

Plataforma de Monitoramento Inteligente da COVID-19

COVID-19 Intelligent Monitoring Platform

Plataforma de Monitoreo Inteligente COVID-19

Recebido: 18/10/2022 | Revisado: 29/10/2022 | Aceitado: 01/11/2022 | Publicado: 07/11/2022

Arlete Teresinha Beuren

ORCID: <https://orcid.org/0000-0001-7565-6184>
Universidade Tecnológica Federal do Paraná, Brasil
E-mail: arletebeuren@utfpr.edu.br

Hugo José Teixeira de Freitas

ORCID: <https://orcid.org/0000-0002-3690-7208>
Universidade Tecnológica Federal do Paraná, Brasil
E-mail: hugofreitas@alunos.utfpr.edu.br

Thiago França Naves

ORCID: <https://orcid.org/0000-0002-3152-1197>
Universidade Tecnológica Federal do Paraná, Brasil
E-mail: naves@utfpr.edu.br

Hugo Andrés Ruiz Flórez

ORCID: <https://orcid.org/0000-0002-2314-1493>
Universidade Tecnológica Federal do Paraná, Brasil
E-mail: hugoflorez@utfpr.edu.br

Gloria Patricia López Sepúlveda

ORCID: <https://orcid.org/0000-0002-5642-1889>
Universidade Tecnológica Federal do Paraná, Brasil
E-mail: gloriap@utfpr.edu.br

Anderson da Silva Soares

ORCID: <https://orcid.org/0000-0002-2967-6077>
Universidade Federal de Goiás, Brasil
E-mail: anderson@inf.ufg.br

Telma Woerle de Lima Soares

ORCID: <https://orcid.org/0000-0002-4927-2221>
Universidade Federal de Goiás, Brasil
E-mail: telma@inf.ufg.br

Resumo

Durante a pandemia da COVID-19 surgiram, por meio da análise e visualização de dados, plataformas de monitoramento para a doença. O foco da maioria das plataformas era o monitoramento a nível de Estado, com informações concentradas em macrorregiões. Com isso, cidades e municípios ficaram sem a devida atenção aos seus dados internos, o que resultou em dificuldade de tomada de decisão e estratégias de enfrentamento personalizadas para suas necessidades locais. Com uma prévia experiência no monitoramento da COVID-19 no Estado do Paraná coletando informações de acesso públicos, foi possível identificar novos tipos de dados e análises que poderiam ser realizadas com informações específicas de cada cidade ou município, no intuito de gerar informações úteis no contexto das características locais, como forma de auxiliar gestores na tomada de decisão. A capacidade de análise e visualização de dados da plataforma proposta, está diretamente relacionada aos desafios de estabelecer um padrão para obtenção dos mais variados tipos de dados encontrados nas organizações internas de cidades e municípios brasileiros. Assim, este trabalho apresenta uma metodologia de monitoramento inteligente dos dados da COVID-19, que compreende as etapas de obtenção, extração, limpeza, análise e a visualização dos dados. Com a análise de dados foi possível investigar ações e definir estratégias com foco nas características locais, gerando uma plataforma que foi transferida para uso entre alguns municípios do Paraná no ano de 2021.

Palavras-chave: Big data; Aprendizado de máquina; COVID-19; Análise de dados.

Abstract

During the COVID-19 pandemic, monitoring platforms for the disease emerged through the analysis and visualization of data. The focus of most platforms was monitoring at the State level, with information concentrated in macro-regions. As a result, cities and municipalities were left without due attention to their internal data, which resulted in difficult decision-making and coping strategies tailored to their local needs. With previous experience in monitoring COVID-19 in the State of Paraná, using public access data, it was possible to identify new types of data and analyzes that could be carried out with specific information for each city or municipality, in order to generate useful information in the context local characteristics, as a way of helping managers in decision making. The platform's ability to analyze and

visualize data is directly related to the challenges of establishing a standard for obtaining the most varied types of data found in the internal organizations of Brazilian cities and municipalities. Thus, this work presents a methodology for intelligent monitoring of COVID-19 data, which comprises the steps of obtaining, extracting, cleaning, analyzing and visualizing the data. With data analysis, it was possible to investigate actions and define strategies focusing on local characteristics, generating a platform that was transferred to some municipalities in Paraná in 2021.

Keywords: Big data; Machine learning; COVID-19; Data analysis.

Resumen

Durante la pandemia de COVID-19 surgieron plataformas de monitoreo de la enfermedad a través del análisis y visualización de datos. El enfoque de la mayoría de las plataformas fue el monitoreo a nivel estatal, con información concentrada en macrorregiones. Como resultado, las ciudades y municipios quedaron sin la debida atención a sus datos internos, lo que resultó en una difícil toma de decisiones y estrategias de afrontamiento adaptadas a sus necesidades locales. Con la experiencia previa en el monitoreo del COVID-19 en el Estado de Paraná, utilizando datos de acceso público, fue posible identificar nuevos tipos de datos y análisis que podrían realizarse con información específica para cada ciudad o municipio, con el fin de generar información útil en las características locales del contexto, como una forma de ayudar a los gerentes en la toma de decisiones. La capacidad de la plataforma para analizar y visualizar datos está directamente relacionada con los desafíos de establecer un estándar para obtener los más variados tipos de datos que se encuentran en las organizaciones internas de las ciudades y municipios brasileños. Así, este trabajo presenta una metodología para el monitoreo inteligente de datos de COVID-19, que comprende los pasos de obtención, extracción, limpieza, análisis y visualización de los datos. Con el análisis de datos, fue posible investigar acciones y definir estrategias con foco en las características locales, generando una plataforma que fue transferida a algunos municipios de Paraná en 2021.

Palabras clave: Big data; Aprendizaje automático; COVID-19; Análisis de datos.

1. Introdução

O ano de 2020 foi marcado pela pandemia da COVID-19, causando altas taxas de contaminação e letalidade (Madeiro, 2020). Nos primeiros meses de pandemia, as universidades se destacaram no sentido de identificar uma medicação e vacinas eficazes para a população (Matheus, et al., 2020).

A COVID-19, causada pelo coronavírus (SARS-CoV-2), foi responsável por mudar drasticamente o mundo em variados aspectos, como, transporte, educação, políticas de saúde, abastecimento de mercadorias e outros (Ribeiro, et al., 2020). Para entender melhor sobre a doença e sua forma de transmissão, a obtenção e a análise de dados relacionados à COVID-19 foi e ainda é essencial para avaliar a situação epidêmica (Ponjavic, et al., 2020).

Uma das primeiras plataformas de análise e visualização de dados a nível estadual da COVID-19 foi o Portal PR (Ito, 2021), que no ano de 2020 estabeleceu um formato para informar a população e fazer observações para possíveis tomadas de decisão no Estado do Paraná. A obtenção de dados macrorregionais do estado foi um desafio na manutenção da plataforma de monitoramento, uma vez que o número de fontes era escasso e os formatos de dados variavam desde arquivos de texto até arquivos em PDF (Ito, 2021).

A análise de dados da COVID-19 provou ser capaz de auxiliar no combate à doença e de prover uma série de soluções promissoras para tal (Ayyoubzadeh, et al., 2020). Isso é evidenciado uma vez que os dados elucidam o entendimento sobre a COVID-19 em termos de estrutura do vírus, automação de testes e criação de vacinas. Além disso, os dados combinados com inteligência artificial podem criar modelos de simulação tanto para a estimativa de surtos quanto para a previsão futura da epidemia, nos quais ajudam as agências de saúde a monitorar a propagação da doença a fim de preparar melhores medidas preventivas (Pham, et al., 2020). Com isso em vista, o desenvolvimento do Portal PR trouxe resultados importantes em termos de auxílio a Secretaria da Saúde (SESA) do Paraná e informação para a população paranaense através de um monitoramento de dados amplo e aberto, utilizando desde análises de casos e comportamento da doença, até previsões de surtos e lotação de leitos.

Nesse contexto, o monitoramento de dados se dá pela exibição dos dados em *dashboards* (PowerBI, 2022), permitindo a tomada de decisões de maneira facilitada e o monitoramento dinâmico através de várias maneiras. Os *dashboards* podem ser adaptados para adicionar ou remover camadas de informações, filtrados por região, e etc. Além disso, essas interfaces facilitam

a visualização do resultado das intervenções realizadas com o objetivo de melhorar a tomada de decisão para alocação de recursos direcionados em localidades específicas.

Com o avanço do monitoramento da pandemia dentro do Estado do Paraná foi observada a necessidade de explorar uma estratégia semelhante para cidades e municípios, com foco em analisar dados locais e auxiliar os gestores na escolha de ações mais assertivas (Ribeiro, et al., 2020). As cidades geram uma grande quantidade de dados internos com características do contexto da localidade e informações com especificidade urbana e até entre os bairros (Pham, et al., 2020). Com o Painel PR efetuando a coleta de dados direcionada aos Estados brasileiros e com foco no Paraná, as cidades e municípios ficaram sem o devido monitoramento de dados da COVID-19.

Assim, para desenvolvimento de uma plataforma de monitoramento para cidades e municípios é estabelecido o mesmo desafio imposto ao Painel PR, a correta obtenção de dados internos dentro das organizações municipais. Para isso, é preciso criar um método que permita capturar dados em diferentes organizações internas, e ao mesmo tempo, facilitar o processo de entrada destes dados pelos gestores da saúde, gerando volume e variedade de informações para que a plataforma de monitoramento consiga produzir análises para o contexto local. Junto com a correta obtenção de dados é preciso estabelecer novas análises que considerem os aspectos locais como bairros e as diferentes áreas de uso pelos habitantes como setores de comércio, parques públicos, transporte e locais de trabalho, para que sejam feitas inferências com valor para guiar a tomada de decisão pelos gestores da saúde.

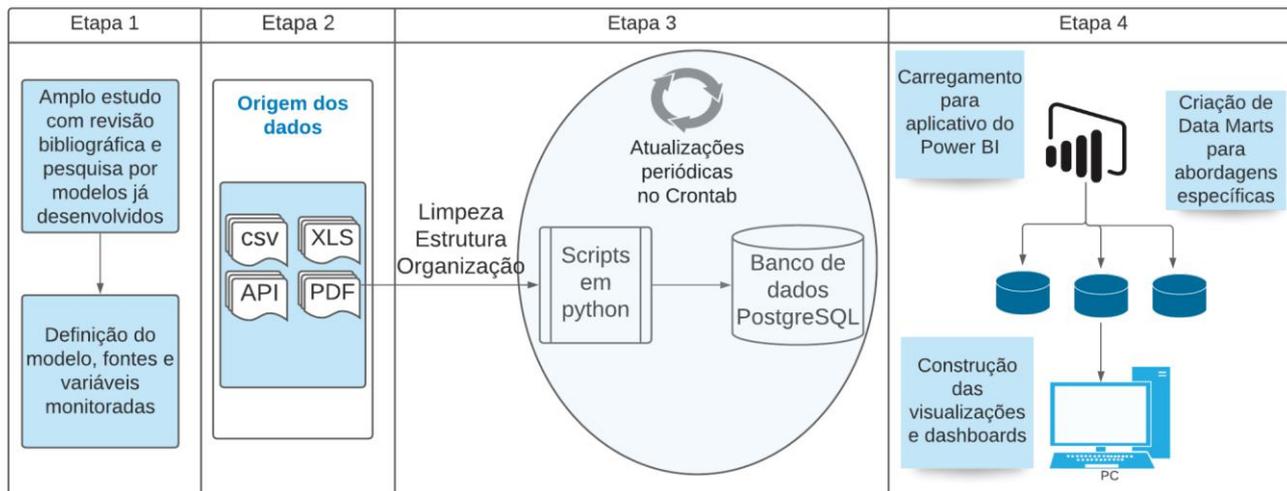
Diante desse cenário, a falta de análise de dados nas cidades e municípios levou à tomada de decisões generalistas que, por sua vez, não levaram em conta as particularidades de cada cidade e/ou região (Alsunaidi, et al., 2021). A partir desse entrave, foi desenvolvida a Plataforma de Monitoramento Inteligente da COVID-19 (DataCovid), um *dashboard* interativo projetado para incluir principalmente as informações das cidades e municípios que seriam atualizadas pelos próprios representantes, com variedade de visualizações e de forma intuitiva.

O DataCovid inclui análise e previsão de casos raramente vistos nas soluções de visualização mais comuns disponíveis até então para a COVID-19, além das análises direcionadas aos contextos locais e internos das cidades e seus bairros e áreas comuns para os habitantes. A exibição dos dados em forma de *dashboards* é de grande ajuda, pois com auxílio visual é possível ter uma melhor compreensão dos dados. Entre os resultados, a plataforma DataCovid foi implantada no ano de 2021 em algumas cidades e municípios para o monitoramento da COVID-19, com auxílio aos gestores de saúde com coleta e processamento de dados diários.

2. Metodologia

A metodologia descreve como foi realizada cada etapa do desenvolvimento da Plataforma de Monitoramento Inteligente de Dados da COVID-19 em cidades e municípios. As etapas são executadas na ordem estabelecida pela arquitetura proposta que pode ser vista na Figura 1, sendo estas na forma das atividades principais de obtenção e extração de dados (etapas 1 e 2), limpeza e análise (etapa 3) e visualização de dados (etapa 4).

Figura 1 – Algoritmo de extração e coleta de dados.



Fonte: Autores.

É importante ressaltar que cada etapa foi executada conforme descrito nas seções 2.1 a 2.5.

2.1 Obtenção de dados

O processo de obtenção se configura pela automatização da coleta de dados por diversas fontes heterogêneas, de modo que a quantidade de dados obtida seja suficiente para lidar com o problema atual (Roh, et al., 2019). Para obter informações dos municípios do Estado do Paraná, foram utilizados como fonte primária os boletins epidemiológicos elaborados diariamente pela SESA. Estes são varridos por meio de algoritmos especializados em extração de dados no formato de PDF, uma vez que este é o formato padrão dos boletins. O processamento é realizado inicialmente medindo as linhas e colunas que compõem as tabelas onde estão informações de novos casos, óbitos e ocupação de leitos.

Com a identificação do padrão esperado das tabelas, é possível navegar por células específicas e obter os seus dados. Em seguida, os quadros com informações gerais de infecção, comorbidades e demográfica por sexo e faixa etária são obtidos. Um desafio da raspagem de dados em formato de PDF é a alteração das tabelas e seus tamanhos e composição que pode acontecer de tempos em tempos, sendo necessário uma nova análise do padrão do documento para ajustes no algoritmo.

O Quadro 1 apresenta as principais variáveis monitoradas pela plataforma e que são obtidas via raspagem automatizada dos boletins epidemiológicos. Dentre as principais variáveis encontram-se a que mostra a quantidade acumulada de casos desde o início da pandemia (Moreira & Pinheiro, 2020), que ajuda a entender a extensão da taxa de transmissão da doença e comparar a situação do país com o estado pandêmico global. A quantidade de óbitos, composta pelos indicadores Óbitos por Idade e Óbitos por Comorbidade, também é importante pois esclarece quais ações devem ser tomadas para diminuir a taxa de mortalidade. Por meio da plataforma, pôde-se constatar que do total de mortes no Paraná, grande parte foi por associação da COVID-19 com alguma comorbidade ou saúde frágil de idosos (Ito, 2021).

Quadro 1 – Algumas das principais variáveis que são analisadas na Plataforma de Monitoramento Inteligente.

VARIÁVEIS	COMPOSIÇÃO
Casos confirmados	Quantidade acumulada de casos desde o início da pandemia.
Novos casos	Quantidade de novos casos do dia atual.
Evolução dos casos	Taxa da evolução de novos casos confirmados do dia atual em relação ao dia anterior.
Letalidade	Proporção entre o número de óbitos e o número total de casos confirmados, ambos no acumulado desde o início da pandemia.
Mortalidade	Proporção entre o número de óbitos e a quantidade de pessoas no local, que pode ser uma cidade, estado ou país..
Média móvel de confirmados	Taxa de evolução dos casos no intervalo de médias de 7 dias, do dia atual e dos seis dias anteriores.
Evolução da média móvel	Diferença da taxa da média móvel entre duas semanas diferentes.
Óbitos por comorbidade	Quantidade de óbitos de pessoas que sofriram alguma doença anterior à COVID-19, quantidades separadas por doenças específicas.
Óbitos por idade	Levantamento dos óbitos por diferentes faixas etárias.
Leitos enfermaria ocupados	Quantidade de leitos de enfermaria ocupados do total existente.
Leitos UTI ocupados	Quantidade de leitos de UTI ocupados do total existente.
Taxa de Ocupação Regiões	Taxa de ocupação de leitos hospitalares por regiões diferentes do estado.
Casos Suspeitos	Quantidade de casos ainda não confirmados à espera do resultado do exame.
Casos Recuperados	Quantidades acumuladas de pessoas que se recuperaram da doença.
Previsão de confirmados	Previsões dia a dia em um intervalo de 20 dias à frente do total de casos confirmados.
Previsão de óbitos	Previsões dia a dia em um intervalo de 20 dias à frente do total de óbitos.
Previsão de leitos	Previsões dia a dia em um intervalo de 20 dias à frente da taxa de ocupação de leitos hospitalares.

Fonte: Adaptado de Ito (2021).

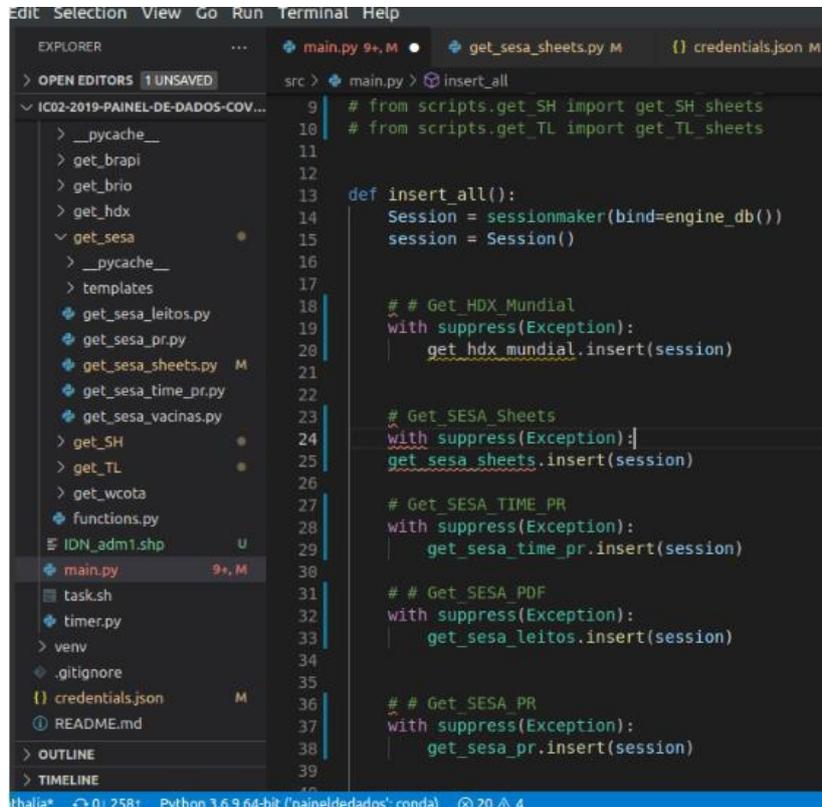
Para a obtenção de dados específicos das cidades e municípios, aqueles que ficam em posse dos gestores de saúde, foi necessário gerar uma Interface de Programação de Aplicação (API). Esta é responsável pela obtenção automatizada de dados preenchidos diariamente pelos gestores da saúde nos sistemas oficiais de auditoria do governo, como o SIVEP-Gripe, o sistema de informação oficial para notificação de casos e óbitos, e no e-SUS VE, ferramenta de registro online de notificação de casos de síndrome gripal. A API foi uma solução necessária devido ao retrabalho que os profissionais de saúde teriam ao preencherem os dados mais de uma vez em diferentes ferramentas. Com a automação do processo, a usabilidade da metodologia de coleta tornou-se adequada para a realidade das cidades sem sobrecarregar ou desviar a função dos gestores de saúde.

2.2 Extração de Dados

Para a extração dos dados, foram utilizadas diversas fontes, no estado do Paraná, a fonte foi a partir dos boletins divulgados pela SESA, onde algoritmos especializados retiraram as informações de seus PDFs e de planilhas. Para os dados em escala internacional, como o Brasil e outros países, o site WCota (Cota, 2020) foi o utilizado, cujos dados são disponibilizados para o processamento de algoritmos. Os dados são retirados diariamente para o uso em futuras previsões.

O algoritmo usado para a extração de dados foi desenvolvido na linguagem de programação python, fazendo uso da biblioteca beautifulsoup, responsável por analisar documentos HTML e XML, como demonstrado na Figura 2.

Figura 2 – Algoritmo de extração e coleta de dados.



```
9 # from scripts.get_SH import get_SH_sheets
10 # from scripts.get_TL import get_TL_sheets
11
12
13 def insert_all():
14     Session = sessionmaker(bind=engine_db())
15     session = Session()
16
17
18     ## Get_HDX_Mundial
19     with suppress(Exception):
20         get_hdx_mundial.insert(session)
21
22
23     # Get_SESA_Sheets
24     with suppress(Exception):
25         get_sesa_sheets.insert(session)
26
27
28     # Get_SESA_TIME_PR
29     with suppress(Exception):
30         get_sesa_time_pr.insert(session)
31
32
33     ## Get_SESA_PDF
34     with suppress(Exception):
35         get_sesa_leitos.insert(session)
36
37
38     ## Get_SESA_PR
39     with suppress(Exception):
40         get_sesa_pr.insert(session)
```

Fonte: Autores.

É possível observar pela imagem como a extração foi automatizada a partir da API que recebia como entrada os dados de login dos sistemas SIVEP-Gripe e e-SUS VEO. As informações do dia atual eram capturadas e enviadas diretamente para o *Data Lake* (DL) da plataforma. O DL armazena todos os dados extraídos e utiliza como tecnologia principal um banco de dados PostgreSQL, escolhido por sua escalabilidade e sua robustez. Diariamente são realizados três agendamentos para extração e coleta dos dados automatizados no servidor Linux da plataforma utilizando um crontab, onde o DL é mantido com os dados atuais para sua posterior limpeza e organização.

2.3 Limpeza de Dados

O processo de extração, transformação e carregamento (ETL) de dados é um tipo de *Data Integration* em três etapas, usado para combinar informações de diversas fontes (Clement, et al., 2020). Nesse processo, os dados da COVID-19 foram retirados (extraídos) de um sistema-fonte que os municípios utilizam junto ao governo, convertidos (transformados) em um formato que possa ser analisado, e armazenados (carregados). Fazer a integração dos dados que são obtidos das fontes do governo de coletas de grupos sem API e dentro das cidades com dados faltantes e duplicados não é uma tarefa simples (Luo, et al., 2020). O Quadro 2 mostra os principais tipos de erros que são corrigidos nas fontes de dados.

Quadro 2 – Erros comuns de dados.

#	Tipo de consulta	Tupla	Attr	Faltante	Atípico
1	X=X(Número).Y=Y	Sim	Não	Sim	Sim
2	X=X(Categoria).Y=Y	Sim	Sim	Sim	Sim
3	X=BIN(X).Y=AGG(Y)	Sim	Não	Sim	Sim
4	X=GRUPO(X).Y=AGG(Y)	Sim	Sim	Sim	Sim

Fonte: Autores.

Por exemplo, se X' é gerado a partir de X por algumas operações de seleção, algumas tuplas podem estar faltando em X' porque elas não correspondem exatamente às condições de seleção devido ao sinônimo e, então, o valor agregado em Y' será incorreto devido a erros em X' . As entradas duplicadas e atributos prejudicam a estimativa de predições da plataforma e os valores faltantes dificultam o acompanhamento da progressão da contaminação.

Para facilitar as análises na plataforma, foram realizadas transformações nos dados, a primeira foi o mapeamento sobre os dados de mobilidade, casos diários, datas dos casos e estratégias de enfrentamentos e correlaciona todos estes em novas tabelas anexas no Data Warehouse. Outra grande transformação foi a adição de valor aos dados, onde todas as informações coletadas de fontes externas que são passadas apenas semanalmente são segmentados em dias, com base na comparação de dias da semana e datas, para padronizar o formato para as predições e algoritmos de estimativa de infecção.

2.4 Análise de Dados

Análise de dados é o processo de aplicação sistemática de técnicas estatísticas e/ou lógicas para descrever, ilustrar, descobrir padrões e avaliar os dados (Salloum, et al., 2016). A plataforma de monitoramento permite avaliar os cenários e, com isso, ajudar no combate a pandemia do COVID-19. Entre as análises disponíveis estão a avaliação da efetividade de ações já executadas pela cidade, monitoramento da mobilidade urbana em zonas públicas da cidade, além das predições de ocupação de leitos e de novos casos, com um horizonte de até 15 dias à frente da data atual. Avaliar a efetividade das ações municipais de enfrentamento a COVID-19 é uma importante análise para indicar quais podem ser realizadas novamente e quais não. Para analisar essa situação, os municípios fornecem a data de início e fim da ação e, com isso, é realizada a análise da média móvel antes, durante e depois deste período.

Caso a média móvel seja menor nos próximos 15 dias após o encerramento da medida, e a evolução da infecção tiver uma redução a partir de 30%, um indicador sinaliza que esta ação impacta no número de casos, e que o município pode adotar mais vezes esse método, se houver uma alta nos casos. Ações como fechamento de comércio, rodízio em bairros e horários controlados de funcionamento de estabelecimentos eram as mais comuns para análise.

A mobilidade da população foi outra análise importante para a observação dos casos. O monitoramento ocorre através de dados gerados por smartphones, como GPS e pontos de acesso de WI-FI, e de forma anônima. Muitas empresas usam os dados de GPS com os de WI-FI e geram uma estimativa do local onde o aparelho se encontra. A Google oferece um serviço de relatórios de mobilidade urbana para o auxílio da pandemia da COVID-19, apesar do serviço ter previsão de ser descontinuado, os relatórios ajudaram a plataforma de monitoramento da COVID-19 a gerar gráficos para ilustrar como estava a tendência de movimentação de pessoas em diferentes áreas e divididos por setores dentro da cidade.

Entre as análises preditivas estão a ocupação de leitos e novos casos que foram feitas utilizando o histórico dos casos dentro do município desde o início da pandemia. Houve uma tentativa de prever a força da infecção da COVID-19. O modelo utilizado foi o *Susceptible-Infected-Removed*, SIR (Blackwood & Childs, 2018) e está representado na equação 1:

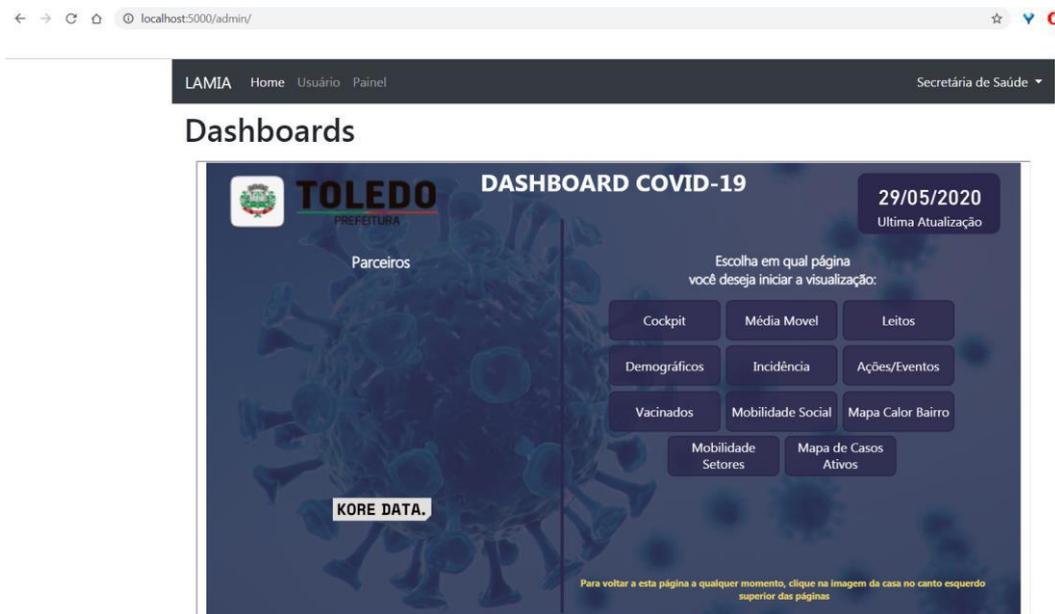
$$dS/dt = -\beta SI, dI/dt = \beta SI - \gamma I, dR/dt = -\gamma I \quad (1)$$

onde S , I e R são: suscetível, infectado e descartado, respectivamente, β é o índice de transmissão e γ é o índice de recuperados. O modelo não apresentou um bom desempenho, já que ele não considera dois cenários da pandemia do COVID-19, o primeiro é que paciente que já contraiu a doença não poderia ser infectados novamente e o segundo é a variação de β e γ . Dado a quantidade de informações, as predições foram realizadas utilizando o histórico de casos do momento com o algoritmo *Random Forest*, que apresentou previsões satisfatórias com uma Raiz quadrada do erro-médio (RMSE) de 0.1654.

2.5 Visualização de Dados

A visualização de dados é a representação gráfica de informações e dados usando elementos visuais como tabelas, gráficos e mapas (Saini et al., 2021). A forma de exibição dos dados na plataforma de monitoramento do coronavírus é completa e possui variações. Algumas das principais maneiras são: o monitoramento da força de infecção por bairros, que exibe um mapa da cidade associado a áreas redondas coloridas em destaque que representam onde existem os surtos, a progressão de casos por bairro, que mostra todos os dados de cada bairro individualmente e a mobilidade urbana na cidade, que informa a mobilidade de cada setor no município de maneira individual (Figura 3).

Figura 3 – Página inicial do *dashboard* que sumariza as opções disponíveis.



Fonte: Autores.

Monitorar a força relacionada à infecção da COVID-19 contribui para mensurar os dados obtidos e de outras análises em conjunto com esta, além de permitir a representação de possíveis cenários futuros que possam vir a ocorrer em termos de alcance da doença (Piccolomini & Zama, 2020). Os benefícios do monitoramento da força de infecção por bairro vão além de apenas ter o conhecimento dessas informações. Com base no histórico dos dados, nota-se que existe uma sobreposição entre áreas onde os grupos de risco estão e bairros com casos confirmados. Portanto, infere-se que outros locais visitados por grupos de risco também podem ser perigosos. Dessa forma, as autoridades podem se precaver contra essas áreas com antecedência e adotar diferentes políticas anti epidêmicas em áreas diferentes para obter o controle eficaz.

O monitoramento da progressão de casos por bairros emprega em seu cerne a utilização da Curva R para representar,

por meio do mapa de calor, os bairros onde o vírus está com maior índice de propagação. A Curva R representa o número médio de casos secundários da doença causados por um único indivíduo infectado durante seu período infeccioso (Figura 4). O R_0 (número básico de reprodução) representa um valor no início da epidemia, tomando que a população toda é suscetível à infecção e nenhuma medida social restritiva foi tomada até o momento. Mesmo que R_0 seja útil para analisar a gravidade geral da epidemia no início, é limitado quando se diz respeito à avaliar uma mudança subsequente causada por uma quebra de comportamento da população. Sendo assim, é utilizado o R_t (número de reprodução instantânea) que permite avaliar as medidas restritivas impostas, fazendo com que a população fique em alerta, o que deve levar a uma diminuição do valor de R_t .

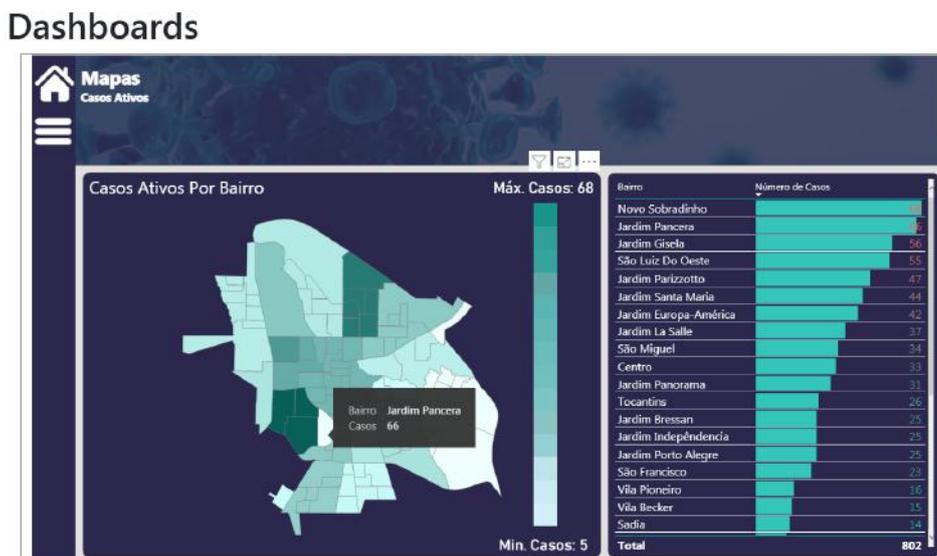
Os dados de mobilidade urbana fornecem informações sobre como as pessoas se deslocam nas áreas metropolitanas, esse tipo de dados fornece estatísticas abrangentes e insights importantes sobre padrões e tendências de transporte (Hu, et al., 2021). Na plataforma de monitoramento os dados de mobilidade espaço-temporal descrevem trajetórias e padrões dos habitantes da cidade, que podem ser coletados de diversas formas. A maneira adotada na plataforma da COVID-19 chama-se *wardriving*, na qual grandes corporações como a Google combinam APs (pontos de acesso/roteadores) Wi-Fi com dados de GPS a fim de melhorar a precisão da localização dos smartphones. Essa tecnologia funciona a partir de permanência, saltos e deslocamentos, ou seja, o tempo que um usuário permaneceu na imediação de uma estação Wi-Fi e movimentou-se entre estações dependendo das diferenças de tempo observadas em um dispositivo em cada estação.

3. Resultados e Discussão

Com a visualização assertiva dos dados coletados conjuntamente à cada cidade é possível exercer monitoramentos inéditos para as cidades e municípios dando embasamento às tomadas de decisão. Com efeito, o LAMIA recebeu o prêmio SESI ODS (SESI-ODS, 2022) de 2020, entregue anualmente a empresas e instituições com projetos de alto impacto dentro da sociedade e seis cidades receberam via transferência tecnológica plataformas customizadas pelos dados coletados diretamente com cada um destes municípios.

A Figura 4 exhibe o *dashboard* dos casos ativos em todos os bairros de uma determinada cidade. Por conseguinte, os benefícios do monitoramento por bairros resultaram em escolhas administrativas de fechamento dessas regiões e isolamentos menores e mais concentrados, com maior redução da força da infecção no próprio bairro e de outros em suas imediações.

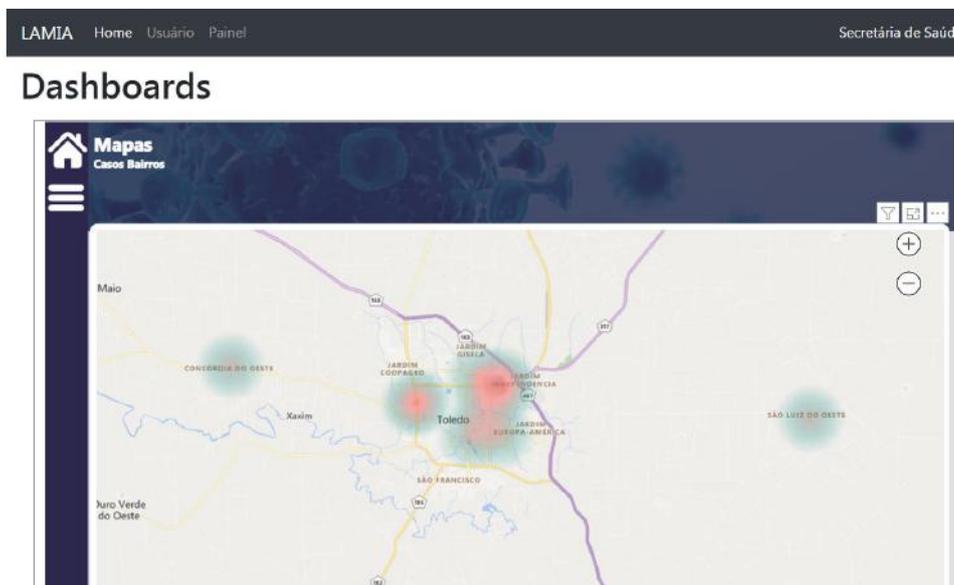
Figura 4 – Dashboard dos casos ativos por bairro.



Fonte: Autores.

A representação da Curva *R* (Figura 5) no formato de mapa de calor traz consigo uma série de benefícios. A projeção, realizada via mapa de calor e a progressão dos casos por bairros, foi realizada a partir da Curva *R* facilitando a interpretação e análise dos dados. Com isso, é possível monitorar famílias e partes ainda menores dentro de um bairro ou comunidade. Sendo assim, o uso dessa tecnologia também contribui na descomplexificação dos procedimentos anti epidêmicos.

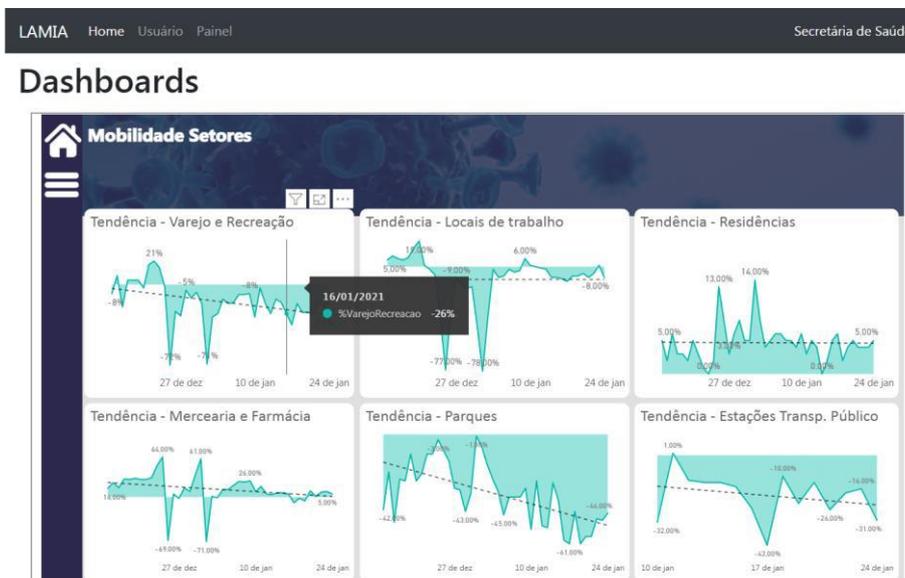
Figura 5 – *Dashboard* do mapa de calor que utiliza a Curva *R*.



Fonte: Autores.

De maneira análoga, o monitoramento da mobilidade urbana dentro das cidades (Figura 6), também apresenta vantagens. Em decorrência da coleta de informações dos fluxos de pessoas em diferentes tipos de segmentos do município, como, por exemplo, em locais de trabalho, estações de transporte público, parques, áreas residenciais ou varejo, é possível acompanhar a proporção da taxa de adesão às medidas restritivas impostas na cidade ou se elas não surtiram efeito. Tendo, dessa forma, uma nova ferramenta disponível para controlar, eficientemente, as ações de contenção da epidemia.

Figura 6 – *Dashboard* da mobilidade nos setores da cidade.



Fonte: Autores.

4. Conclusão

Este trabalho apresenta uma plataforma de monitoramento de dados da COVID-19 para monitorar cidades e municípios, diferente de outras plataformas que focam apenas em Estados. Ressalta-se o benefício da plataforma para o gerenciamento de ações em pontos estratégicos que podem ser visualizados por meio da plataforma, e que podem ser constatadas em relação à sua eficiência. Foram utilizados algoritmos especializados para a obtenção dos dados em plataformas do governo, e foi gerada uma API para automatizar o processo da obtenção. Os resultados foram utilizados por municípios do Paraná para monitorar a evolução do quadro da COVID-19 e tomada de decisão utilizando Big Data e Inteligência Artificial, integrados no processo ETL de dados para extrair, transformar e armazenar os dados.

Aplicando ferramentas como *dashboards* interativos para exibição dos dados de variadas maneiras e de técnicas como Curva *R* e *wardriving* para melhor organização e coleta das informações. Nos experimentos os resultados mostram uma boa performance na gestão e monitoramento de grandes volumes de dados. Com isso, a solução apresentada se apresenta como uma ótima opção para gestão de hospitais e postos de saúde, em especial aos que recebem pacientes contaminados com o Coronavírus, para que possam ter previsões e visualizem melhor o atual estado viral da doença e os direcionam para uma melhor solução.

Para trabalhos futuros, melhorias podem ser adicionadas aos *dashboards*, além de disponibilizar a plataforma para outros Estados do Brasil. Vale destacar que a plataforma pode ser ajustada para o monitoramento de outras doenças com potencialidades de ocasionar uma pandemia.

Referências

- Alsunaidi, S. J., Almuhaideb, A. M., Ibrahim, N. M., Shaikh, F. S., Alqudaihi, K. S., Alhaidari, F. A., Khan I. U., Aslam, N., & Alshahrani, M. S. (2021). Applications of Big Data Analytics to Control COVID-19 Pandemic. *Sensors (Basel)*, 21(7), 2282. 10.3390/s21072282
- Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Niakan Kalthori, S. R. (2020). Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR public health and surveillance*, 6(2), 18828. 10.2196/18828
- Blackwood, J. C., & Childs, L. M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics* 2018, 5(1), 195.10.1080/23737867.2018.1509026
- Clement, F., Kaur, A., Sedghi, M., Krishnaswamy, D., & Punithakumar, K. (2020). Interactive Data Driven Visualization for COVID-19 with Trends, Analytics and Forecasting. *2020 24th International Conference Information Visualisation(IV)*, 593-598. 10.1109/IV51561.2020.00101
- Cota, W. (2020). Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.362>
- Hu, T., et al. (2021). Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges. *International Journal of Digital Earth*, 14(9), 1126-1147. <https://doi.org/10.1080/17538947.2021.1952324>
- Ito, G. C. (2021). *Informar para superar: retratos sobre a COVID-19 na web*. Livros Editora UniFil. http://periodicos.unifil.br/index.php/livros_unifil/article/view/1478
- Luo, Y., Chai, C. Qin, X., Tang, N., & Li, G. (2020). Interactive cleaning for progressive visualization through composite questions. *IEEE. 2020 36th International Conference on Data Engineering (ICDE)*. [S.l.: s.n.], 733-744. 10.1109/ICDE48307.2020.00069
- Madeiro, C. (2020). Covid-19 já é a maior causa de mortes no Brasil registrada em um único ano. *Uol*. <https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2020/09/15/com-133-mil-obitos-covid-ja-tem-recorde-como-causa-morte-no-pais-em-um-ano.htm>
- Matheus, D., et al. (2020). Universidades públicas de São Paulo juntas no combate à covid-19. *Jornal da USP*. <https://jornal.usp.br/institucional/universidades-publicas-de-sao-paulo-juntas-no-combate-a-covid-19/>
- Moreira, A., & Pinheiro, L. (2020). OMS declara pandemia de coronavírus. *G1 – Bem-Estar*. <https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/11/oms-declara-pandemia-de-coronavirus.ghtml>
- Pham, Q., Nguyen, D., Huynh-The, T., Hwang, W., & Pathirana, P. (2020). Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. *IEEE access, Institute of Electrical e Electronics Engineers*, 8, 130820-130839. 10.1109/ACCESS.2020.3009328
- Piccolomini, E. L., & Zama, F. (2020) Monitoring Italian COVID-19 spread by a forced SEIRD model. *PLoS one*, 15(8), e0237417. <https://doi.org/10.1371/journal.pone.0237417>
- Ponjavic, M., Karabegović, A., Ferhatbegović, E., Tahirović, E., Uzunović, S., Travar, M., Pilav, A., Mulić, M., Karakaš, S., Avdić, N., Mulabdić, Z., Pavić, G., Bičo, M., Vasilj, I., Mamić, D., & Hukić, M. (2020). Spatio-temporal data visualization for monitoring of control measures in the prevention of the spread

of COVID-19 in Bosnia and Herzegovina. *Med Glas (Zenica)*, 17(2), 265–274. 10.17392/1215-20

PowerBI: Dashboards para usuários de negócios do serviço do Power BI (2022). *Microsoft Ignite*. <https://learn.microsoft.com/pt-br/power-bi/consumer/end-user-dashboards>

Ribeiro, M., Nisi, V., Prandi, C., & Nuneset, N. (2020). A data visualization interactive exploration of human mobility data during the COVID-19 outbreak: a case study 2020 *IEEE Symposium on Computers and Communications (ISCC)*, 1–6. 10.1109/ISCC50000.2020.9219552

Roh, Y., Heo, G., & Whang, S. E. (2019). A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 33(4), 1328-1347. 10.1109/TKDE.2019.2946162

Saini, S. K., Dhull, V., Singh, S., & Sharma, A. (2020). Visual exploratory data analysis of covid-19 pandemic. *5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*. 1-6. 10.1109/ICRAIE51050.2020.9358331

Salloum, S., Dautov, R., Chen, X., et al. (2016). Big data analytics on Apache Spark. *International Journal of Data Science and Analytics*, 1(3), 145-164, <https://doi.org/10.1007/s41060-016-0027-9>

SESI-ODS: Projeto desenvolvido na UTFPR de Santa Helena ganha Prêmio SESI ODS (2020). *O Presente*. <https://www.opresente.com.br/municipios/projeto-desenvolvido-na-utfpr-de-santa-helena-ganha-o-premio-sesi-ods-2020/>