

# **Aplicação de Técnica de Machine Learning para previsão de casos de Covid nos próximos anos**

Viviane Ferreira de Oliveira<sup>1</sup>

## **INTRODUÇÃO**

O mundo presencia, desde janeiro de 2020, a pandemia do Covid-19, que desde então constitui uma Emergência de Saúde Pública de Importância Internacional (OMS, 2020). Contudo, os indicadores do Covid-19 mostram que atualmente há uma tendência de redução da circulação da doença, graças, sobretudo, ao uso das vacinas. De acordo com Tedros Ghebreyesus, diretor geral da Organização Mundial da Saúde, o fim da pandemia está próximo, com isso, afirma que o mundo nunca esteve em “melhor posição para acabar com a pandemia”, porém, destaca que os governos não devem desacelerar as ações contra a proliferação do vírus. Dessa forma, muito se pensa sobre o que poderá ocorrer com a pandemia do Coronavírus no futuro, para tanto, muitas pesquisas a respeito da predição de cenários com a utilização de dados relativos ao Covid-19, foram e estão sendo desenvolvidas, tais como “Utilização de aprendizagem de máquina para predição do diagnóstico do Covid-19” (BORTOLINI, VINÍCIUS, 2020), “Predição de casos de COVID-19 nos municípios de santa catarina utilizando redes neurais recorrentes” (Vianna, L.S., Busana, J.A., 2021) e “Estimação e predição dos casos de COVID-19 nas metrópoles brasileiras” (Sousa G.J.B et al, 2020).

## **MINERAÇÃO DE DADOS**

A mineração de dados (Data Mining) pode ser entendida como o processo de encontrar padrões e correlações em grandes conjuntos de dados para prever resultados. Tem sua base fundamentada em três disciplinas interligadas, a estatística, a inteligência artificial e o machine learning. Ademais, Azuaje (2006), apresenta a mineração de dados como um instrumento que tem como objetivo principal, "identificar padrões e relacionamentos informativos relevantes, interessantes e potencialmente novos em grandes conjuntos de dados para apoiar a tomada de decisões e a descoberta de conhecimento".

A análise preditiva consiste na aplicação de algoritmos para entender a estrutura dos dados existentes e gerar regras de previsão. Esses algoritmos podem ser usados em um cenário não supervisionado, no qual apenas preditores (covariáveis) estão disponíveis em um conjunto de dados, ou em problemas supervisionados, quando,

---

<sup>1</sup>Viviane Ferreira de Oliveira. Acadêmico do Curso de Sistemas de Informação. Centro Universitário Paraíso (UnifAP). Rua da Conceição, 1228 - São Miguel - CEP: 63.010-465 - Juazeiro do Norte, CE. E-mail: vivianeferreira18@aluno.fapce.edu.br

além dos preditores, também está disponível uma resposta de interesse, responsável por orientar a análise (SANTOS, H.G.D. et al, 2019). Isto posto, métodos preditivos desempenham um papel de grande significância para que se possa ter uma antecipação e até mesmo prevenção de eventos.

Historicamente, alguns modelos foram desenvolvidos para prever a ocorrência de eventos de interesse para a saúde da população, geralmente derivados de modelos lineares, considerados algoritmos de aprendizado de máquina mais simples. No Brasil, a utilização desses algoritmos em saúde pública ainda é incipiente (SANTOS, H.G.D. et al, 2019). Como exemplo, o estudo preditivo para diagnóstico de diabetes não diagnosticada, de Olivera et al., a partir de dados de 12.447 adultos, utilizando cinco algoritmos de machine learning (regressão logística, redes neurais, naive bayes, método dos K vizinhos mais próximos e random forest).

## **MACHINE LEARNING NO COVID**

O aprendizado de máquina ou machine learning (ML) tem como objetivo principal, construir um sistema de computador que aprenda com um banco de dados pré-definido e gere, ao final, um modelo de predição, previsão, classificação ou detecção.

O machine learning aplicado à saúde é um ponto muito importante quando se trata do bem estar da população, assim, é evidente que a cada dia, a tecnologia evolui e na área da saúde isso não é diferente, hoje podemos contar com sistemas que são capazes de diagnosticar, tratar e prevenir patógenos.

Com o início da pandemia iniciou-se um estudo pela equipe do Laboratório de Big Data e Análise Preditiva em Saúde (LABDAPS) da Faculdade de Saúde Pública da Universidade de São Paulo (USP), trabalho coordenado pelo Dr. Alexandre Dias Porto Chiavegatto Filho, no qual utiliza dados laboratoriais, clínicos e demográficos para treinar cinco algoritmos de aprendizado de máquina. Este estudo utilizou uma amostragem de 1.040 pacientes diagnosticado com Covid-19, e ao final do estudo foi dito que 70% dos pacientes foram utilizados para o aprendizado da máquina para que ela pudesse ter uma base de informações e com isso poder gerar os diagnósticos mais assertivos sobre os pacientes, foi aplicado ao sistema os 30% restantes, 25,5% foram enviados a UTI (Unidade de Terapia Intensiva), 10,2% necessitam de auxílio de ventilação mecânica intensiva, e 9,4% faleceram.

Segundo Fernando Fernandes (2020), cientista da computação e primeiro autor da pesquisa dos 5 algoritmos, os que tiveram maior aproveitamento no desempenho durante os testes foram o random forest (RF), o extra trees e o extreme gradient

boosting (XGBoost), “o algoritmo poderia ser aplicado, por exemplo, no pronto atendimento, onde já há coleta de dados mínima para uso da ferramenta, como a solicitação de hemogramas. Cabe ressaltar que o algoritmo, por si só, não toma a decisão, apenas fornece estimativas de risco para dar subsídios para melhores decisões. O médico terá sempre a palavra final”, comentou Fernando.

## **MÉTODOS DE MACHINE LEARNING**

A livre tradução significa aprendizagem de máquina, faz parte do conceito de inteligência artificial oferecendo um aprendizado a partir do contato com certo volume de dados e experiências. Quando um sistema está em conjunto com a tecnologia de Machine Learning, passa a ser capaz de identificar padrões, oferecer respostas e tomar decisões com o mínimo de participação humana, como cada sistema tem sua especificidade a tecnologia é dividida em diferentes formas de construção e aprendizado.

### **1. Aprendizado supervisionado:**

Sendo esse o que mais se aproxima do aprendizado humano e menos complexo, o sistema recebe um conjunto de dados já tratados, a partir disso ele aprende a fazer comparações e mostrar o “certo” e “errado”, o que permite com que ele aprenda da mesma forma que um humano faz uma tarefa.

### **2. Aprendizado não-supervisionado:**

Já dessa forma nosso sistema não recebe informações rotuladas, sendo ele o responsável por conferir e observar padrões nos dados, podendo fazer isso de várias formas, detectando anomalias ou clusterizando e procurando semelhanças em conjuntos.

### **3. Aprendizado semi-supervisionado:**

Utilizando essa técnica, o sistema consegue lidar com dados rotulados ou não, geralmente é aplicada quando o custo para rotular dados é alto.

### **4. Aprendizado por reforço:**

É baseado na lógica de tentativa e erro, dessa forma o sistema aprende quais ações são melhores para serem tomadas conforme seus erros, ou seja, o aprendizado é através de sua experiência.

# ALGORITMO DO MACHINE LEARNING

## 1. REGRESSÃO LINEAR

Esse algoritmo é chamado dessa forma porque existe uma reta traçada entre dois pontos no meio de um diagrama de dispersão, serve para achar uma relação entre esses pontos. A regressão Linear é dividida em dois tipos, sendo eles regressão linear simples e regressão linear múltipla:

### 1.1. Regressão linear simples

Modelo mais simples da regressão linear, utilizando apenas uma variável de desfecho e outra variável preditora, ela é utilizada quando é preciso descrever uma relação linear entre duas variáveis, por exemplo quando queremos prever o valor de uma variável pelo valor da outra.

### 1.2. Regressão Linear múltipla

Esta última é utilizada em praticamente 100% dos casos em relação à regressão linear simples, pois um modelo útil de Data Science segue lógicas multivariadas.

## 2. REGRESSÃO LOGÍSTICA

A regressão logística é um método usado para lidar com problemas de classificação, analisar diferenças de objetos para poder classificar essa diferença e falar onde ela melhor se encaixa. Esse método usa princípios de estatística, probabilidade e regressão logística, é dividido em três tipos, regressão logística binomial, regressão logística ordinal e regressão logística multinomial.

## **2.1. Regressão logística binominal**

Na regressão logística binomial os objetos são classificados em dois grupos, entre sim ou não, por exemplo quando alguém recebe uma ligação de um número que não está salvo nos contatos alguns celulares perguntam “Esse número é spam ou não ?” Se a resposta for sim, o sistema coloca aquele número numa lista de spam e o usuário não receberá mais ligações desse número. Caso a resposta seja não, o número poderá ligar normalmente mesmo não estando salvo como contato.

## **2.2. Regressão logística ordinal**

Esse método trabalha com o conceito de categorias ordenadas, podendo classificar os objetos em três ou mais classes, sendo elas boa, neutra ou ruim, por exemplo: o grau de satisfação do paciente com o atendimento é insatisfatório, satisfeito ou muito satisfeito.

## **2.3. Regressão logística multinomial**

No método de regressão logística multinomial os objetos são nomeados em três ou mais categorias, que não possuem ordem entre si, por exemplo: esse objeto é um lápis, caneta ou uma borracha? As regressões podem ser usadas em aplicações como:

- Pontuação de Crédito
- Medindo as taxas de sucesso de campanhas de marketing
- Prever as receitas de um determinado produto
- Prever se haverá um terremoto em um dia específico

### 3. ÁRVORES DE CLASSIFICAÇÃO E REGRESSÃO

A representação do modelo da árvore de decisão é uma árvore binária. Cada nó representa uma única variável de entrada (x) e um ponto de divisão nessa variável (assumindo que a variável seja numérica), os nós de cada folha da árvore tem uma variável de saída (y) usada na previsão, as previsões são feitas percorrendo as divisões da árvore até chegar a uma folha e gerar o valor da classe nessa folha.

As árvores são muito rápidas para fazer previsões. Elas também costumam ser precisas para uma ampla gama de problemas e não exigem nenhuma preparação especial para seus dados.

Como exemplo, podemos considerar uma amostra de 30 alunos de uma escola, com três variáveis: sexo (masculino ou feminino), classe (IX ou X) e altura (160 cm a 180 cm). Digamos também que dos 30 alunos, 15 deles jogam tênis no recreio. A partir disso, como podemos criar um modelo para prever quem vai jogar tênis durante o recreio? Neste problema, precisamos dividir os alunos que jogam tênis no recreio com base nas três variáveis à nossa disposição. Ela dividirá os alunos com base nos valores das três variáveis e identificará a variável que cria os melhores conjuntos homogêneos de alunos.

## REFERÊNCIAS

**HISTÓRICO da pandemia de COVID-19** - OPAS/OMS | Organização Pan-Americana da Saúde. Disponível em:

<<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>>. Acesso em: 30 set. 2022.

**OMS DIZ que fim da pandemia pode estar próximo.** Disponível em:

<https://news.un.org/pt/story/2022/09/1801061>. Acesso em: 30 set. 2022.

**O QUE é mineração de dados?** Disponível em:

<[https://www.sas.com/pt\\_br/insights/analytics/mineracao-de-dados.html](https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html)> Acesso em: 30 set. 2022.

**Algoritmo do machine learning**, disponível em :

<<https://ilumeo.com.br/todos-posts/2020/06/22/um-tour-pelos-10-principais-algoritmos-d-e-machine-learning>> acessado em: 03 de outubro de 2022.

Azuaje, F. Witten IH, Frank E: **Data Mining: Practical Machine Learning Tools and Techniques** 2ª edição. *Biomed Eng OnLine* 5, 51 (2006).

<https://doi.org/10.1186/1475-925X-5-51>

DA SILVA, D. **Machine learning: tudo o que você precisa saber sobre a tecnologia de aprendizagem de máquina**. Zendesk Português Zendesk, , 1 jan. 1970. Disponível em: <<https://www.zendesk.com.br/blog/machine-learning/>>. Acesso em: 30 set. 2022

SANTOS, H. G. D. et al. **Machine learning for predictive analyses in health: an example of an application to predict death in the elderly in São Paulo, Brazil**.

*Cadernos de saude publica*, v. 35, n. 7, p. e00050818, 2019.

**Modelos de aprendizado de máquina podem prever piora clínica na covid-19.**

Disponível em:

<[https://portugues.medscape.com/verartigo/6506437?src=mkm\\_ptmkt\\_210618\\_mscmrk\\_ptexcnws\\_nl&uac=292657PN&impID=3448072&faf=1](https://portugues.medscape.com/verartigo/6506437?src=mkm_ptmkt_210618_mscmrk_ptexcnws_nl&uac=292657PN&impID=3448072&faf=1)>. Acesso em: 1 out. 2022.