

## Capítulo III – Uma solução *web* e *mobile* para predição e análise de sentimentos sobre o avanço da Covid-19

Alexandre Marcos de Souza Junior<sup>7</sup>

Rubens Fujita Hirata<sup>8</sup>

Anderson Paulo Ávila Santos<sup>9</sup>

Fabio Takeshi Matsunaga<sup>10</sup>

Lucas Busatta Galhardi<sup>11</sup>

### RESUMO

A comunidade acadêmica tem se mobilizado para buscar soluções para realizar o monitoramento dos avanços da pandemia da Covid-19. Diante do exposto, o objetivo do presente projeto é realizar a projeção do avanço da Covid-19 de modo que os dados sejam acessíveis pela população por meio de uma aplicação *web* e *mobile*, a partir dos quais serão disponibilizados gráficos sobre dados oficiais. Estes dados serão processados por meio de regressão linear, para realizar estimativas sobre avanços. Além disso, será utilizada uma biblioteca em linguagem Python para realizar a análise de sentimentos da população sobre a pandemia, cujo dados serão buscados em postagens do Twitter. De acordo com os resultados, o número de óbitos representa em torno de 2% dos casos totais, com uma faixa etária entre 41 e 59 anos, sendo a maioria do sexo feminino.

**Palavras-chave:** API, *back-end*, *front-end*, Inteligência Artificial, *Machine Learning*.

### **A web and mobile solution for predicting and analyzing sentiments about the advance of Covid-19**

#### ABSTRACT

The academic community has mobilized to seek solutions to monitor the advances of the Covid-19 pandemic. The main objective of this project is to carry out the projection of the advance of Covid-19 so that the data are accessible by the population. For this, a web and mobile application will be developed to provide graphics on official data.

<sup>7</sup> Graduando em Engenharia de Software.

<sup>8</sup> Graduando em Engenharia de Software.

<sup>9</sup> MSc. Anderson Paulo Ávila Santos. E-mail: anderson.avila@sistemafiep.org.br

<sup>10</sup> MSc. Fabio Takeshi Matsunaga. E-mail: fabio.matsunaga@sistemafiep.org.br

<sup>11</sup> MSc. Lucas Busatta Galhardi. E-mail: lucas.galhardi@sistemafiep.org.br

These data will be processed through linear regression, to make estimates of disease progress. In addition, the data will be processed through a library in Python language to carry out the analysis of the population's feelings about the pandemic, whose data will be searched on Twitter posts. According to the results, the number of deaths represents around 2% of the total cases, with an age group between 41 and 59 years, the majority being female.

**Key-words:** API, Artificial Inteligente, back-end, front-end, Machine Learning.

## 1. INTRODUÇÃO

O Brasil sofre em 2020 com a chegada de uma pandemia originada na China. Esta pandemia é efeito de uma doença causada pelo coronavírus SARS-CoV-2, que apresenta um quadro clínico que varia de infecções assintomáticas a quadros respiratórios graves (MINISTÉRIO DA SAÚDE, 2020a).

O evento abalou fortemente a sociedade brasileira, tanto em aspectos econômicos, como o fechamento de 1,1 milhão de vagas de emprego e retração de 5,89% do PIB do mercado brasileiro (G1, 2020), mas também sociais, com o óbito de mais de 131.274 falecidos e 4.315.858 pessoas infectadas em 13/09/2020, segundo Ministério da Saúde (2020b).

Diante desta tragédia no cenário mundial, a comunidade acadêmica e científica tem realizado diversas pesquisas e publicações sobre o assunto. No trabalho de Neves e Barros (2020), é utiliza um modelo matemático não linear proposto por Gompertz para verificar se as medidas impostas pelos governantes estão tendo resultados, já que este tipo de análise é aplicado em estudos populacionais interpretativos, levando a deduzir a velocidade e a aceleração dos casos.

Já Fernandes (2020) relata que os modelos matemáticos epidemiológicos usados em pandemias passadas, foram pouco eficazes de predizer números reais. De forma simplificada, obtendo resultados estatísticos, buscou dados dessa pandemia em 30 países que já haviam atingido o pico máximo e estavam ao menos com 5 dias de queda ou estabilização, tendo 18 desses 30 países considerados concluídos o ciclo completo da pandemia por apresentar novos casos menor que 70% dos seus picos de novos caso por dia. O Brasil tem o terceiro maior número de leitos de UTI do mundo, atrás do EUA e da Alemanha.

Diante do cenário atual da pandemia na região de Londrina-PR com base em dados clínicos disponibilizado pelo órgão público SESA-PR (Secretaria da Saúde Paraná), o objetivo do presente projeto é realizar a projeção do avanço da doença por meio de uma aplicação *web* e *mobile*. Para isso, serão coletados dados oficiais sobre a Covid-19 na base de dados do SESA-PR, os quais passarão por processo de filtragem, processamento e aplicação de regressão linear, um algoritmo de *Machine Learning* para predição dos dados e análise de sentimentos da população em rede social ao enfrentar tal pandemia e o isolamento social.

## 2. FUNDAMENTAÇÃO TEÓRICA

Com a frase “Sem precedentes!”, Giannella e Velho (2020) desenvolveram a pesquisa sobre o tema levantando a seguinte questão: “quando isso vai acabar?”. Tal questão é uma dúvida emergente em momento de larga e acelerada produção de dados, que mesmo com o excesso de informação, não parece aclarar a visualização. Segundo Giannella e Velho, a Matemática e o Design da Informação baseadas em evidências, podem ser: i) a evidenciação de eventos concretos e finalizados, com conjuntos de dados com início, meio e fim e ii) a evidenciação de eventos controlados e em processo, com conjuntos de dados atualizados em tempo real em ambiente controlado.

Contudo, a pandemia de Covid-19, descreve um fenômeno de evento em processo em ambiente não controlado, ou parcialmente controlado, baseados em velocidade, imprevisibilidade e leitos disponíveis. Giannella e Velho, fizeram uma nova abordagem de visualização de dados hipotéticos, evidenciando eventos concretos usando conceitos e modelos matemáticos em simulações para chegar ao gráfico de achatamento da curva com o uso da função exponencial, da função logarítmica que é inversa da função exponencial e da função Logística, que apresenta o decrescimento até se chegar a um patamar aceitável, utilizando o modelo SIR (Susceptíveis, Infectadas e Removidas) em equações não lineares.

Farias *et al.* (2020) descrevem a utilização da Regra de Bayes para representar um número esperado de novos casos infecciosos secundários gerados a partir de um caso infectado em uma população suscetível, a partir do número de contatos que cada

indivíduo infectado realiza com outros indivíduos em uma unidade de tempo médio de duração do processo infeccioso.

### 3. METODOLOGIA

#### 3.1 Pesquisa inicial e tecnologias

O sistema efetuará predições sobre o avanço da doença em Londrina, com dados comparativos e informativos dos casos confirmados na região, além de efetuar uma análise em rede social de pessoas na cidade. A partir destas informações, serão feitas as predições do sentimento geral da cidade ao enfrentar a pandemia e o isolamento social.

Com a intenção de coletar dados verídicos e estáveis sobre a região de Londrina, utilizou-se a plataforma SESA (Secretaria do Estado da Saúde), que disponibiliza as informações oficiais sobre a Covid-19. O órgão atende a região pertinente e a estrutura dos dados foram considerados de fácil utilização e disponibilizados diariamente.

Tratando-se de aplicações web e mobile, muitas tecnologias estão disponíveis para uso no mercado e no meio acadêmico. Para o armazenamento de dados, foi utilizado o banco de dados MongoDB, por se tratar de um banco NoSQL e permitir o armazenamento de dados diretamente no formato JSON (*JavaScript Object Notation*), o qual permite que os dados sejam acessados e manipulados em quaisquer aplicações, independente da tecnologia e da linguagem de programação.

Para *back-end*, foi utilizada a linguagem de programação Python, devido a disponibilidade de bibliotecas para operações e cálculos matemáticos com predição de dados por meio de regressão linear, além de possuir recursos para realizar a análise dos sentimentos da população. A conexão entre o banco de dados MongoDB com a aplicação Python foi realizada com a biblioteca Pymongo.

Na parte de *front-end* web foi utilizado JavaScript, HTML, CSS e React JS para construção das páginas, além de componentes como React Google Charts para criação dos gráficos. O ambiente de execução NodeJS foi aplicado para subir a aplicação e realizar testes. Para aplicação mobile, foi utilizado o React Native para o desenvolvimento das telas e das funcionalidades e a plataforma Expo para a simulação da aplicação. A comunicação entre o *front-end* e o *back-end* foi realizada

por meio de uma arquitetura API Rest, que permite com que tanto o *front-end* web como o mobile sejam integrados com o mesmo *back-end*. Esta integração com a API foi realizada por meio da biblioteca Flask Cors.

### 3.2 Documentação e planejamento de desenvolvimento

Baseado nas informações contidas nos dados SESA-PR, foi definido as seguintes funcionalidades do sistema que são apresentados por meio de gráficos na forma de dashboard:

- **Representação de avanço de casos (confirmado, recuperado e óbito) e predição de óbito:** esta funcionalidade da aplicação é responsável por exibir os registros de casos confirmados, recuperados e óbitos na cidade de Londrina, além de utilizar algoritmo de *Machine Learning* para predizer os casos de óbitos da cidade.
- **Representação de afetados agrupado por sexo masculino ou feminino:** por meio de gráfico, representa um agrupamento de casos confirmados filtrados pelo sexo juntamente com atendimentos efetuados na cidade de Londrina.
- **Relação de atendimentos de pacientes agrupado por cidade:** exibe um agrupamento de casos confirmados, filtrados pela cidade de residência do infectado juntamente com atendimentos efetuados na cidade de Londrina.
- **Representação de infectados, agrupados por idade:** demonstra o agrupamento de casos confirmados, filtrados pela idade juntamente com atendimentos efetuados na cidade de Londrina.
- **Análise de sentimentos:** exibe resultado de análise de sentimento, executado por algoritmo de *Machine Learning* que classifica os posts do Twitter em "Positivo", "Negativo" e "Neutro" filtrados pela região do Brasil.

As seguintes funcionalidades, foram separadas com foco na comunicação e apresentação de informações sobre a estrutura do sistema ao usuário.

- **Quadro de comunicado:** esta funcionalidade tem por objetivo, exibir comunicados da aplicação para o usuário, trazendo informações e esclarecimentos da falta de dados, informações ou erros.

- **Informações do projeto:** tem como objetivo explicar ao usuário, como os dados exibidos no dashboard foram obtidos.

Para detalhar como deveria ser feito o desenvolvimento, processos e tecnologia a ser seguida durante desenvolvimento, foi construído documentos para cada caso de uso do sistema, onde havia descrições do objetivo do caso, público alvo, atores, descrição, pré-condições, fluxos básicos, fluxos alternativos, fluxos de exceção, pós-condição, requisitos funcionais e não funcionais. Para ter uma visão geral da aplicação foi construído um diagrama de caso de uso.

#### 4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Antes de realizar processamento e armazenamento dos dados, foi necessário efetuar tratamento das informações contidas nos arquivos disponibilizados pelo SESA. Este tratamento permitiu filtrar informações irrelevantes para aplicação, como dados de outras cidades e dados sem possibilidade de uso.

A estrutura das informações em formato CSV (*Comma-separated values*), que ao longo da pandemia tiveram alterações, necessitaram de implementação de código para tratar essas alterações, mapeando e localizando as informações relevantes. Todo processo de coleta foi executado manualmente, por meio do *download* os arquivos CSV do site da SESA, renomeando de acordo com sua data no padrão dd\_MM\_yyyy (dd=dia, MM=mês, yyyy=ano). Por exemplo, um arquivo disponibilizado no dia 04 de Dezembro de 2020 recebe o nome de "04\_12\_2020".

Um dos arquivo disponibilizado pela plataforma que foram processados foi o que representava os casos e óbitos (Figura 1), que é utilizado no processamento de registro de casos confirmados, recuperados e óbitos pelo Covid-19. Outro arquivo processado é o que mostra os dados gerais, onde é trabalhado os dados de casos confirmados agrupado por idade, sexo e cidade.

Figura 1 – Arquivo CSV de casos e óbitos.

```

IBGE;RS;MACRO;Município;Casos;Óbitos;Recuperados;Em_investigacao
4100103;18;NORTE;Abatiá;93;1;86;0
4100202;02;LESTE;Adrianópolis;191;4;168;0
4100301;02;LESTE;Agudos do Sul;122;1;77;0
4100400;02;LESTE;Almirante Tamandaré;2728;65;719;14
4100459;11;NOROESTE;Altamira do Paraná;21;2;16;1
4100509;12;NOROESTE;Altônia;237;3;173;1
4100608;14;NOROESTE;Alto Paraná;196;1;146;0
4100707;12;NOROESTE;Alto Piquiri;124;4;108;2
4100806;17;NORTE;Alvorada do Sul;150;5;7;0

```

Fonte: O autor (2020).

Através da API Tweepy, uma biblioteca do Python, foram pesquisados tweets filtrados pela região do Brasil, com palavra-chave “covid” e incluiu-se o parâmetro “extended” no “tweet\_mode”, para que seja retornado toda mensagem da postagem. O retorno da API é anexado e armazenado no banco de dados.

Para gerar os dados de evolução da doença na cidade contendo os casos de confirmados, óbitos e recuperados, foram utilizados todos dados já salvo no banco referente ao arquivo “Casos e Óbitos”. Estes dados, são chamados pelo *front-end* que interage com a API, buscando os dados salvos no banco e retornando no formato JSON, onde o front-end carrega e gera os gráficos utilizando React Google Charts.

Para gerar os dados de predição de óbito, foi utilizado a biblioteca Scikit Learn, onde consumiu-se todos os dados existentes de casos confirmados (variável independente) e óbitos (variável dependente) registrado no sistema, conforme mostra a Figura 2. Estes dados foram passados no modelo de regressão linear como dados para treino e teste. Os dados de predição, foram preditos por uma média de evolução de casos confirmados dos últimos cinco dias, que foram acrescentados a partir do último registro, simulando a evolução dos casos nos dias seguintes. Os dados simulados foram passados como parâmetro no modelo de regressão (variável independente) já treinado que retornou a predição de óbito para cada dia. Estes dados foram salvos no banco e ficam disponível para consumo do front-end através da API que segue o mesmo processo descrito na comunicação entre API e *front-end*.

Figura 2 – Evolução dos casos.

```

},
{
  "Confirmado": 13846,
  "Data": "11142020",
  "Obito": 320,
  "Recuperado": 10375
},
{
  "Confirmado": 13907,
  "Data": "11152020",
  "Obito": 320,
  "Recuperado": 10392
},
{
  "Confirmado": 13960,
  "Data": "11162020",
  "Obito": 321,
  "Recuperado": 10419
},
{
  "Confirmado": 14098,
  "Data": "11172020",
  "Obito": 321,
  "Recuperado": 10568
}
]

```

---

Fonte: O autor (2020).

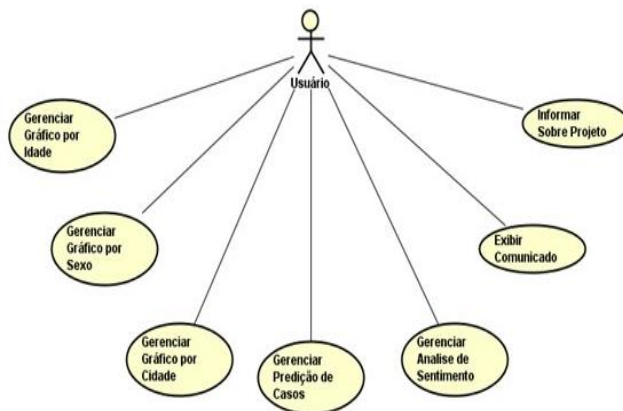
Os dados de casos confirmados classificados pelo sexo, foram obtidos através de uma consulta no banco referente aos dados salvos do tipo “Geral”, onde foram contabilizados os valores de “M” para masculino e “F” para feminino. Após totalizar os valores, foram anexados e salvos no banco, ficando disponível para consumo.

Referente aos casos confirmados agrupados pela idade, seguiu-se o mesmo processo dos casos classificados pelo sexo, porém são contabilizados o número de registros dentro das faixas 0-10, 11-17, 18-28, 29-40, 41-59 e maior que 60 anos. Para os dados do tipo atendimento por cidade, utilizaram-se registros salvos no banco referente aos dados “Geral”, a contabilização dos dados leva em consideração o município de atendimento, que no caso foi a cidade Londrina e o município de residência. Por fim, os dados foram salvos no banco de dados e disponíveis para o consumo.

As análises de sentimentos, utilizam-se os tweets salvos no banco que processam uma segunda camada de filtro. Primeiramente, foram retirados links, pontos e vírgulas das mensagens, já que estes elementos são considerados resíduos para análise e posteriormente são retirados artigos definidos e indefinidos das frases. Os resultados foram manipulados com a biblioteca TextBlob, que por característica, realiza processamento computacional de linguagens naturais, no qual os dados são

traduzidos para idioma inglês para que funcione com a biblioteca. Para cada texto, é chamado a função de análise de sentimento do TextBlob, que retorna a polaridade da mensagem com valores entre 1, 0, -1, onde 1 é considerado positivo, -1 negativo e 0 neutro. A polaridade da mensagem foi salva no banco e por fim, é realizado a contabilização, ficando disponível para consumo.

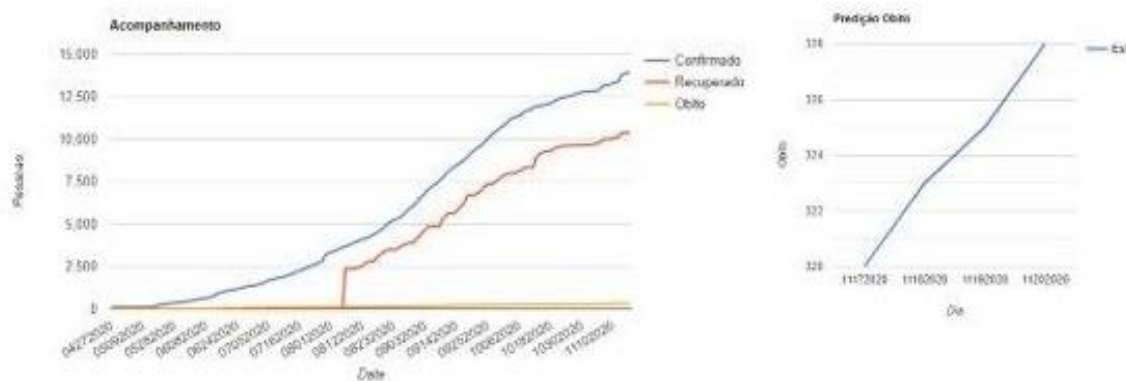
Figura 3 – Diagrama de casos de uso do sistema.



Fonte: O autor (2020).

Após todo processo de desenvolvimento, este projeto chegou ao resultado proposto, cumprindo todos os requisitos e seguindo toda documentação estabelecida. Os requisitos funcionais estão sendo ilustrados no diagrama de casos de uso da Figura 3, um exemplo de um gráfico de acompanhamento da Covid-19 e a respectiva predição na Figura 4 e um dashboard da aplicação *web* e *mobile* na Figura 5.

Figura 4 – Gráfico de acompanhamento dos casos confirmados, ocorrência de óbitos e recuperações, junto com a predição realizada.



Fonte: O autor (2020).

Figura 5 – Dashboard da aplicação *web* (à esquerda) e *mobile* (à direita).



Fonte: O autor (2020).

Pode-se observar na Figura 5 que na cidade de Londrina, o número de casos de óbitos representa em torno de 2% dos registros totais e que a faixa etária com maior número de infecção estão entre 41 e 59 anos. O sexo feminino possui maior número de casos confirmados.

## 5. CONSIDERAÇÕES FINAIS

De acordo com a pesquisa realizada neste projeto, as estimativas relacionadas aos avanços da Covid-19 tiveram suas premissas em cálculos estatísticos matemáticos. O desafio foi aplicar tecnologias digitais, incluindo *Machine Learning*, inteligência artificial e finalizando com aplicação *web* e *mobile*, a fim de disponibilizar a população os aplicativos para os usuários.

A qualificação de dados, a vulnerabilidade populacional e diferentes ações dos gestores públicos, tornaram desafiadora a utilização da predição por algoritmos de aprendizagem de máquina. É importante frisar que para melhor eficácia, há a dependência de grande volume de dados para o aprendizado, ausentes no início desta pandemia.

Por fim, apesar do momento difícil na história da humanidade, a necessidade de devolver soluções na velocidade necessária, fez com que esse surto de Covid-19 proporcionasse um grande avanço no desenvolvimento intelectual dos cientistas e

pesquisadores, um avanço muito positivo. Certamente, esses esforços já trouxeram benefícios e deverá acelerar a adoção de soluções digitais na saúde e principalmente, salvar vidas.

## REFERÊNCIAS

FARIAS, C. M.; MEDRONHO, R. A.; TRAVASSOS, G. H. Avaliação do comportamento da Covid-19 no estado do Rio de Janeiro e seus municípios com base em R(o) calculado a partir das evoluções anteriores de R dos casos notificados à Secretaria de Estado de Saúde-RJ. **Grupo de Trabalho Multidisciplinar da UFRJ sobre a Coronavirus Disease 19 (COVID-19)**. Rio de Janeiro, 2020, 18p.

FERNANDES, J. L. Covid-19 no Brasil: aprendendo a andar no escuro sem deixar nada para trás. **Arquivos Brasileiros de Cardiologia**, v. 114, n. 6, jun. 2020.

G1. Entenda os impactos da pandemia de coronavírus nas economias global e brasileira, 2020a. Disponível em: <<https://g1.globo.com/economia/noticia/2020/02/26/entenda-os-impactos-do-avancodo-coronavirus-na-economia-global-e-brasileira.ghtml>>. Acessado em: 26 ago. 2020.

GIANNELLA, J.; VELHO, L. Visualização em Tempo de Conora Vírus – parte 1. **Laboratorio VISGRAF**: Instituto de Matemática Pura e Aplicada. Rio de Janeiro, 2020, 23p.

MINISTÉRIO DA SAÚDE (Brasil). Sobre a Doença: O que é Covid-19. [S. I.], 2020a. Disponível em: <<https://coronavirus.saude.gov.br/sobre-a-doenca#o-que-e-covid>>. Acessado em: 26 ago. 2020.

MINISTÉRIO DA SAÚDE (Brasil). Painel Coronavírus [S. I.], 2020b. Disponível em: <<https://covid.saude.gov.br>>. Acessado em: 26 ago. 2020.

NEVES, K. D; BARROS, N. Análise preditiva de casos confirmados de Covid-19 no brasil e em oito países baseada no modelo não linear de Gompertz. **SciELO Peru**, v. 202, 2020.

SECRETARIA DA SAÚDE (Paraná). Boletim – Informe Epidemiológico Coronavírus (COVID-19) – Arquivos CSV (Dados Abertos) [S. I.], 2020. Disponível em: <<http://www.saude.pr.gov.br/Pagina/Coronavirus-COVID-19>>. Acessado em: 26 ago. 2020.