

Projection of Covid-19 in state of Minas Gerais using artificial neural network

O. Fernanda², M. Camila¹, O. Filipe³, Edyene Oliveira¹

¹*Centro Universitário UNA*

Rua dos Guajajaras, 175 - Centro, Belo Horizonte - MG, 30180-100

²*Universidade Federal de Minas Gerais*

Av. Pres. Antônio Carlos, 6627, Pampulha, Belo Horizonte - MG, 31270-901

³*Centro Universitário Newton Paiva*

Rua Paulo Piedade Campos, 420, Buritis, Belo Horizonte - MG, 30494-060

camilacmariano@hotmail.com, Fernandakmo@gmail.com, Felps931@gmail.com

Abstract. Coronavirus is a disease that, since the end of 2019, has been contaminating people in all countries and continents. As a result of its appearance, scientists and researchers around the world are conducting studies and disseminating information from data on virus behavior in the human body to data on contagion and prevention. These research have been decisive in the fight against Covid-19, because according to WHO (World Health Organization) the tendency is that the virus continues to circulate for some years. Therefore, the exhibition of auxiliary information helps people and governments to take actions to reduce Contagion. Thus, the present work aims to project the impact of the pandemic in the State of Minas Gerais, considering the government data to try control the pandemic, such as periods when there was social isolation and social flexibility. For the realization of this project, the use of Artificial Neural Networks (ANN) was made. To train the ANN, data from public government databases, both national and international, were used. This data was pre-processed using Python and Matlab programming language. In addition, they were carried out in the comparative project between the ANN algorithms, aiming to find the best performance among them.

Keywords: first palavra-chave, second keyword, third keyword (até 5 palavras-chave).

1 Introdução

Desde março de 2020 estamos lidando com a pandemia do Corona Vírus no Brasil. Trata-se de uma doença respiratória que pode levar à morte por várias razões, entre elas o comprometimento do pulmão. A OMS (Organização Mundial da Saúde) declarou essa doença como uma pandemia e, desde então, o mundo enfrenta uma emergência de saúde pública, sendo o Brasil, atualmente, o segundo país com maior número de pessoas infectadas e mortas [1].

Devido à quantidade de pessoas doentes ao mesmo tempo acabou sobrecarregando o sistema de saúde, fazendo com que o Ministério da Saúde recomendasse que apenas os casos mais graves fossem testados e tratados na área hospitalar. Em Minas Gerais, o Governador do Estado decretou estado de calamidade pública em 20 de março até o final de 2020¹. Atualmente, Minas Gerais é o segundo estado com a menor taxa de infectados do país.

A evolução das infecções tem variado de acordo com as respostas dos governos mediante a crise, em alguns casos parecendo estar contida e em outros atingindo proporções gravíssimas [2]. A disseminação da doença ocorre por meio de contato entre pessoas, o que requer a construção de planos rígidos de enfrentamento e contingência. Deste modo, torna-se essencial conseguir prever os casos estimados, a fim de auxiliar planos de contenção e proteção necessários.

Entre os métodos utilizados para acompanhar o estado da doença no país, a inteligência artificial é uma delas. O presente trabalho propõe o uso de Redes Neurais Artificiais para projeção do número de casos de contágio e óbitos causados por COVID-19. O estudo foi realizado no estado de Minas Gerais, região sudeste do Brasil, a partir de dados públicos do governo.

2 Trabalhos correlatos

A pandemia do Corona vírus gerou inúmeros trabalhos de pesquisas, sobretudo pelo fato de ser uma doença desconhecida. Existem vários trabalhos relacionados à previsão de Covid-19, porém, a maioria com enfoque no Brasil e no mundo.

No caso específico de utilização de redes neurais artificiais foi encontrado um trabalho de Jonas et al [3]. Neste artigo os autores implementaram um modelo de rede neural para prever contágios e mortes no estado do Pará, Brasil.

O autor Andre Luiz et al [4] realizaram pesquisa usando modelos de regressão não lineares na análise de dados de óbitos por Covid-19 nos estados de Pernambuco e Ceará. Segundo os autores, embora o número de mortes ainda não estabilizou, os modelos implementados alcançaram bons ajustes ao serem comparados com o valor real na época da escrita do artigo.

Outros autores como Nadia et al [5] fizeram uso de modelos matemáticos para preverem novos casos de Covid-19 em Minas Gerais. Os autores levaram em consideração variáveis como diferentes taxas de isolamento social. A previsão deles foi realizada no período de 30 dias.

¹ <http://www.agenciaminas.mg.gov.br/noticia/romeu-zema-decreta-calamidade-publica-estadualiza-medidas-e-fecha-divisas>

3 – Referencial Teórico

3.1 – Regressão Linear

Regressão linear é um modelo utilizado com o objetivo de se prever uma variável dependente de saída tendo previamente os dados independentes de entrada. É utilizada em diversos projetos em que se é necessário ter a estimativa de algum valor, como por exemplo a previsão do limite de cartão de crédito de alguém em vista de seu salário, ou a altura de uma criança baseado na altura de seus pais [6].

Isso é feito utilizando a expressão matemática indicada na equação 01. A variável y é o valor de saída que se deseja prever, na equação é a variável dependente. A variável x representa o conjunto de dados ou características, logo, é uma variável independente. Os coeficientes angular e linear a e b tem como objetivo auxiliar a preservação da relação entre x e y . Na Equação 1 pode ser vista forma matemática da regressão linear simples.

$$y = a * x + b \tag{1}$$

Entretanto, esta abordagem não é indicada para projetos que apresentam comportamento não linear. Para esses, deve-se fazer uso de um modelo que apresenta mais coeficientes em cada parcela, caracterizando uma regressão polinomial [7]. A Figura 01 demonstra o comportamento do número de óbitos causados pelo Covid-19 no estado de Minas Gerais até 31 de julho. É possível observar que este gráfico apresenta um comportamento curvilíneo, portanto, visando diminuir o erro, foi utilizado neste trabalho a regressão polinomial que correspondeu melhor ao comportamento da doença.

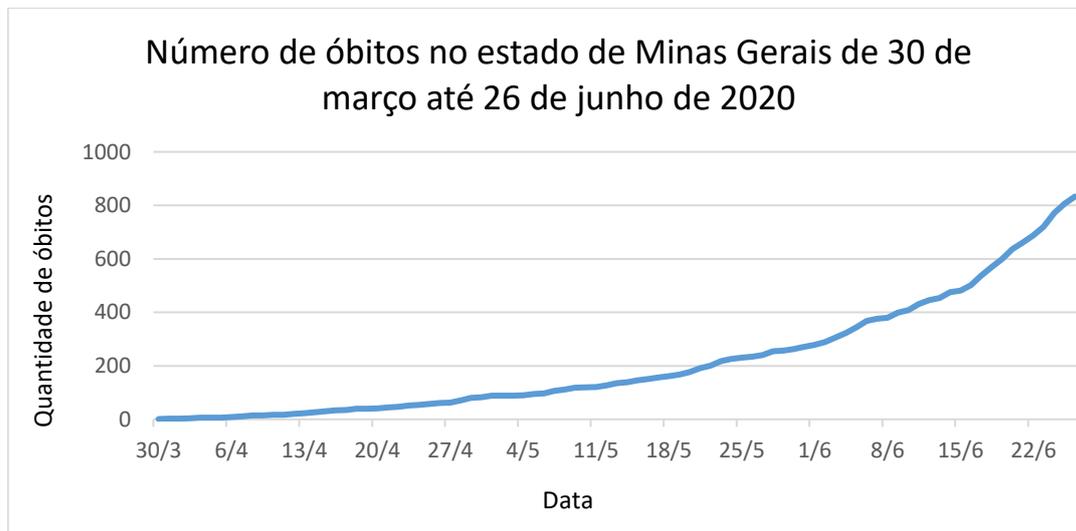


Figura 01 – Número de óbitos registrados pelo governo de Minas Gerais no período entre 30/03/2020 e 22/06/2020.

3.2 – Regressão Polinomial

A regressão polinomial tem o objetivo de prever uma variável dependente de saída tendo previamente

os dados independentes de entrada. Entretanto, ela é capaz de estimar um valor que segue um comportamento não linear [7].

A regressão polinomial é frequentemente usada também como um meio de testar a não linearidade na relação entre X e Y. Como os polinômios podem descrever uma quantidade ampla de curvas, um teste de não linearidade pode ser conduzido determinando se a adição de potências sucessivas ou conjuntos de potências de X melhoram o ajuste do modelo em um grau estatisticamente significativo [8].

A Equação 02 demonstra a função de um modelo de polinômio de grau 4, em que a variável y é o valor de saída que se deseja prever, a variável “ x ” são os dados de entrada e os coeficientes a, b, c, d , e são os valores que se busca encontrar que irão melhor preservar a relação entre x e y .

$$y = a * x + b * x^2 + c * x^3 + d * x^4 + e \quad (2)$$

3.3 – Função de custo

A função custo é uma função que tem como objetivo calcular o quão longe está o resultado desejado em relação ao resultado calculado pela rede neural. Essa discrepância de valores ocorre devido ao erro que se obtém durante o aprendizado da rede, em que os pesos que serão multiplicados com a entrada são inicialmente aleatórios. Dessa forma, a função custo irá atuar calculando a diferença entre o valor alvo e o valor real. Com isto, o objetivo do modelo é encontrar o menor valor de função de custo possível, atingindo assim o valor de saída mais próximo do desejado [9].

Neste projeto, foram utilizadas as funções de custo Mean Squared Error (MSE) e Mean Absolute Error (MAE). O MAE, como dita o próprio nome, calcula o erro absoluto que foi obtido. Já o MSE calcula a média de erros da rede neural elevado ao quadrado. Dessa forma, os valores de erro que são muito elevados poderão ser colocados em evidência devido a sua parte quadrática [10]. A Equação 02 demonstra a função de perda MSE.

$$custo = MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

onde:

- $\frac{1}{n} \sum_{i=1}^n$ - É a média dos valores
- y_i - É o valor previsto pela rede neural
- \hat{y}_i - É o valor alvo

4 – Metodologia

Este trabalho teve como principal objetivo a criação de uma inteligência artificial capaz de prever a quantidade de casos e de óbitos futuros causados pelo Coronavírus da população do estado de Minas Gerais. Para tal, foi utilizada a linguagem de programação Python inserida no ambiente “Anaconda”, em que se fez

uso da regressão polinomial para fazer a previsão dos casos de Corona vírus.

Os dados de entrada utilizados para o treinamento do modelo foram retirados a partir de bases disponibilizadas em fontes do governo². Entretanto, esses dados tiveram que ser devidamente pré-processados antes de serem utilizados no programa.

Conforme pode ser observado na Tabela 01, para este projeto foram utilizadas duas colunas: A data dos casos como sendo o conjunto de entrada (a esquerda), e, para a predição do número de óbitos, foram utilizados a quantidade de mortos até o dia 26 de junho de 2020 (a direita).

O conjunto de datas foi remodelado para uma sequência de 0 a 88, representando respectivamente o dia 30 de março até o dia 26 de junho. Foi utilizado o mesmo raciocínio para a predição do número de casos.

Tabela 01 – Exemplo do conjunto de dados de entrada

Dia	Número de óbitos
84	688
85	720
86	771
87	806
88	833

A regressão polinomial foi implementada utilizando a biblioteca de código aberto “Scikit Learn” do Python. O polinômio utilizado no projeto que apresentou melhor resultado foi de grau 4, este se adequou ao comportamento da doença em questão. Portanto, a definição foi realizada de forma empírica.

Para a previsão do número de casos futuros, foi criado um vetor com apenas o dia ou dias adiante a partir da última data do treinamento. O modelo foi treinado com os dados correspondentes ao período entre 30/03/2020 a 26/06/2020.

Dessa forma, se o objetivo fosse prever o número de casos de Coronavírus para 15 de agosto, o valor a ser passado para formação do vetor em questão seria de 50. Esta é a metodologia, pois passaram-se 50 dias desde a data que foi usada para o treinamento da regressão, ou seja, 26/06/2020.

² <https://covid.saude.gov.br/>

5 – Resultados

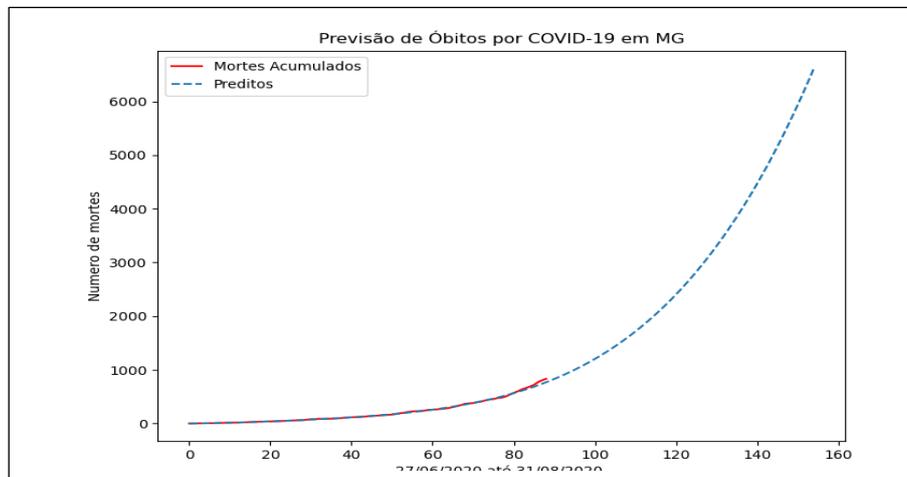


Figura 2 – Resultado da previsão de óbitos em MG no período entre 27/06 e 31/08

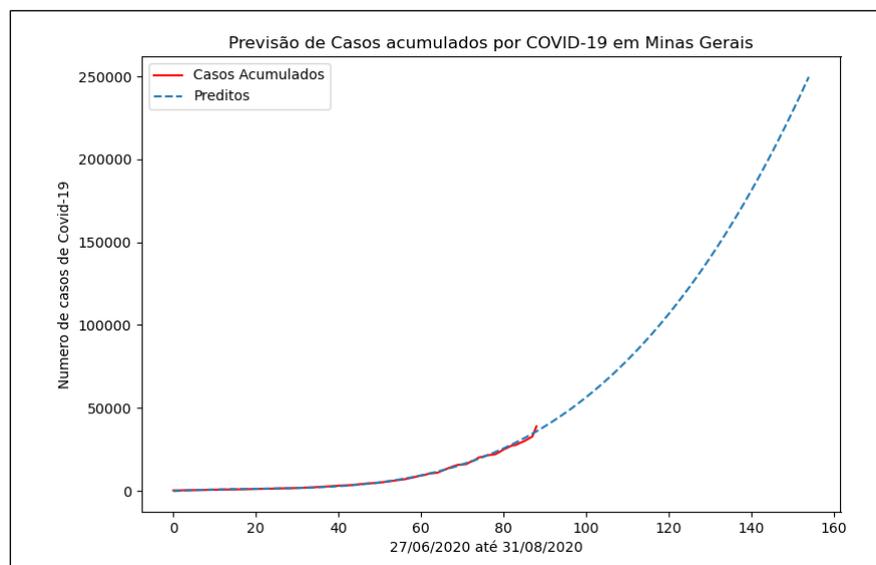


Figura 3 – Resultado da previsão de casos de Covid-19 em MG no período entre 27/06 e 31/08.

A Figura 2 mostra um exemplo de previsão de óbitos realizado pelo modelo implementado. A previsão foi realizada a partir de 27/06/2020 a 31/08/2020, ou seja, um período de 66 dias. Do início dos registros da pandemia no Brasil que foi 30/03/2020 até a data de 26/06/2020 somaram 88 dias.

O resultado da previsão de óbitos previsto foi acima do real. Em 31/08/2020 o número real registrado pelas autoridades foi de 5.335 enquanto o modelo previu 6.613 óbitos por Covid-19. A diferença foi de 1.278 sendo necessário rever o modelo.

A Figura 3 mostra previsão de pessoas contaminadas pelo Covid-19 no mesmo período, ou seja, entre

27/06/2020 e 31/08/2020. Os resultados mostraram previsão acima do real, em 66 dias o modelo previu 249.643 enquanto o real registrado pelas autoridades em 31/08/2020 é de 216.557 óbitos ³.

Nos dois casos o modelo realizou previsão acima do real tanto em casos de óbitos quanto em casos de contaminados pelo Covid-19. Esse comportamento pode ter ocorrido por dois fatores sendo, por exemplo, atuação por parte dos governantes, como, isolamento social, conscientização de medidas de higiene e uso de máscaras de proteção.

Além disso, o modelo utilizado é exponencial, logo seria interessante para trabalhos futuros implementar modelos que trabalham com dados temporais. Um destes modelos é a rede neural LSTM (*Long short-term memory*). Outro método que pode ser aplicado são funções de surto, ou também chamadas de função pulso, que são conhecidas por descrever o comportamento de pandemias.

Referencias

- [1] Y. R. Guo *et al.*, "The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status," *Military Medical Research*. 2020.
- [2] L. Enjuanes, "Coronavirus replication and reverse genetics: Preface," in *Current Topics in Microbiology and Immunology*, 2005.
- [3] "Redes neurais artificiais na previsão de contágio e óbitos por covid-19: um estudo noestado do Pará, Brasil," *IJDR Int. J. Dev. Res.*, pp. 35416–35421, 2020.
- [4] A. L. P. dos Santos *et al.*, "Análise e previsão da evolução do número de óbitos por COVID-19 do estado de Pernambuco e Ceará utilizando modelos de regressão," *Res. Soc. Dev.*, 2020.
- [5] N. G. Sousa, A. de O. Cardoso, R. F. Cardoso, and A. G. Utsumi, "Análise da dinâmica de transmissão da COVID-19 em Minas Gerais: Modelagem e Simulação," *Res. Soc. Dev.*, 2020.
- [6] I. N. Silva, D. H. Spatti, and R. A. Flauzino, *Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas*. 2010.
- [7] C. M. Bishop, *Machine Learning and Pattern Recoginiton*. 2006.
- [8] R. B. Darlington, *Regression and linear models*. 1990.
- [9] S. Haykin, "Redes neurais: princípios e prática," *Bookman*, 2017.
- [10] T. M. Mitchell, "Machine Learning 10-601," *Mach. Learn.*, 2011.

³ http://coronavirus.saude.mg.gov.br/images/Boletim-Epidemiologico_COVID-19_31.08.2020.pdf