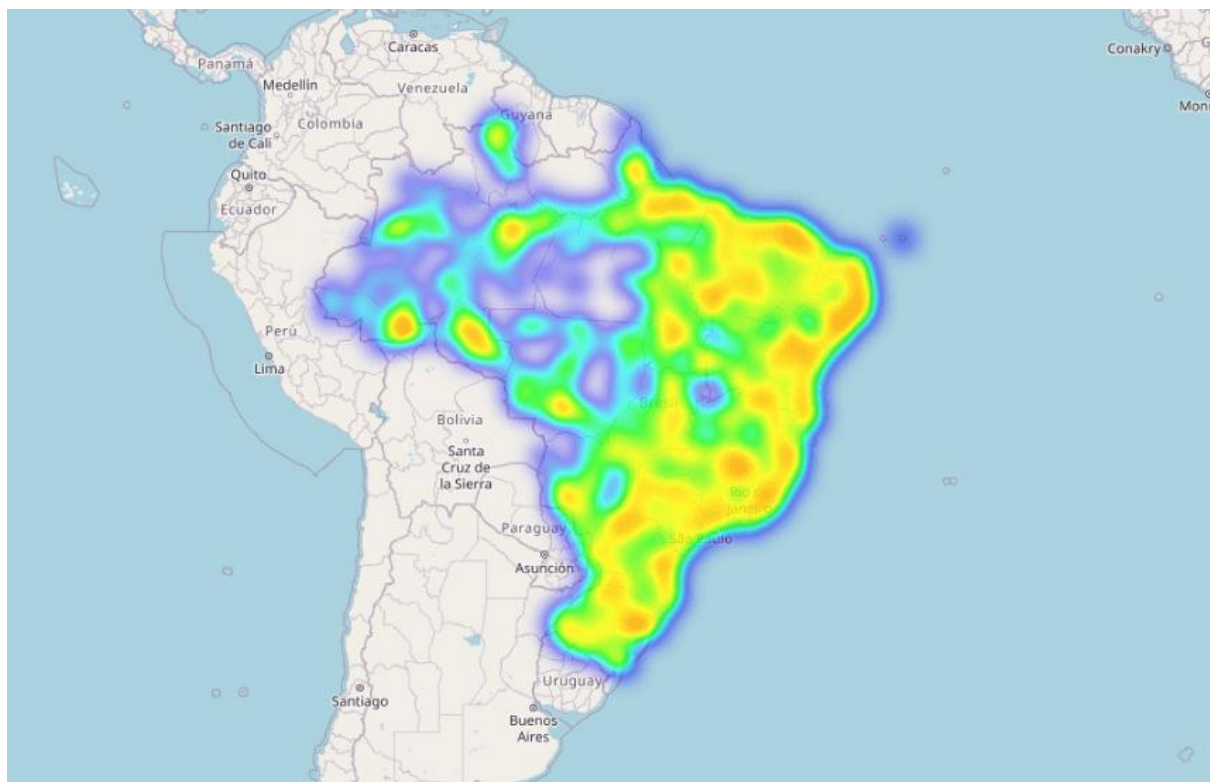
$$x = \text{Log}_a(b)$$

Onde x é o expoente da base a, então x é o logaritmo de b na base a.

3.2 MAPA DE CALOR

Assim como o mapa coroplético, o mapa de calor utiliza cores para representar dados mais e menos expressivos. Com o mapa de calor é possível enxergar as regiões mais afetadas, representadas por cores quentes, enquanto as regiões com menos casos de Covid-19 possuem cores mais frias. Diferente do mapa anterior que usava as fronteiras dos estados para expor os dados, o mapa de calor utiliza microrregiões com a mesma finalidade, representado na Figura 2.

Figura 2 - Mapa de Calor Covid-19 no Brasil

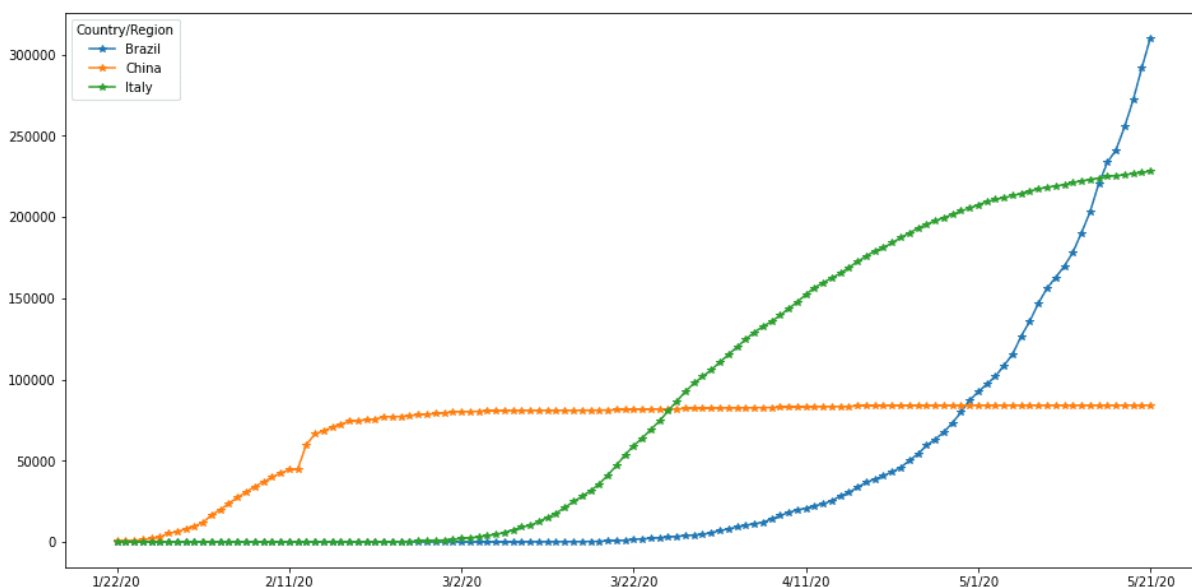


Fonte de dados: Secretarias de Saúde

3.3 GRÁFICOS COMPARATIVOS

A progressão do Sars-cov-2 foi semelhante em todos os países acometidos pela covid-19, comparar a curva de casos é importante para se chegar à dimensão do ponto em que se está da contaminação. Os casos acumulados têm subida exponencial e após a redução das contaminações se alcança o chamado platô, ou seja, a soma constante dos últimos casos acumulados. A China foi o primeiro país a chegar no platô, por conseguir isolar a epicentro da doença, mas países como o Brasil que permanecem com suas atividades normais, têm aumentado o crescimento de casos sem ter ainda uma real visualização do platô, o gráfico da Figura 3 é uma comparação de casos entre Brasil, China e Itália, os dois últimos países passaram pelo processo de aumento exponencial de casos e agora estão voltando as suas rotinas comerciais e pessoais. Esses gráficos são o resultado da aplicação das bibliotecas Pandas, Numpy e Matplotlib.

Figura 3 - Gráfico comparativo de casos acumulados no Brasil, Itália e China



Fonte de dados: Worldmeter

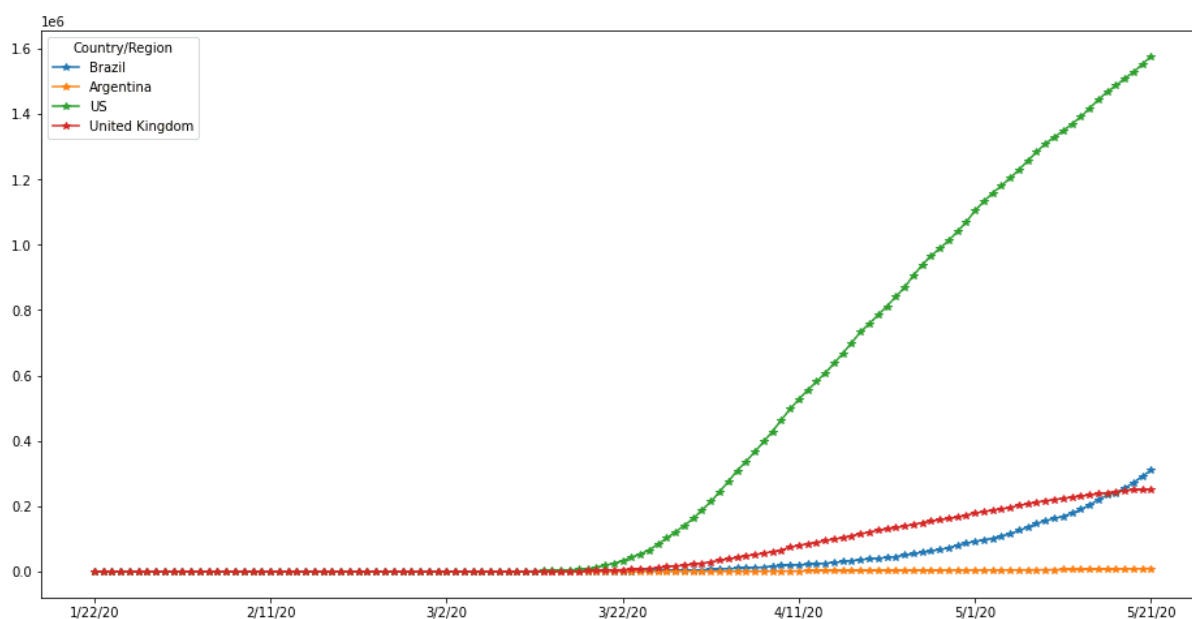
O eixo x desse gráfico é o avanço da doença desde o janeiro, o eixo y representa a quantidade casos. Nota-se que o Brasil, em maio ultrapassou a China e logo em seguida a Itália, esse dado é extremamente relevante, pois mostra como as



contaminações são muito rápidas. O Brasil em menos de um mês passou de 100.000 para 300.000 casos.

Em outro comparativo, na Figura 4 entre Brasil, Estados Unidos, Reino Unido e Argentina, podem ser observadas como as recomendações da OMS surtem efeito quando são feitas e tem resultados significativos quando tomadas rapidamente. A Argentina e o Reino Unido tiveram resultados positivos pois assumiram a necessidade do distanciamento social, enquanto os Estados Unidos e o Brasil, ainda não tiveram por parte dos seus governantes um ações efetivas quanto ao avanço da doença.

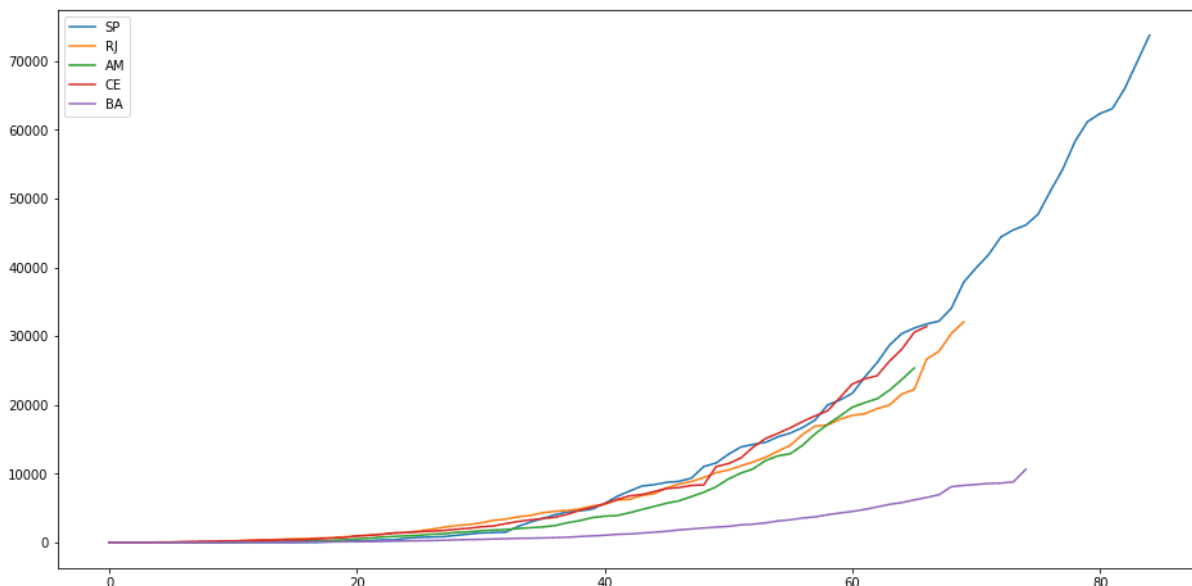
Figura 4 - Gráfico comparativo Brasil, Estados Unidos, Reino Unido e Argentina



Fonte de dados: Worldmeter

Em uma escala local, a comparação entre os estados também mostra como as ações de combate a pandemia são eficazes. O primeiro caso ocorreu no estado de São Paulo e logo se espalhou por todos os estados. Contudo, a população brasileira permanece cética sobre o avanço da pandemia, em SP o número de pessoas contaminadas é significativo, tendo em vista a proporção de habitantes em comparação com outros estados, na Figura 5, o gráfico compara os estados de São Paulo, Amazonas, Rio de Janeiro, Ceará e Bahia.

Figura 5 - Gráfico comparativo Estados brasileiros



Fonte de dados: Secretarias de Saúde

O Brasil é um país com dimensões continentais, conhecido por sua grande diversidade cultural e biológica, mas possui um dos piores índices em questões de saneamento básico, educação básica. Esses fatores são favoráveis para o avanço da covid-19 no país. No Rio de Janeiro, o portal Voz das Comunidades (2020), organizou um painel com dados da covid-19 apenas nas favelas, os dados obtidos mostram o avanço da doença, 2% de todos os casos do Rio de Janeiro está localizado nas comunidades.

3.4 GRÁFICOS DE PREVISÃO

A partir de dados temporais e a crescente exponencial da covid-19, desenvolver gráficos de previsão, os modelos matemáticos atribuídos, são embasados em regressão linear. Alguns estudos possibilitam mensurar a capacidade das UTIs com o decorrer da pandemia, para evitar o colapso no SUS (FIDALGO, 2020).

O gráfico de previsão utiliza uma margem de erro em seus dados, utilizando a biblioteca Prophet, as datas (ds), a menor margem (yhat_lower), a mediana dos dados (yhat) e a maior margem (yhat_upper), como é possível ver na Tabela 1.



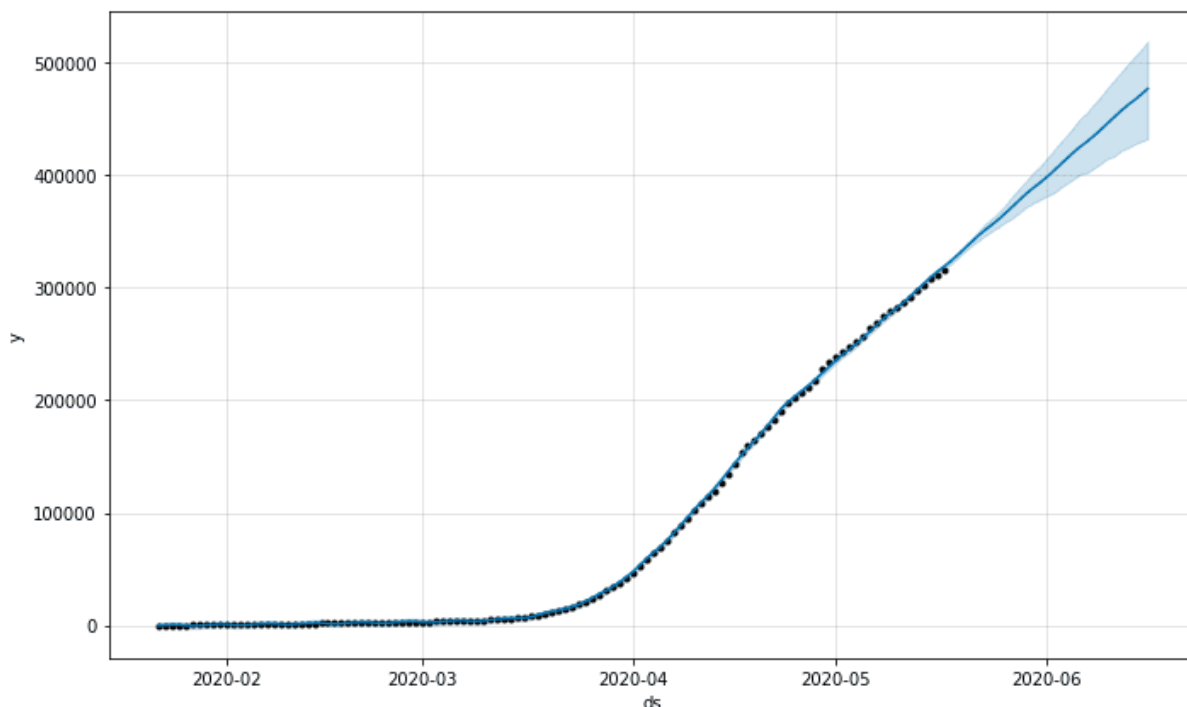
Tabela 1 - Previsão por Covid-19 através da biblioteca Prophet

ds	yhat_lower	yhat	yhat_upper
01/06/2020	381630,1891	397873,7434	414286,4328
02/06/2020	383988,723	403258,7465	421360,8957
03/06/2020	388493,6658	409031,2021	428786,2591
04/06/2020	392865,7724	414579,2587	435410,971
05/06/2020	396657,6343	420161,9484	442570,6225
06/06/2020	400946,1081	425280,3656	450389,0583
07/06/2020	402296,9173	429847,6031	455818,1056
08/06/2020	406431,2319	434783,2885	463054,3009
09/06/2020	410359,9382	440168,2916	469691,7447
10/06/2020	415040,2623	445940,7472	477478,2626
11/06/2020	416789,0184	451488,8039	484284,4346
12/06/2020	421820,7204	457071,4936	491655,2796
13/06/2020	424461,927	462189,9107	498549,165
14/06/2020	427672,3128	466757,1482	505309,4528
15/06/2020	430238,8531	471692,8337	511796,4759
16/06/2020	432655,2221	477077,8368	519328,771

Fonte de dados: CSSE Jonhs Hopkins



Figura 6 - Previsão por Covid-19 através da biblioteca Prophet



Fonte de dados: CSSE Jonhs Hopkins

Partindo do histórico de óbitos por covid-19 no mundo, o Prophet oferece essa tabela, como é possível observar até os meados do mês de junho a previsão é de cerca de 477.077 óbitos no mundo, com a margem entre 432.655 a 519.328 óbitos. O gráfico, apresenta esses dados de uma forma mais simples.

3.5 QUARENTENA INFORMADA

Ocorrendo a necessidade de uma aplicação que concentre informações e sobre as pesquisas em desenvolvimento e que ao mesmo tempo pudesse informar o público em geral fora do âmbito acadêmico, foi desenvolvido o aplicativo PWA 'Quarentena Informada[3]'. Através da plataforma Glide App, o aplicativo foi construído com três pilares: a busca por informações com embasamento científico, dados reais sobre a pandemia e orientações em acordo com as apresentadas pela OMS. Os aplicativos passaram a fazer parte do nosso cotidiano e levar informação através dessas tecnologias ajuda a propagar o conhecimento científico.



O aplicativo é uma ferramenta aberta e gratuita, com cerca de 300 usuários é atualizado diariamente com gráficos contendo informações de casos confirmados, óbitos confirmados e pessoas recuperadas da covid-19. O aplicativo passa por um processo de aplicação de ferramentas para pessoas com deficiência visual, através de um assistente de voz.

4. CONSIDERAÇÕES FINAIS

A pandemia do novo coronavírus é o principal desafio de uma geração, pelo menos nos últimos 50 anos, uma doença que parou as principais economias mundiais e ao mesmo tempo tornou evidente as falhas nos setores econômicos, sanitários e sociais. Dando início também a uma revolução tecnológica, tornando cada pessoa mais próxima do ambiente digital. Mas é preciso ter responsabilidade para prosseguir, sob um envolvimento de informações.

A informação atualmente é o mecanismo mais poderoso, com a ajuda da ciência de dados, esses elementos passam a ser mais claros. Apresentar ferramentas é tão importante quanto apresentar os dados em si, pois ao tornar algo mais próximo do público geral, criando um vínculo entre a sociedade e a ciência, novas ferramentas surgirão e é preciso estar preparado para usá-las da maneira correta e com um objetivo efetivo.

REFERÊNCIAS

BRASIL. Coronavírus - SUS. **Governo do Brasil**, 2020. Disponível em: <<https://www.gov.br/pt-br/apps/coronavirus-sus>>. Acesso em: 10 Março 2020.

CASTELLS, M. **Fim de milênio - A Era da Informação**. 3°. ed. São Paulo: Paz e Terra, v. I, 2010.

CHERIFA, M.; PIRRACCHIO, R. O que todo intensivista deveria saber sobre Big Data e aprendizado da máquina na unidade de terapia intensiva. **Revista Brasileira de Terapia Intensiva**, São Paulo, 20 Janeiro 2019. 3.



CSGICTV. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **Nature Microbiology**, Maryland, 2 Março 2020. 9.

FERNANDES, N. Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy. **SSRN Website**, 2020. Disponível em: <<https://dx.doi.org/10.2139/ssrn.3557504>>. Acesso em: 27 Março 2020.

FIDALGO, M. Como a ciência de dados vem ajudando na luta contra a COVID-19. **Criteo**, 2020. Disponível em: <<https://www.criteo.com/br/insights/como-a-ciencia-de-dados-vem-ajudando-na-luta-contra-o-covid-19/>>. Acesso em: 25 Abril 2020.

GORELICK, M.; OZSVALD, I. **High Performance Python: Practical Performant Programming for Humans**. 2º. ed. Sebastopol: O'Reilly Media, v. I, 2020.

OMS/WHO. WHO announces COVID-19 outbreak a pandemic. **World Health Organization Europe**, Genebra, p. 20, 2020. Disponível em: <<http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>>. Acesso em: 15 Março 2020.

REINKE, F. L. Startup para eventos: tecnologia, relacionamento e gamificação como estratégia para a organização de eventos. **Biblioteca Digital Unijuí**, Ijuí, 16 abr. 2020. 22.

RODRIGUES, R. Mais de 60% dos brasileiros não sabem reconhecer notícia falsa. **Kaspersky Daily**, 2020. Disponível em: <<https://www.kaspersky.com.br/blog/fake-news-brasil-pesquisa/14060/>>. Acesso em: 5 Março 2020.

APÊNDICE - REFERÊNCIAS DE NOTA DE RODAPÉ

2. São protocolos de isolamento que geralmente impedem que pessoas, informações ou carga deixem uma área.



MULTIDISCIPLINARY SCIENTIFIC JOURNAL

**NÚCLEO DO
CONHECIMENTO**

REVISTA CIENTÍFICA MULTIDISCIPLINAR NÚCLEO DO
CONHECIMENTO ISSN: 2448-0959

<https://www.nucleodoconhecimento.com.br>

3. Disponível em: <https://quarentenainformada.glideapp.io>

Enviado: Maio, 2020.

Aprovado: Junho, 2020.

RC: 52582

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/ciencia-de-dados>