



PROJETO DE GRADUAÇÃO

**USO DE PREDIÇÃO PARA TRIAGEM DE
PACIENTES A PARTIR DO PADRÃO DE
SINTOMATOLOGIA ASSOCIADO À COVID-
19**

**SAMUEL DE MATOS PEREIRA
MATRÍCULA 15/0064730**

Brasília, 24 de setembro de 2022

UNIVERSIDADE DE BRASÍLIA

**FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

PROJETO DE GRADUAÇÃO

USO DE PREDIÇÃO PARA TRIAGEM DE PACIENTES A PARTIR DO PADRÃO DE SINTOMATOLOGIA ASSOCIADO À COVID-19

Por

SAMUEL DE MATOS PEREIRA
MATRÍCULA 15/0064730

Relatório submetido como requisito parcial para obtenção do grau de Engenheiro de Produção

Banca Examinadora

Prof. Ari Melo Mariano, Ph.D. - UnB/EPR
(Orientador)

Prof^ª. Dra. Márcia Terezinha Longen Zindel -
UnB/EPR

Prof^ª. MSc. Maíra Rocha Santos - UnB/FAV

Brasília, 24 de setembro de 2022

AGRADECIMENTOS

Agradeço a Deus por todas as bênçãos colocadas sobre mim durante minha jornada acadêmica e pessoal.

À minha mãe, Lília, por ter me apoiado em absolutamente todas as decisões que já tomei e em todos os momentos difíceis, por sempre me incentivar a buscar novos caminhos e sonhos. Obrigado por nunca deixar me faltar nada, educar-me da melhor maneira possível, orientar-me e amar-me incondicionalmente.

À minha namorada, Joyce, que sempre me apoiou na minha vida pessoal, acadêmica e profissional, principalmente nessa reta final.

Ao meu orientador, Prof. Ari Melo Mariano, Ph.D., por toda paciência que teve comigo, por acreditar e confiar no meu trabalho, e por sua enorme contribuição para minha formação e meu projeto de graduação.

A todos os colegas que fiz durante o curso e amigos mais próximos, por toda ajuda e momentos de descontração proporcionados, que amenizaram as etapas mais difíceis dessa longa caminhada.

Ao observatório de dados, por fornecer os dados necessários para a minha pesquisa.

À banca examinadora, por aceitar avaliar o meu trabalho.

A todos os professores e outras pessoas da UnB que contribuíram de alguma forma para minha formação em Engenharia de Produção.

RESUMO

De acordo com a literatura, vários modelos de predição de COVID-19 foram desenvolvidos como forma de apoio ao combate do Coronavírus que causou uma pandemia em 2020. Este estudo teve como objetivo apresentar uma proposta de triagem de pacientes a partir do padrão de sintomatologia associado à COVID-19. De modo a alcançar esse objetivo, foi realizada pesquisa do tipo exploratória quantitativa, por meio de um modelo de regressão logística binária, que foi desenvolvido seguindo o ciclo de vida de dados Crisp-DM utilizando o *software* RapidMiner. O algoritmo foi alimentado por uma base de dados secundária de 26.822 respostas referente à sintomatologia apresentada por pacientes da região sul do estado da Bahia, Brasil, infectados pela COVID-19. O modelo mostrou sensibilidade de 99,93% para detectar a COVID-19, a partir de nove sintomas impactantes para o resultado: confusão, saturação de O₂<95%, falta de olfato, falta de paladar, expectoração/catarro no peito, tosse seca, fadiga, dispneia/falta de ar e rinorreia/coriza. Após a análise da relação entre a sintomatologia e o resultado positivo para COVID-19, foi proposta uma aplicação prática para o modelo, aplicando-o ao protocolo de Manchester, como método de triagem complementar.

Palavras-chave: COVID-19; Ciência de Dados; modelo de predição; regressão logística; Crisp-DM.

ABSTRACT

According to the literature, several COVID-19 prediction models were developed to support the fight against the coronavirus that caused a pandemic in 2020. This study aimed to present a triage proposal for patients based on the associated symptomatology pattern. to COVID-19. In order to achieve this objective, a quantitative exploratory research was carried out, using a binary logistic regression model, which was developed following the Crisp-DM data life using the RapidMiner software. The algorithm was fed by a secondary database of 26,822 responses regarding the symptoms presented by patients from the southern region of the state of Bahia, Brazil, infected by COVID-19. The model showed a sensitivity of 99.93% to detect COVID-19, from nine symptoms impacting the result, namely: confusion, O₂ saturation<95%, lack of smell, lack of taste, expectoration/phlegm in the chest, dry cough, fatigue, dyspnea/shortness of breath and rhinorrhea/coryza. After analyzing the relationship between the symptomatology and the positive result for COVID-19, practical applications were proposed for the model, in situations of diagnosis of other diseases, together with health professionals.

Keywords: COVID-19; Data Science; prediction model; logistic regression; Crisp-DM.

SUMÁRIO

1.	INTRODUÇÃO.....	10
1.1.	Problema da pesquisa.....	11
1.2.	Justificativa	11
1.3.	Objetivos	12
1.3.1.	Objetivo geral.....	12
1.3.2.	Objetivos específicos.....	12
1.4.	Estrutura dos capítulos.....	12
2.	TEORIA DO ENFOQUE META-ANALÍTICO	14
2.1.	Primeira etapa: preparação da pesquisa	14
2.2.	Segunda etapa: Apresentação e inter-relação dos dados	15
2.3.	Terceira etapa: Detalhamento, modelo integrador e validação por evidências.....	23
3.	REFERENCIAL TEÓRICO	31
3.1.	Indústria 4.0.....	31
3.1.1.	Os nove pilares da Indústria 4.0	31
3.1.2.	<i>Smart buildings</i>	34
3.2.	Data Science	35
3.3.	COVID-19.....	37
3.4.	Regressão logística	37
3.5.	Modelo Crisp-DM	38
4.	METODOLOGIA.....	40
4.1.	Tipo de pesquisa	40
4.2.	Local do estudo.....	40
4.3.	Objeto de estudo	40
4.4.	Instrumento de coleta de dados	40
4.5.	Amostra e critérios de inclusão e exclusão	40
4.6.	Ferramenta de análise e tratamento dos dados	41
4.7.	Estrutura da metodologia	41
5.	RESULTADOS E ANÁLISES	43
5.1.	Entendimento do negócio.....	43
5.2.	Compreensão dos dados.....	43
5.3.	Preparação dos dados.....	46
5.3.1.	Buscar valores faltantes	46
5.3.2.	Substituir dados faltantes	46
5.3.3.	Variável dependente	46
5.3.4.	Variáveis independentes	46
5.4.	Modelagem.....	46
5.5.	Avaliação	50

5.6. Implantação	51
6. CONSIDERAÇÕES FINAIS, LIMITAÇÕES E FUTURAS LINHAS DE PESQUISA .	54
REFERÊNCIAS BIBLIOGRÁFICAS	55
APÊNDICES.....	62
APÊNDICE A.....	62

LISTA DE FIGURAS

Figura 1: Modelo TEMAC	14
Figura 2: Publicações no decorrer dos anos	16
Figura 3: Publicação por país.....	17
Figura 4: Áreas que mais publicaram.....	18
Figura 5: Frequência de palavras-chave no tempo.....	19
Figura 6: Autores que mais publicaram.....	20
Figura 7: Autores mais citados	21
Figura 8: Mapa de calor das palavras-chave em títulos, resumos e palavras-chave	23
Figura 9: Mapa de calor de <i>co-citation</i>	24
Figura 10: Mapa de calor do <i>coupling</i>	25
Figura 11: Modelo integrador	30
Figura 12: Nove pilares da Indústria 4.0	32
Figura 13: Surgimento da Ciência de Dados	36
Figura 14: Etapas do Crisp-DM.....	39
Figura 15: Estrutura da metodologia.....	42
Figura 16: Impactos da COVID-19 no Brasil e na Bahia.....	43
Figura 17: Etapas do desenvolvimento do modelo de regressão logística	47
Figura 18: Subprocessos do modelo de regressão logística.....	48
Figura 19: Sintomas mais impactantes para o resultado positivo de COVID-19	49
Figura 20: Modelo de predição aplicado ao protocolo de Manchester	52

LISTA DE QUADROS

Quadro 1: Lista de artigos mais relevantes.....	25
Quadro 2: Principais modelos de predição encontrados.....	27
Quadro 3: Definição dos dados.....	45

LISTA DE TABELAS

Tabela 1: Resultado da regressão logística.....	49
Tabela 2: Matriz confusão	50

1. INTRODUÇÃO

Os avanços tecnológicos possibilitaram o desenvolvimento de máquinas e dispositivos que sequer eram possíveis há cinco anos atrás. A evolução da tecnologia possibilitou processamento e computação cada vez mais rápidos, além de permitir que dispositivos cada vez menores possuam mais memória por um custo menor (CHAPMAN *et al.*, 2015). Neste contexto, Schwab *et al.* (2016) discorre sobre o surgimento da Indústria 4.0, que se caracteriza por uma revolução digital na qual a internet torna-se muito mais móvel e ubíqua, através de sensores mais potentes, menores e mais baratos, por *machine learning* e inteligência artificial. Além disso, essa revolução integra três *clusters* principais: físico, digital e biológico.

Em conjunto com essa evolução tecnológica, as Tecnologias da Informação e Comunicação (ICTs) tiveram participação fundamental na melhoria da eficácia de vários processos relacionados à assistência médica, o que gerou o conceito de *eHealth* a partir das aplicações dessas ICTs (ACETO; PERSICO; PESCAPÉ, 2018). Desse modo, diagnosticar sintomas, encontrar profissionais especialistas da área de saúde e armazenar tais dados e informações se tornou cada vez mais prático devido à internet e ao uso das tecnologias *eHealth* e *mHealth* – dispositivos móveis de saúde, dispositivos de telemedicina, portais de saúde e bancos de dados para armazenamento de informações pessoais de saúde (HARRINGTON; KOON; ROGERS, 2020).

A Indústria 4.0 vem transformando as áreas de produção e de serviços por intermédio das várias tecnologias da informação e comunicação. No contexto da saúde, tecnologias como Internet das Coisas, Computação em Nuvem e Big Data estão direcionando a *eHealth* para a Healthcare 4.0 (AHMAD *et al.*, 2022). Ao mesmo tempo, a digitalização do sistema de saúde produziu grande quantidade de dados clínicos, intensificando o desenvolvimento da ciência de dados na medicina (SANCHEZ-PINTO; LUO; CHURPEK, 2018). Com esse cenário surgindo, a utilização de técnicas de *machine learning*, a partir do uso intensivo de dados, tornou-se presente em esferas científicas, tecnológicas e comerciais, possibilitando tomadas de decisão fundamentadas em evidências em âmbitos como a assistência médica (JORDAN; MITCHELL, 2015). A aplicação desses novos métodos analíticos viabilizou, também, avanços no desenvolvimento de modelos de assistência médica impulsionados por essa grande quantidade de dados e novas tecnologias (GINSBURG; PHILLIPS, 2018).

A saúde inteligente (*smart health*) compõe-se dessas várias tecnologias que buscam realizar monitoramento remoto e intervenções clínicas a partir de redes de sensores corporais e técnicas de análise de dados. Além disso, a saúde inteligente se respalda em várias tecnologias da Internet das Coisas (IoT) como ferramentas de computação, sensores e *hardwares* de rede e

comunicação (WANG *et al.*, 2015). Além disso, Islam *et al.* (2015) enfatiza o fato de a Internet das Coisas possuir muitas aplicações médicas, tais como monitoramento remoto de saúde, programas de condicionamento físico e atendimento a idosos, o que possibilita tratamentos e medicações realizados em casa, por meio de sensores, dispositivos médicos de diagnóstico e imagem e outros dispositivos inteligentes que são partes fundamentais da IoT.

Em 2019, o mundo enfrentou uma grande crise de saúde pública, causada pela COVID-19. Desde então, inúmeros danos estão sendo reparados. Antes da pandemia, abordagens de *eHealth* e telemedicina ainda estavam sendo testadas a partir de vários estudos com enfoque no uso de soluções digitais (ECCLESTON *et al.*, 2020). Nesse cenário de pandemia, Fagherazzi *et al.* (2020) aborda a maneira como essas soluções digitais permitiram consultas por vídeo, constatação de sintomas *on-line*, prescrições *on-line*, bem como uma triagem prévia do paciente para diminuir a sobrecarga dos hospitais e reduzir a proliferação do vírus.

Com o objetivo de amenizar a sobrecarga dos hospitais e prestar melhor atendimento aos pacientes, Wynants *et al.* (2020) destaca a importância de diagnósticos e prognósticos eficientes da doença. Nesse contexto, modelos de predição, que combinam várias variáveis para ponderar a probabilidade de pessoas estarem infectadas e os riscos associados à doença, são de grande relevância para que os profissionais da saúde possam alocar, de forma eficiente, os recursos de saúde que são limitados (WYNANTS *et al.*, 2020).

1.1. Problema da pesquisa

Com a extrema urgência por métodos que otimizassem o sistema de saúde e seus recursos frente à pandemia, vários estudos de modelos de predição da COVID-19 foram desenvolvidos. Porém, apesar de inúmeros modelos apresentarem uma discriminação moderada a ótima, muitos apresentam alto risco de viés devido ao *overfitting* do modelo, à calibração inadequada do modelo, ao uso de dados de fontes inapropriadas e aos relatórios imprecisos (WYNANTS *et al.*, 2020).

Dessa forma, a questão motivadora para a realização da pesquisa é: como a sintomatologia está associada à infecção de COVID-19?

1.2. Justificativa

A pandemia causada pela COVID-19 provocou danos para a sociedade que se estendem aos dias atuais e ainda estão sendo reparados, desde a morte de várias pessoas ao redor do mundo até os prejuízos econômicos causados pelo *lockdown*, que foi extremamente necessário para controle da propagação da doença. Dessa forma, modelos eficazes de predição foram de suma

importância para determinar a alocação otimizada de recursos da saúde e, conseqüentemente, reduzir parte desses danos.

Este trabalho se justifica cientificamente por tratar de um tema recente e preocupante, com elevado crescimento no número de publicações sobre predição de COVID-19 de 2020 para 2021, com potencial de crescimento ainda maior para 2022, como mostrado na Figura 2. Além disso, o número de publicações no Brasil ainda é pequeno comparado ao de países que mais publicaram, de acordo com a Figura 3, sendo importante a realização do trabalho com dados brasileiros para o crescimento científico do país na área deste estudo.

A Engenharia de Produção está evoluindo de forma alinhada a diversas tecnologias da Indústria 4.0. Assim, estudos como este contribuem para essa evolução a partir da utilização de métodos matemáticos e estatísticos abordados na Engenharia de Produção e integrados a tais tecnologias em ascensão. Além disso, este trabalho estuda modelos que cooperam diretamente para o aperfeiçoamento da gestão de pessoas e processos na esfera da saúde, otimizando diversas áreas do setor.

1.3. Objetivos

1.3.1. Objetivo geral

O presente estudo tem por objetivo apresentar uma proposta de triagem de pacientes a partir do padrão de sintomatologia associado à COVID-19.

1.3.2. Objetivos específicos

A fim de alcançar o objetivo geral, propõem-se os seguintes objetivos específicos:

- identificar os principais modelos de predição da COVID-19 por meio da revisão da literatura;
- propor um modelo conceitual com base nas principais contribuições da literatura;
- aplicar um modelo preditivo via regressão logística a partir da sintomatologia de pacientes que realizaram o exame de COVID-19;
- validar o modelo via matriz de confusão.

1.4. Estrutura dos capítulos

O trabalho está estruturado da seguinte forma: no Capítulo 2, tem-se a revisão bibliográfica sobre o assunto estudado, através da Teoria do Enfoque Meta-Analítico

Consolidado. No Capítulo 3, foi realizado o referencial teórico, contextualizando historicamente e mostrando os principais conceitos de Indústria 4.0 e Data Science, fornecendo base para o trabalho. No Capítulo 4, tem-se a metodologia utilizada no estudo. O Capítulo 5 apresenta os resultados e as análises do modelo de regressão logística desenvolvido e aplicado, além de possíveis aplicações práticas. Por fim, o Capítulo 6 apresenta as considerações finais, limitações do trabalho e sugestões de futuras linhas de pesquisa.

2. TEORIA DO ENFOQUE META-ANALÍTICO

Buscar estudos dentro de bases literárias e filtrá-los de forma relevante é fundamental para o melhor embasamento de uma pesquisa e, conseqüentemente, para a realização de um trabalho significativo dentro da literatura acadêmica.

A metodologia de revisão bibliográfica utilizada no presente trabalho foi a do modelo de Teoria do Enfoque Meta-Analítico (TEMAC). Segundo Mariano e Rocha (2017), o TEMAC é um método de revisão sistemática integrador, com o objetivo de destacar as interações entre diferentes estudos e seus respectivos autores. O modelo é composto por três etapas: preparação da pesquisa utilizando múltiplas bases de dados, apresentação e inter-relação dos dados e, por último, detalhamento, a concepção de um modelo integrador e validação por evidências. Essas etapas estão ilustradas na Figura 1, a seguir:

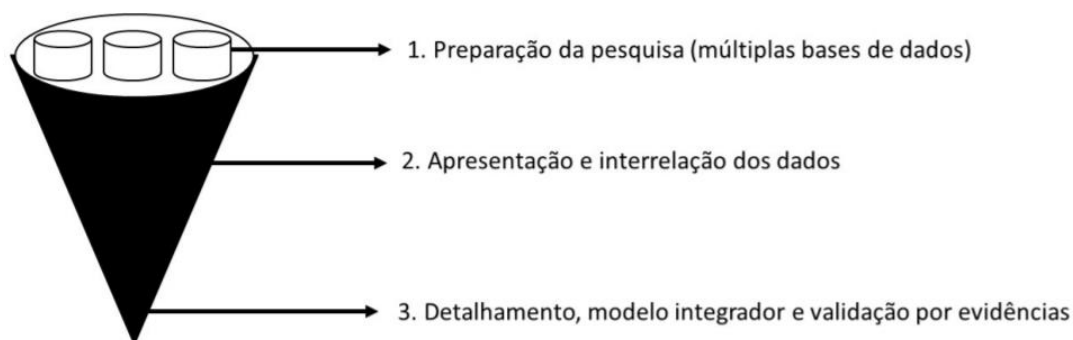


Figura 1: Modelo TEMAC
Fonte: Mariano e Rocha (2017)

2.1. Primeira etapa: preparação da pesquisa

Na primeira etapa do TEMAC, foi empregada a base de dados Web of Science, a qual, de acordo com a Clarivate, é o banco de dados de citações de editores independentes mais confiável do mundo.

A princípio, os seguintes termos foram buscados no campo de pesquisa da base Web of Science: “Big Data analytics”, “Data Analy*” ou “Data Science”, sem delimitação de tempo, buscando-se em todos os campos. A forma como o termo “Data Analy*” foi aplicado garante que sejam encontrados resultados que comecem com as palavras *data analy*, como: Data Analysis, Data Analytics etc. A busca por esses termos visou encontrar quantos estudos relacionados à Análise de Dados, Ciência de Dados e Análise de Big Data foram publicados na Web of Science até o momento para, posteriormente, pesquisar como a área de ciência de dados foi utilizada no contexto da pandemia. Nessa busca inicial, foram encontrados 369.714 resultados, o que mostra

a importância do tema e o quanto essa área vem sendo estudada.

Após a primeira pesquisa, foram incluídos, na busca, os termos “Sars-cov-2” ou “COVID-19”, e foram encontrados 3.851 resultados.

Com o objetivo de refinar a pesquisa ainda mais, foram incluídos os termos “control” ou “prediction” para encontrar estudos que mostram como a ciência de dados colaborou com o controle e a previsão da COVID-19. Nessa nova busca, foram encontrados 855 resultados.

2.2. Segunda etapa: Apresentação e inter-relação dos dados

Nessa etapa do TEMAC, serão apresentadas informações sobre os 855 resultados encontrados anteriormente e a inter-relação entre esses dados, focando em algumas das muitas variáveis retiradas dos resultados. Dentre elas, serão relacionadas: estudo mais antigo, evolução do tema ano a ano, países que mais publicaram, áreas que mais publicam, frequência de palavras-chave, agências que mais financiam a pesquisa, universidades que mais publicam, principais tipos de documento, principais idiomas dos documentos, revistas que mais publicaram sobre o tema, autores que mais publicaram vs. autores que mais foram mencionados e documentos mais citados.

Ao analisar os 855 resultados encontrados, constatou-se que o estudo mais antigo foi de 2020. Apesar de ser bem recente, isso já era esperado devido ao foco da pesquisa ser no contexto da pandemia causada pela COVID-19, que teve início em 2020 com a descoberta do vírus em 2019. O artigo, *Forecasting of Covid-19 Cases Using Machine Learning Approach*, de Kumar S. e Veer K., visa prever os casos de COVID-19 na Índia utilizando abordagens de *machine learning* e propôs um modelo de regressão, com base no conceito de Support Vector Machine, para prever número de mortes, número de casos de recuperação e total de casos confirmados para os próximos 30 dias. Os resultados do estudo mostraram que o modelo de regressão polinomial obteve pontuação de precisão acima de 95%, pontuação linear acima de 90% e pontuação RBF (Radial Basis Function) acima de 85% em previsão de mortes cumulativas, casos confirmados e casos de recuperação.

De acordo com o gráfico da Figura 2, que mostra o número de publicações no decorrer dos anos, é possível perceber a evolução do número de trabalhos publicados sobre o tema. É evidente que só houve publicações a partir de 2020 devido à COVID-19 ter surgido recentemente.



Figura 2: Publicações no decorrer dos anos
 Fonte: O autor. Dados: ISI Web of Science

O gráfico revela grande quantidade de publicações desde o surgimento da COVID-19, com 249 publicações em 2020. Além disso, percebe-se rápido aumento no número de publicações, com mais que o dobro de estudos publicados em 2021 em relação a 2020, sendo fácil relacionar esse aumento a todos os riscos e as consequências trazidas pelo coronavírus. Outro fator importante é a diminuição do número de publicações em 2022 em relação a 2021, porém com uma média mensal de publicações ainda maior do que em 2020, considerando que o presente estudo teve início em fevereiro de 2022, mostrando que o tema ainda possui considerável relevância.

A partir do mapa coroplético, representado na Figura 3, nota-se que os países que mais publicaram foram Estados Unidos, Índia e Austrália, com aproximadamente 27%, 9% e 7% de participação, respectivamente.



Figura 3: Publicação por país
 Fonte: O autor. Dados: ISI Web of Science

O Brasil teve uma participação significativa em publicações, totalizando 40 estudos, que representam quase 5% do total. Um dos principais estudos é o de Dayan *et al.* (2021) que propõe um modelo baseado em *federated learning* para prever a necessidade futura de oxigênio de pacientes sintomáticos com COVID-19, a partir de dados laboratoriais, sinais vitais e radiografias de tórax fornecidos por vinte instituições ao redor do mundo, incluindo uma do Brasil. Usando dados exclusivamente do Brasil, tem-se o trabalho de Ranzani *et al.* (2021) que estuda a eficácia da vacina CoronaVac na população acima de 70 anos do estado de São Paulo, e o estudo de Raymundo *et al.* (2021) que investiga a incidência da COVID-19 nos municípios brasileiros e analisa a sua relação com as variáveis sociodemográficas, com o objetivo de entender o contexto social e a propagação da doença no país.

Na Figura 4, é possível visualizar as 15 áreas que mais publicaram sobre o tema. Nota-se que as principais áreas são: saúde pública, ambiental e ocupacional e ciência da computação, representando 14,3% e 13,5% das publicações, respectivamente.

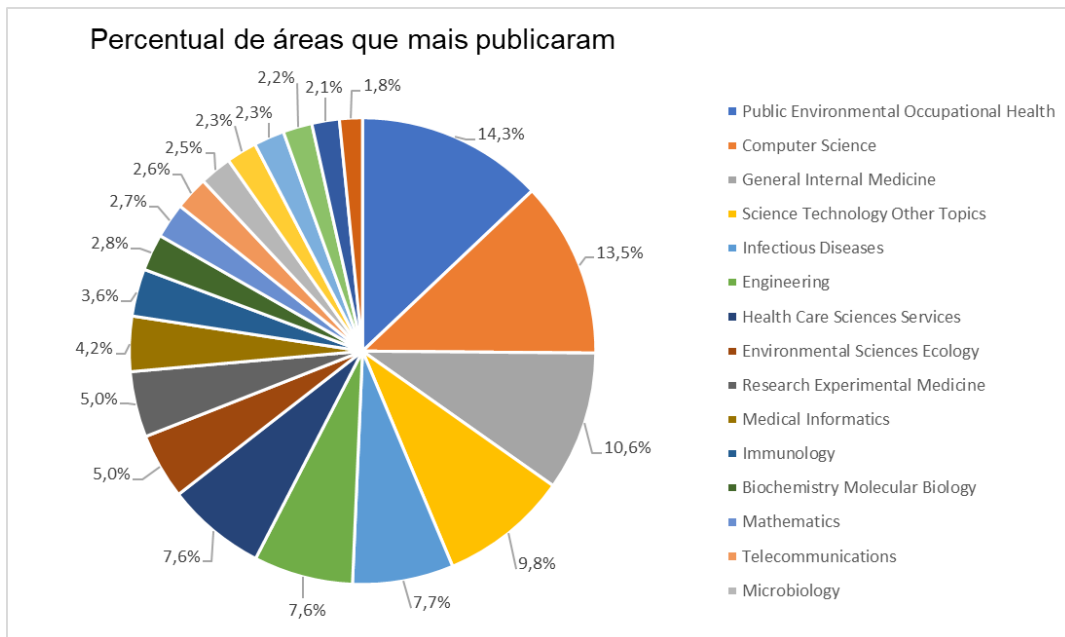


Figura 4: Áreas que mais publicaram
 Fonte: O autor. Dados: ISI Web of Science

Outras áreas também publicaram sobre o tema, com participações menores, como: medicina interna, ciência e tecnologia, doenças infecciosas, engenharia e diversas outras. Essa extensa gama de áreas já era esperada pelo autor, visto que o tema do trabalho é bastante amplo. Porém, muitas delas pertencem ao âmbito da saúde e as outras estão relacionadas com ciência de dados e tecnologia, proporcionando mais direção para o desenvolvimento do estudo.

A partir das palavras-chave mais utilizadas nas publicações no decorrer dos anos, é possível analisar quais palavras, estudos e até áreas estão relacionadas com o tema pesquisado. De acordo com a Figura 5, é possível verificar quais foram as palavras-chave mais frequentes nos resultados encontrados. Na imagem, os círculos representam as palavras-chave e as linhas representam o vínculo entre elas. Para representar a progressão dos anos, em ordem crescente, foi utilizada uma escala de cores.

da União Europeia, além de financiar pesquisas por toda a Europa. Além disso, as universidades que mais tiveram publicações foram: University of London, Harvard University, University of Oxford e University of California System, respectivamente.

Ao analisar quais foram os tipos de documentos mais publicados, percebeu-se que os mais usados foram de artigos e artigos de revisão, respectivamente. A imensa maioria dos trabalhos encontram-se em língua inglesa, porém algumas publicações estão, em ordem decrescente, em espanhol, alemão, seguidos de uma publicação em português e russo, cada. Dos 855 resultados, a revista *International Journal of Environmental Research and Public Health*, da Suíça, foi a que teve mais publicações, 18 no total. A revista possui fator de impacto de 5 anos de primeiro quartil (Q1) em saúde pública, ambiental e ocupacional.

Acerca dos autores que mais publicaram sobre o assunto, encontram-se Jie Xu, com 12 publicações, Yadong Wang e Haiyan Yang, com 11 trabalhos publicados, os quais foram desenvolvidos em conjunto pelos três autores. Além disso, os três estão entre os autores mais citados, comprovando a relevância deles para o tema. Jie Xu teve 39 citações, Yadong Wang, 27 e Haiyan Yang, 27.

Outros autores com mais publicações foram: Li Shi, com 7 publicações, e Hongjie Hou e Ying Wang, ambos com seis publicações, como pode ser observado na Figura 6.

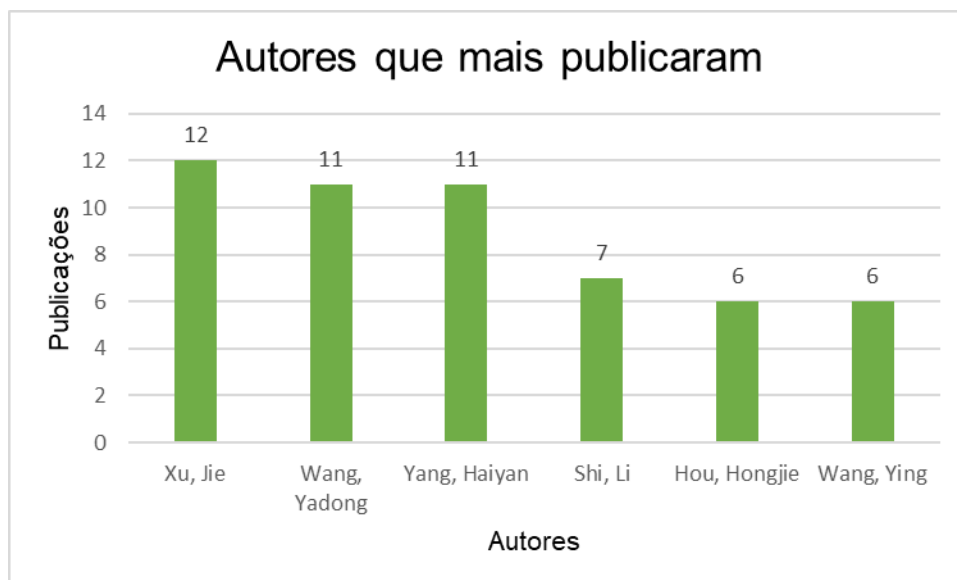


Figura 6: Autores que mais publicaram
Fonte: O autor. Dados: ISI Web of Science

O principal aspecto entre os estudos de maior relevância dos três autores que mais publicaram foi a utilização de métodos de meta-análise e técnicas estatísticas para identificar relações entre pacientes infectados com COVID-19 e outras doenças respiratórias, como asma,

embolia pulmonar e tuberculose. Pela análise, foram verificadas as possíveis influências da COVID-19 sobre essas doenças e o risco de morte desses pacientes. Em um dos estudos, foi validado que a tuberculose está fortemente associada a um risco aumentado de gravidade e mortalidade entre pacientes com COVID-19.

Contudo, a publicação de Jie Xu com mais citações, 12 no total, foi o único trabalho dele da busca que não foi concebido em conjunto com os outros dois autores, Yadong Wang e Haiyan Yang. A publicação em questão, *Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach*, visa analisar o uso de *federated learning*, técnica de *machine learning*, para prever a mortalidade em pacientes hospitalizados com COVID-19 dentro de sete dias. Ao final, concluiu-se que a técnica se revela promissora para o desenvolvimento de modelos preditivos robustos, sem comprometer a privacidade do paciente.

Os seis autores que mais publicaram também estão entre os que mais foram citados, como demonstrado na Figura 7:

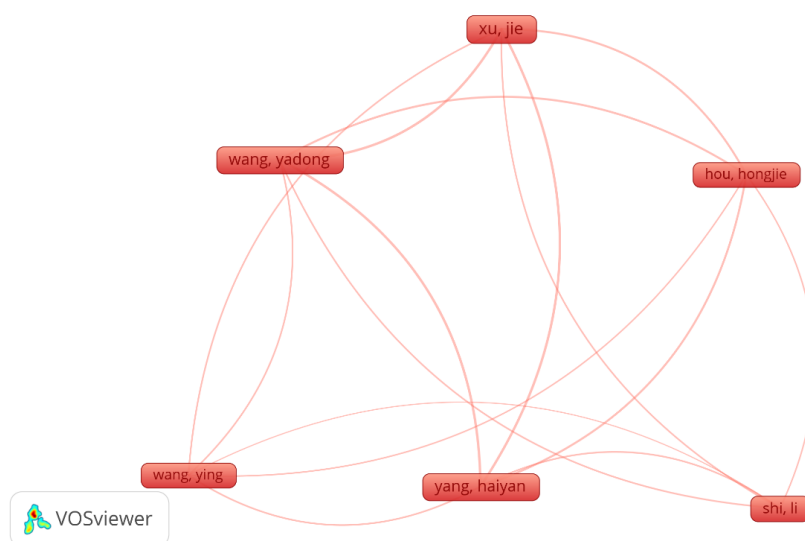


Figura 7: Autores mais citados
Fonte: VosViewer. Dados: ISI Web of Science

Analisando a figura, nota-se a forte correlação entre os autores – todos eles têm pelo menos uma publicação desenvolvida em conjunto com cada um dos outros autores. O trabalho mais citado de Ying Wang, com 152 citações, não teve a participação dos outros autores mencionados. Nesse artigo, *Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China*, os autores propõem um modelo determinístico de dinâmica de doenças infecciosas para estimar a tendência epidêmica em Wuhan, na China, no período de

fevereiro de 2020.

Apesar de ter apenas cinco publicações, o autor mais referido foi Wei Wang, com 512 citações. Em seu trabalho mais expressivo, *Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China*, foram analisados os dados de *contact survey*, de Wuhan e Shanghai, e *contact-tracing*, da província de Hunan. A partir dessa análise, foi criado um modelo de transmissão para estudar o impacto do distanciamento social e do fechamento de escolas. Foi descoberto que somente o distanciamento social seria suficiente para controlar a COVID-19. Além disso, apesar de o fechamento de escolas não ser suficiente para interromper a transmissão por conta própria, ele pode reduzir o ápice de incidência de 40% a 60% e desacelerar o avanço da epidemia.

Analisando as publicações mais citadas, o artigo *Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal* ocupa o primeiro lugar, com 969 citações. Esse trabalho tem como principal objetivo revisar e avaliar a validade e utilidade de modelos de previsão para diagnosticar a COVID-19 em pacientes com suspeita de infecção, para prognóstico de pacientes com COVID-19 e para detecção de pessoas que estão infectadas pelo vírus, com risco aumentado ou internadas no hospital devido à doença. A revisão aponta que quase todos os modelos de previsão publicados são mal explanados e com alto risco de viés. Apesar disso, foram encontrados dois modelos promissores, um de diagnóstico e outro de prognóstico.

Além de trabalhos com abordagem orientada para a área da saúde que tiveram muitas citações, outro importante estudo foi o artigo *Effect of non-pharmaceutical interventions to contain COVID-19 in China*, que teve 409 citações. Os autores dessa publicação desenvolveram uma estrutura de modelagem para simular diferentes cenários de surto e intervenção em toda a China, para validar o uso de intervenções não farmacêuticas para conter a COVID-19.

O mapa de calor da Figura 8 apresenta as palavras-chave contidas nos títulos e resumos de todas as publicações acerca do tema. O mapa é determinante para destacar as principais palavras que estão sendo utilizadas e, com isso, identificar os principais tópicos dentro do tema principal.

o tema. O mapa de calor do *co-citation* pode ser visualizado na Figura 9:

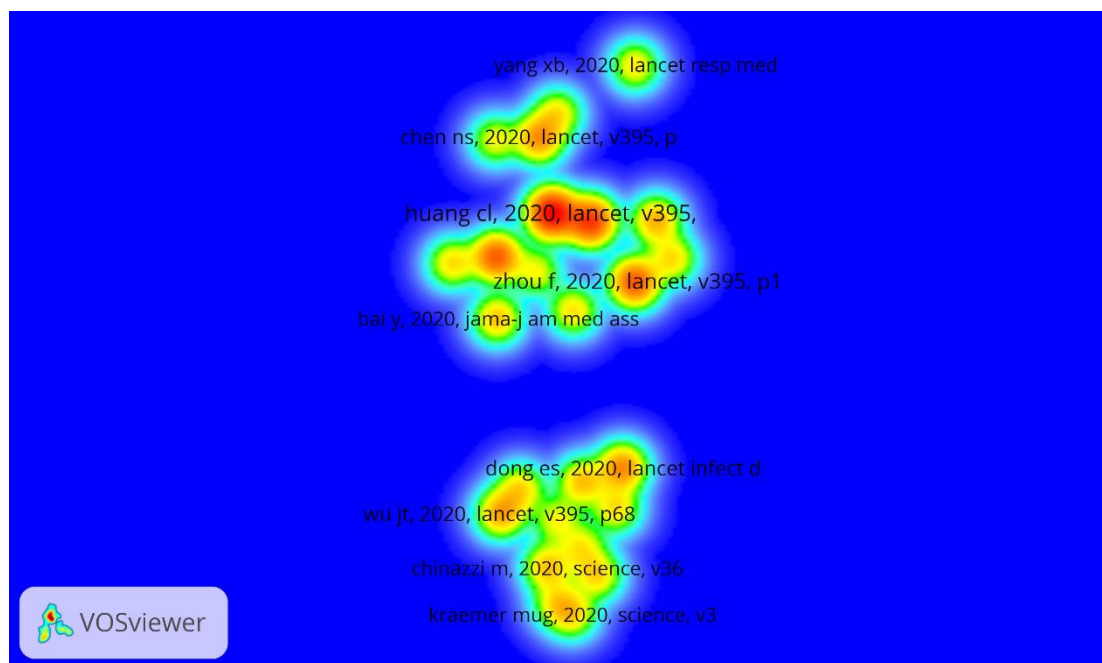


Figura 9: Mapa de calor de *co-citation*
Fonte: VosViewer. Dados: ISI Web of Science

A primeira abordagem analisada é a dos autores da parte de cima do mapa, e os estudos mais significativos foram os de Chen, Huang e Zhou. A linha de tendência desse grupo de autores foi a investigação dos principais aspectos epidemiológicos e clínicos da COVID-19, que, na época, ainda era muito recente. Foram realizadas análises de dados para mensurar as principais características e sintomas dos pacientes infectados de Wuhan, como: idade, sexo, pacientes com doenças crônicas, pacientes com manifestação febril, sintomas de tosse, falta de ar etc. Em um dos estudos de Zhou, houve uma análise mais completa, utilizando métodos de logística univariável e multivariável para identificar os principais fatores de risco associados ao óbito hospitalar de pacientes infectados pelo Coronavírus.

Já a segunda vertente analisada é referente aos autores Dong, Wu, Chinazzi e Kraemer, que estão localizados na parte de baixo do mapa. A abordagem desses autores é focada na elaboração de um panorama geral de evolução da, até então, epidemia. No trabalho de Dong, foi desenvolvido um *dashboard* interativo *on-line* com o propósito de visualizar e rastrear casos relatados de COVID-19 em tempo real, para todos os países afetados. Outros estudos utilizaram modelos de metapopulação de transmissão de doença para simular a evolução epidêmica da COVID-19 e analisar os efeitos das restrições de mobilidade como medida de controle contra sua disseminação.

A análise *coupling* traz o ponto de vista dos principais *fronts* de pesquisa a partir da observação dos *clusters* de autores que são formados, em que estão presentes os artigos que citam

os mesmos trabalhos. Seguindo essa premissa, o mapa de calor de *coupling* foi concebido e apresentado na Figura 10:

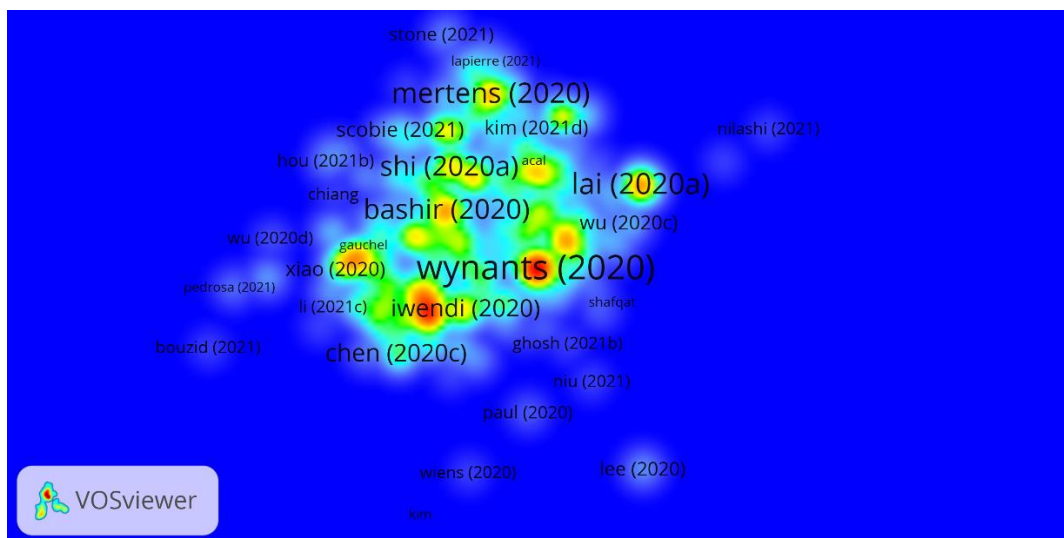


Figura 10: Mapa de calor do *coupling*
 Fonte: VosViewer. Dados: ISI Web of Science

Analisando o mapa, nota-se a formação de vários *clusters* de autores. A princípio, dois núcleos mais acentuados se destacam – dos autores Wynants e Iwendi.

A pesquisa de Wynants *et al.* (2020) refere-se a *review Prediction models for diagnosis and prognosis of Covid-19 infection: systematic review and critical appraisal*, que já foi abordada anteriormente por ser o trabalho mais citado da busca realizada e pela relevância do estudo para o tema. Em suma, o trabalho busca validar modelos de previsão para diagnosticar a doença de Coronavírus 2019. Já o trabalho de Iwendi *et al.* (2020) propõe um modelo *random forest* impulsionado pelo algoritmo de *machine learning* AdaBoost. O modelo utiliza dados geográficos, demográficos, de viagem e de saúde para predizer a gravidade do caso de infecção do paciente e o possível resultado final, de recuperação ou morte.

Outro núcleo importante é referente ao estudo de Mertens *et al.* (2020), que busca identificar preditores de medo do Coronavírus, logo após o surto ser declarado uma pandemia pela Organização Mundial da Saúde (OMS). Tal ponto de vista evidencia a preocupação com a saúde mental mundial em um cenário de pandemia.

Após realizar o detalhamento do tema, foram listadas as pesquisas mais importantes para o presente trabalho, como pode ser observado no Quadro 1.

Quadro 1: Lista de artigos mais relevantes

Título e ano	Autores	Abordagem
<i>Federated learning of electronic health records to</i>	Vaid, Akhil; Jaladanki, Suraj K.; Xu, Jie; Teng, Shelly;	O estudo analisa o uso de <i>federated learning</i> , técnica de

Título e ano	Autores	Abordagem
<i>improve mortality prediction in hospitalized patients with COVID-19: machine learning approach</i> (2021)	Kumar, Arvind; Lee, Samuel; Somani, Sulaiman; Paranjpe, Ishan; De Freitas, Jessica K.; Wanyan, Tingyi; Johnson, Kipp W.; Bicak, Mesude; Klang, Eyal; Kwon, Young Joon; Costa, Anthony; Zhao, Shan; Miotto, Riccardo; Charney, Alexander W.; Boettinger, Erwin; Fayad, Zahi A.; Nadkarni, Girish N.; Wang, Fei; Glicksberg, Benjamin S.	<i>machine learning</i> para prever a mortalidade em pacientes hospitalizados com COVID-19 dentro de 7 dias. Ao final, o artigo concluiu que a técnica se mostra promissora no desenvolvimento de modelos preditivos robustos, sem comprometer a privacidade do paciente.
<i>Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China</i> (2020)	Wang, Huwen; Wang, Zezhou; Dong, Yinqiao; Chang, Ruijie; Xu, Chen; Yu, Xiaoyue; Zhang, Shuxian; Tsamlag, Lhakpa; Shang, Meili; Huang, Jinyan; Wang, Ying; Xu, Gang; Shen, Tian; Zhang, Xinxin; Cai, Yong	O artigo utiliza um modelo determinístico de dinâmica de doenças infecciosas com o objetivo de modelar e prever o número de casos de COVID-19 em Wuhan, China, em fevereiro de 2020.
<i>Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China</i> (2020)	Zhang, Juanjuan; Litvinova, Maria; Liang, Yuxia; Wang, Yan; Wang, Wei; Zhao, Shanlu; Wu, Qianhui; Merler, Stefano; Viboud, Cecile; Vespignani, Alessandro; Ajelli, Marco; Yu, Hongjie	Esse estudo analisou dados de <i>contact survey</i> , de Wuhan e Shanghai, e <i>contact-tracing</i> , da província de Hunan. A partir dessa análise, foi criado um modelo de transmissão para estudar o impacto do distanciamento social e do fechamento de escolas no controle da COVID-19.
<i>Prediction models for diagnosis and prognosis of Covid-19 infection: systematic review and critical appraisal</i> (2020)	Wynants, Laure; Van Calster, Ben; Bonten, Marc M. J.; Collins, Gary S.; Debray, Thomas P. A.; De Vos, Maarten; Haller, Maria C.; Heinze, Georg; Moons, Karel G. M.; Riley, Richard D.; Schuit, Ewoud; Smits, Luc J. M.; Snell, Kym I. E.; Steyerberg, Ewout W.; Wallisch, Christine; Van Smeden, Maarten	O estudo visa revisar e validar modelos de previsão para diagnosticar a COVID-19 em pacientes com suspeita de infecção, para prognóstico de pacientes com COVID-19 e para detecção de pessoas que estão infectadas pelo vírus com risco aumentado ou internadas no hospital devido à doença.

Título e ano	Autores	Abordagem
<i>Effect of non-pharmaceutical interventions to contain COVID-19 in China</i> (2020)	Lai, Shengjie; Ruktanonchai, Nick W.; Zhou, Liangcai; Prosper, Olivia; Luo, Wei; Floyd, Jessica R.; Wesolowski, Amy; Santillana, Mauricio; Zhang, Chi; Du, Xiangjun; Yu, Hongjie; Tatem, Andrew J.	O trabalho desenvolveu uma estrutura de modelagem para simular diferentes cenários de surto e intervenção em toda a China, para validar o uso de intervenções não farmacêuticas no controle do surto da COVID-19.
<i>Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China</i> (2020)	Huang, Chaolin; Wang, Yeming; Li, Xingwang; Ren, Lili; Zhao, Jianping; Hu, Yi; Zhang, Li; Fan, Guohui; Xu, Jiuyang; Gu, Xiaoying; Cheng, Zhenshun; Yu, Ting; Xia, Jiaan; Wei, Yuan; Wu, Wenjuan; Xie, Xuelei; Yin, Wen; Li, Hui; Liu, Min; Xiao, Yan; Gao, Hong; Guo, Li; Xie, Jungang; Wang, Guangfa; Jiang, Rongmeng; Gao, Zhancheng; Jin, Qi; Wang, Jianwei; Cao, Bin	Estudo voltado para análises de dados para mensurar principais características e sintomas dos pacientes infectados pelo Coronavírus, de Wuhan, como: idade, sexo, pacientes com doenças crônicas, pacientes com manifestação febril, sintomas de tosse, falta de ar etc.
<i>An interactive web-based dashboard to track COVID-19 in real time</i> (2020)	Dong, Ensheng; Du, Hongru; Gardner, Lauren	Os autores desse trabalho desenvolveram um <i>dashboard</i> interativo <i>on-line</i> com o propósito de visualizar e rastrear casos relatados de COVID-19 em tempo real, para todos os países afetados.

Fonte: O autor

Com o intuito de estudar diferentes modelos de predição das principais características e sintomas de pessoas infectadas com COVID-19, foram elencados os principais modelos para aplicá-los no presente estudo e, assim, analisar como eles se comportam em diferentes cenários, como pode ser verificado no Quadro 2, o qual contém os títulos dos trabalhos, os modelos utilizados e as descobertas resultantes das aplicações desses modelos.

Quadro 2: Principais modelos de predição encontrados

Título	Modelo	Descobertas
<i>Individualizing risk</i>	Modelo logístico	Menor risco de positivo para

Título	Modelo	Descobertas
<i>prediction for positive Coronavirus Disease 2019 testing: results from 11,672 patients</i>	multivariável para prever resultado de testes (<i>swab</i>) de COVID-19, com base em dados demográficos, comorbidades, histórico de imunização, sintomas, histórico de viagens, variáveis de laboratório, medicações identificadas antes do teste, variáveis contínuas comparadas pelo teste Mann-Whitney U. e variáveis categóricas comparadas pelo teste <i>chi-square</i> (X^2).	COVID-19 em indivíduos asiáticos do que em indivíduos brancos. Menor risco observado para vacina pneumocócica polissacarídica e vacina contra gripe. Maior risco para indivíduos com baixo nível socioeconômico. Risco reduzido para indivíduos sob efeito das drogas Melatonina, Carvedilol e Paroxetina.
<i>Risk stratification of patients admitted to hospital with Covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score</i>	4C Mortality Score – modelo de pontuação de estratificação de risco para prever a mortalidade em pacientes internados com COVID-19. Inclui oito variáveis: idade, sexo, número de comorbidades, frequência respiratória, saturação de oxigênio, nível de consciência, nível de ureia e proteína C-reativa.	Discriminação do modelo foi igual para homens e mulheres. Discriminação do modelo foi melhor para todos os grupos étnicos não-brancos em comparação ao grupo étnico branco.
<i>Clinical features of patients infected with 2019 novel Coronavirus in Wuhan, China</i>	<i>Boxplots</i> (diagramas de caixa) foram desenhados para descrever as concentrações plasmáticas de citocinas e quimiocinas. As variáveis contínuas foram expressas como mediana (IQR) e comparadas com o teste U de Mann-Whitney. As variáveis categóricas foram expressas em número (%) e comparadas pelo teste do X^2 entre os grupos cuidados na UTI e não cuidados na UTI.	73% dos pacientes eram homens. 32% tinham doenças subjacentes, incluindo diabetes, hipertensão e doença cardiovascular. Idade média: 49 anos. Sintomas comuns no início da doença: febre 98%, tosse 76%, mialgia ou fadiga 44%. Dispneia ou falta de ar desenvolvida em 55% dos pacientes (tempo médio do início da doença até a dispneia foi de 8 dias). 63% dos pacientes tiveram linfopenia.

Título	Modelo	Descobertas
		<p>100% dos pacientes tiveram pneumonia com achados anormais no tórax.</p> <p>32% dos pacientes foram internados na UTI.</p> <p>15% dos pacientes morreram.</p>
<p><i>Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study</i></p>	<p>Regressão multivariável. Cinco variáveis consideradas para a análise multivariável: contagem de linfócitos, teste d-dímero, pontuação SOFA (Sequential Organ Failure Assessment), doença cardíaca coronária e idade.</p>	<p>32 pacientes dos 191 precisaram ser submetidos à ventilação mecânica invasiva, dos quais 31 (97%) morreram. Na análise univariada, a chance de óbito intra-hospitalar foi maior em pacientes com diabetes ou doença coronariana.</p> <p>No modelo de regressão logística multivariável, foi descoberto que idade avançada, pontuação Sofa mais alta e d-dímero maior que 1 µg/mL na admissão foram associados a maiores chances de morte.</p>

Fonte: O autor

Para finalizar o TEMAC, propõe-se um modelo integrador de todas as informações obtidas até o momento, como ilustrado na Figura 11.

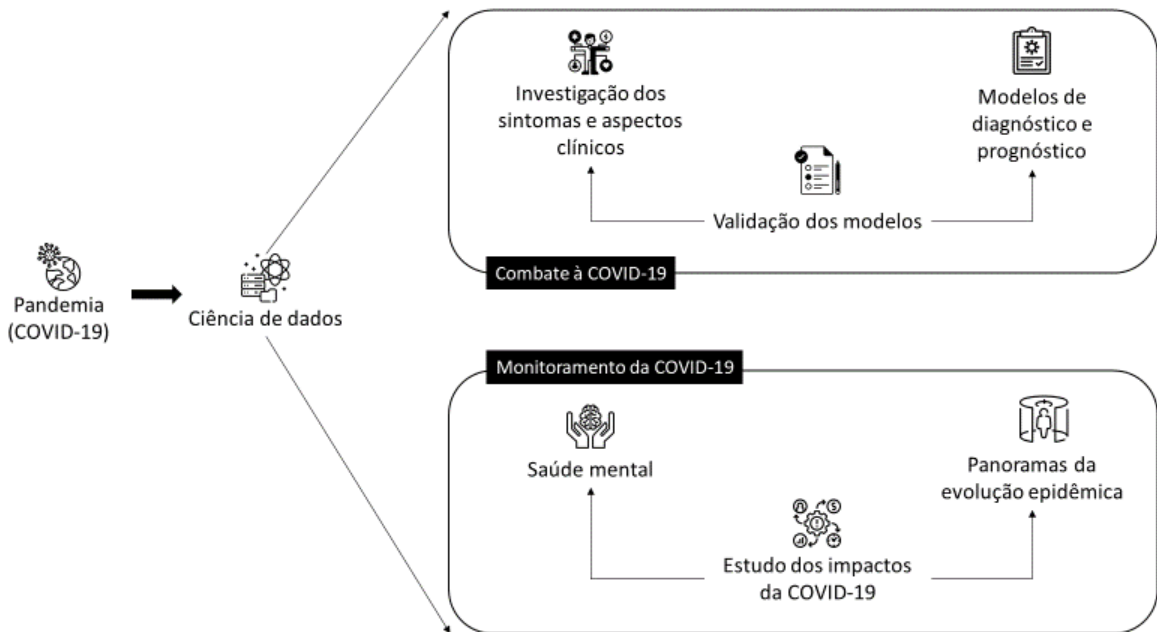


Figura 11: Modelo integrador
Fonte: O autor

O modelo integrador propõe a utilização de duas abordagens principais que associam o uso da ciência de dados para combate à COVID-19, bem como o monitoramento do vírus no cenário mundial. Para o combate à COVID-19, foram desenvolvidos modelos que identificam os principais sintomas e aspectos clínicos causados pelo vírus, bem como modelos que fazem o diagnóstico e prognóstico da doença; além disso, foram encontrados estudos com o objetivo de validar esses modelos para um controle mais eficaz da disseminação do vírus. Já a abordagem de monitoramento da COVID-19 tem como principal foco o estudo dos impactos causados pela pandemia, buscando entender os principais sentimentos (ansiedade, medo, angústia etc.) da população mundial naquele momento e os danos que isso poderia causar para a saúde mental das pessoas. Também várias pesquisas buscaram trazer um panorama geral da evolução epidêmica do surto, identificando dados quantitativos de casos de infecção, recuperação e mortes.

3. REFERENCIAL TEÓRICO

3.1. Indústria 4.0

Indústria 4.0 é um termo que surgiu na Alemanha e representa a quarta Revolução Industrial. O Ministério Federal de Educação e Pesquisa do governo alemão transformou o termo “Indústria 4.0” como epônimo para utilizá-lo em um futuro projeto no contexto da estratégia de alta tecnologia do país (LASI *et al.*, 2014).

De acordo com Schwab (2016), o processo de transição da terceira Revolução Industrial para a quarta iniciou-se com a virada do século e baseia-se em uma chamada revolução digital. Segundo o autor, a quarta Revolução Industrial caracteriza-se pela internet muito mais móvel e “onipresente”, por sensores mais poderosos e menores custando ainda menos, pela inteligência artificial e pelo aprendizado de máquinas. Em seu livro, Schwab (2016) destaca três megatendências impulsionadas pela quarta Revolução Industrial – física, digital e biológica – causando grande impacto na economia, nos negócios, no governo, na sociedade e no indivíduo.

Segundo Schwab (2016), a quarta Revolução Industrial cria um cenário em que sistemas de fabricação físicos e virtuais colaboram entre si de forma global e flexível, por meio da habilitação de fábricas inteligentes. Nesse contexto, a fabricação inteligente integra ativos de fabricação de hoje com os de amanhã, a partir de sensores, tecnologias de comunicação, plataformas de computação, controle, modelagem intensiva de dados, simulação e engenharia preditiva (KUSIAK, 2017). Além disso, Kusiak (2017) enfatiza que a fabricação inteligente aplica conceitos de sistemas ciberfísicos (CPS), Internet das Coisas (IoT), computação em nuvem, computação orientada a serviços, inteligência artificial e ciência de dados.

Na Indústria 4.0, tecnologias relacionadas ao Big Data como inteligência artificial, aprendizado de máquina e tecnologia robótica estão exercendo papel fundamental (SONG; ZHU, 2017). Além disso, os autores apontam que especialistas da área acreditam que essas tecnologias terão impactos significativos nas nossas vidas e nos ajudarão a enfrentar novos desafios que surgem a cada dia.

3.1.1. Os nove pilares da Indústria 4.0

Para transmitir uma visão geral da Indústria 4.0, Rüßmann *et al.* (2015) define seus nove pilares: *additive manufacturing* (manufatura aditiva), *augmented reality* (realidade aumentada), *autonomous robots* (robôs autônomos), *Big Data and analytics, the cloud* (computação em nuvem), *cybersecurity* (cibersegurança), *horizontal and vertical system integration* (integração

horizontal e vertical de sistemas), *Internet of Things* (Internet das Coisas) e *simulation* (simulação), como pode-se conferir na Figura 12:

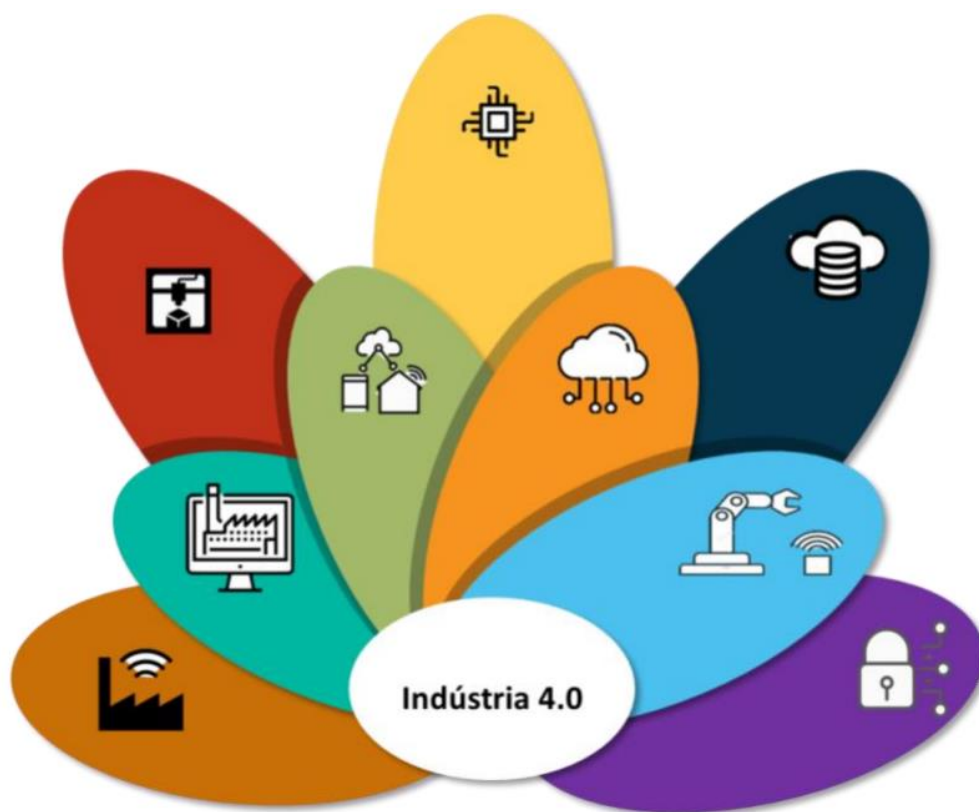


Figura 12: Nove pilares da Indústria 4.0
Fonte: Mariano *et al.* (2019)

A manufatura aditiva, também conhecida como impressão 3D, é definida por produtos customizados de acordo com os requisitos dos clientes (ERBOZ, 2017). Nessa perspectiva, Rüßmann *et al.* (2015) sugere que esses métodos de manufatura aditiva serão empregados para produzir pequenos lotes desses produtos customizados que fornecem certas vantagens, tais como complexidade e *design* leves. Dessa forma, esse tipo de manufatura tem grande potencial para melhorar a logística de transporte e sistemas de estoque.

De acordo com Erboz (2017), a realidade aumentada é definida como a tecnologia que possibilita a interação entre o mundo virtual e seus usuários, em que o mundo virtual está sendo projetado em partes do mundo real. É uma tecnologia promissora, mas que ainda é considerada muito nova. Todavia, no futuro, as companhias farão uso muito mais abrangente da realidade aumentada para proporcionar informações em tempo real aos trabalhadores e, com isso, aprimorar a tomada de decisões e os métodos de trabalho (RÜßMANN *et al.*, 2015).

Robôs estão se tornando cada vez mais autônomos, cooperativos e flexíveis. Em certo ponto, eles começarão a se comunicar com outros robôs, além de trabalhar de forma segura com

humanos e aprender com eles (RÜßMANN *et al.*, 2015). Nesse contexto, Aiman *et al.* (2016) aponta que o uso de robôs está se expandindo para englobar várias aplicações: fabricação, logística, gerenciamento de escritório etc. Além disso, os robôs poderão ser controlados remotamente, possibilitando que as fábricas operem 24 horas por dia, enquanto humanos possuem carga horária limitada.

O conceito de Big Data aplica-se a conjuntos de dados grandes, diversos e complexos que influenciam na tomada de decisão organizacional de uma empresa em relação ao seu estratégico (ERBOZ, 2017). Além disso, é caracterizado pelos 5Vs: volume – tamanho dos dados; velocidade – velocidade de criação, processamento, análise e armazenamento dos dados; variedade – tipos de dados, fontes e modos; veracidade – qualidade, confiabilidade e incerteza dos dados; valor – descoberta de valor acionável, alto retorno de investimento, aumento de destaque para clientes ou produtos e inovação em negócios (SONG; ZHU, 2015).

Segundo Ahmed *et al.* (2017), a integração de dados e *analytics* possibilitam que as empresas inovem em seus processos de negócios. De forma específica, essas organizações podem utilizar ferramentas de análise de dados para transformar grande quantidade de dados coletados em *insights* valiosos. Por isso, Frank, Dalenogare e Ayala (2017) concluem que Big Data e *analytics* são facilitadores essenciais para aplicações avançadas da Indústria 4.0, visto que a inteligência do sistema depende da grande quantidade de dados acumulados (Big Data) e da capacidade de análise com métodos avançados (*analytics*).

Ambrust *et al.* (2009) determina o conceito de *cloud computing*, ou computação em nuvem, como duas partes interligadas: uma refere-se às aplicações enviadas como serviços pela internet e a outra, ao *hardware* e aos sistemas do *software*, que estão nos *data centers* e fornecem esses serviços.

Segundo Erboz (2017), há três modelos de computação em nuvem: *Software* como serviço (SaaS), no qual o cliente precisa pagar para ter acesso ao serviço, como Google *drive*, Gmail e Dropbox; Plataforma como Serviço (PaaS), em que os clientes têm acesso às suas aplicações na nuvem, como desenvolvedores de *softwares* e aplicativos; Infraestrutura como Serviço (IaaS), que oferece atividades básicas, como capacidade de *hardware* em memória, armazenamento e processamento.

De acordo com Erboz (2017), a cibersegurança ainda é um ponto com várias adversidades devido aos impactos que ataques terroristas podem gerar para o ambiente de negócios. Isso faz com que organizações procurem métodos cada vez mais sofisticados para se protegerem desses ataques mal-intencionados.

“Cibersegurança é o conjunto de tecnologias e processos projetados para proteger

computadores, redes, programas e dados contra ataques, acesso não autorizado, alteração ou destruição.” (BUCZAK; GUVEN, 2016, p. 1). Ainda de acordo com Buczak e Guven (2016), sistemas de cibersegurança são formados por sistemas de segurança de rede e sistemas de segurança de computador (*host*). Para proteger esses ativos, são utilizadas ferramentas como *firewall*, antivírus e sistemas de detecção de intrusão (IDS).

Para Rüßmann *et al.* (2015), a Indústria 4.0 permitirá que organizações, departamentos, funções e capacidades se interliguem cada vez mais, uma vez que as redes de integração de dados entre as empresas estão cada vez mais avançadas, tornando cadeias de valor mais automatizadas. Nesse mesmo contexto, surgem as fábricas inteligentes, que são definidas como a integração de objetos físicos com sistemas de informação, através de um sistema ciberfísico de integração (CPS) (WANG *et al.*, 2016).

De acordo com Al-Fuqaha *et al.* (2015), a Internet das Coisas é caracterizada pela conexão entre objetos físicos e a internet, que se tornou possível a partir da evolução de Radio-Frequency IDentification (RFID), sensores inteligentes, tecnologias de comunicação e protocolos de internet. A ideia principal é que esses desenvolvimentos colaborem entre si, sem interferência dos seres humanos, proporcionando novos tipos de aplicações. No contexto de campos de trabalho e doméstico, a IoT tornará possíveis aplicações como domótica, *assisted living*, e-saúde e aprendizado aprimorado (ATZORI; IERA; MORABITO, 2010).

No contexto de Indústria 4.0, Rüßmann *et al.* (2015) estabelece que simulações 3D de produtos, materiais e processos de produção possibilitarão que operadores otimizem o tempo de configuração das máquinas e aumentem a qualidade da produção. Isso se torna possível por meio de testes e aprimoramento das configurações dessas máquinas em um modelo virtual que utiliza dados em tempo real para espelhar o mundo físico. Como sugere Erboz (2017), essas ferramentas de simulação propiciam um ambiente de produção sustentável, pois elas dão suporte às atividades associadas à produção. Além disso, a gestão do chão de fábrica fica mais eficaz com a habilidade de autoconfiguração dessas ferramentas.

3.1.2. Smart buildings

O contexto de Indústria 4.0 cria, ainda, possibilidades para conceitos como cidades inteligentes, fábricas inteligentes, hospitais inteligentes e vários outros “serviços inteligentes”. Segundo Zanella *et al.* (2014), as IoTs urbanas são idealizadas para dar suporte à perspectiva de cidades inteligentes, que, por sua vez, têm como propósito oferecer serviços de valor agregado para a gestão da cidade e para os cidadãos, por meio da utilização de tecnologias de última geração. Além de oferecer melhorias na qualidade de vida da população, os sistemas de cidades inteligentes

também promovem aumentos significativos na gestão de ativos, tais como: sistemas de transporte inteligentes, gestão de iluminação pública, gestão de semáforos, monitoramento ambiental, vigilância, dentre outros (MINOLI; SOHRABY; OCCHIOGROSSO, 2017).

No viés da saúde, o conceito de hospitais inteligentes também oferece inúmeros benefícios para a sociedade. Devido à natureza ubíqua da Internet das Coisas, todos os elementos do sistema de saúde (indivíduos, medicamentos, aparelhos etc.) podem ser monitorados e administrados continuamente (RAHMANI, 2018). Nesse contexto, a identificação e o rastreamento automático de pessoas e dispositivos biomédicos em hospitais, o monitoramento em tempo real dos parâmetros fisiológicos dos pacientes para detecção antecipada de deterioração clínica e a combinação correta entre pacientes e medicamentos são apenas algumas das várias possibilidades que sistemas inteligentes possuem para dar suporte e aprimorar tanto a assistência médica quanto os processos biomédicos (CATARINUCCI *et al.*, 2015).

3.2. Data Science

Paralelamente ao surgimento dessas novas tecnologias que caracterizam a Indústria 4.0, a produção de dados passou a aumentar exponencialmente, o que causou o surgimento da era do Big Data. Os desafios ocasionados por ele começam com a explosão de dados, que aumenta de forma gigantesca, e chegam às necessidades de novas tecnologias com enormes capacidades de processamento de dados.

Segundo Laney (2001), os desafios relacionados ao gerenciamento de dados são definidos em três dimensões, conhecidas como os 3 V's do Big Data: volume, variedade e velocidade. Em uma abordagem semelhante, Gartner (2012) o define do seguinte modo: “*Big data* são ativos de informações de alto volume, alta velocidade e/ou alta variedade que exigem formas econômicas e inovadoras de processamento de informações que permitem insights aprimorados, tomada de decisões e automação de processos.”

Com o passar do tempo, outras duas dimensões são bastante mencionadas: veracidade e valor. Song e Zhu (2015) definem os 5 V's como: (1) volume como o tamanho dos dados, que pode escalar para *terabytes*, *petabytes* ou ainda mais; (2) velocidade como a velocidade de criação, processamento, análise e armazenamento desses dados; (3) variedade como os diferentes tipos de dados, fontes e formas de lidar; (4) veracidade como qualidade, confiabilidade e incerteza nos dados; e (5) valor sendo a descoberta de conhecimento acionável, alto retorno de investimento, maior relevância para os clientes ou produtos e inovações em operações e processos de negócios.

De acordo com Song e Zhu (2015), com o desafio de lidar, de forma efetiva, com os vários problemas trazidos pelo Big Data, surgiu uma disciplina chamada Ciência de Dados, a qual

é conhecida como a disciplina que educa pessoas capazes de enfrentar esses desafios da era do Big Data. Segundo Provost and Fawcett (2013), a Ciência de Dados é definida como um conjunto de princípios fundamentais que apoiam e orientam a extração de informações e conhecimento dos dados. Essa definição de Ciência de Dados é ilustrada pela Figura 13, que mostra o surgimento da área como forma de lidar com esses desafios trazidos pelo Big Data em consequência da digitalização de sistemas e o surgimento de novas tecnologias da Indústria 4.0.

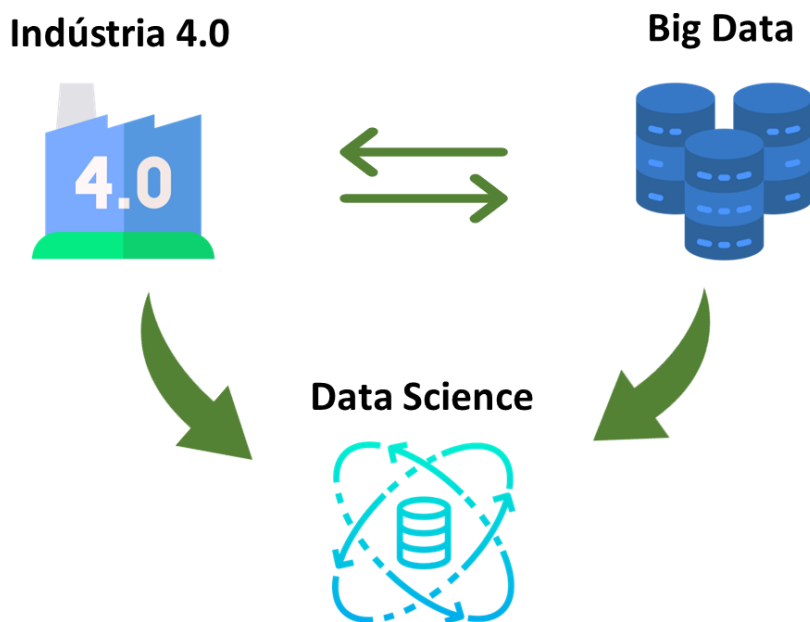


Figura 13: Surgimento da Ciência de Dados
Fonte: O autor

Em consequência da popularização da Ciência de Dados com o surgimento do Big Data, o emprego da disciplina acabou tendo maior foco no tratamento de Big Data. Porém, Song e Zhu (2015) enfatizam que programas de Ciência de Dados devem ensinar as pessoas como trabalhar com pequenas quantidades de dados. A análise de dados pequenos é aquela que não precisa, necessariamente, de tecnologias específicas do Big Data, mas de técnicas gerais como: estatísticas, mineração de dados, *machine learning*, visualização etc. (SONG; ZHU, 2015).

Nessa perspectiva, Choi, Wallace e Wang (2018) elencam duas técnicas bastante utilizadas para análise de Big Data, a estatística e o aprendizado de máquinas. A estatística é um campo de estudo consolidado que estabelece uma estrutura científica para coletar, analisar e tirar inferências e conclusões, através de métodos como a regressão estatística e a análise estatística multivariada. O aprendizado de máquina é utilizado para fornecer algoritmos para que computadores sejam capazes de adquirir conhecimento e tomar decisões após aprenderem com dados fornecidos previamente. Alguns exemplos de técnicas de aprendizado de máquinas são:

aprendizagem profunda, *support vector machine* (SVM), aprendizagem rápida, *distributed machine learning*, aprendizado de ontologias e modelos como MapReduce.

3.3. COVID-19

Segundo Cui, Li e Shi (2019), o Coronavírus já havia causado duas pandemias desde o início do século XXI, sendo a primeira o Coronavírus da Síndrome Respiratória Aguda Grave (SARS-CoV), em 2002 e 2003, e a segunda do Coronavírus da Síndrome Respiratória do Oriente Médio (MERS-CoV), em 2012. Em dezembro de 2019, um novo tipo de infecção viral surgiu na cidade de Wuhan, província de Hubei, China, “[...] e os dados iniciais de sequenciamento genômico desse vírus não correspondem aos CoVs sequenciados anteriormente, sugerindo uma nova cepa de CoV (2019-nCoV), que agora foi denominada síndrome respiratória aguda grave CoV-2 (SARS-CoV-2).” (DHAMA *et al.*, 2020, p. 1).

Com o objetivo de realizar o diagnóstico da doença causada pelo Coronavírus, instituições fizeram detecções em amostras respiratórias através de sequenciamento de última geração e métodos de RT-PCR em tempo real (ZHOU *et al.*, 2020).

Dhama *et al.* (2020) elenca os sintomas mais comuns causados pela COVID-19: febre, tosse, dispneia, expectoração, cefaleia e mialgia ou fadiga. Porém, o quadro clínico da doença parece ser abrangente, incluindo infecção assintomática, doença leve do trato respiratório superior e pneumonia viral grave com insuficiência respiratória (ZHOU *et al.*, 2020).

Desde a declaração oficial da OMS sobre a contaminação por Coronavírus subir de nível para pandemia, o cenário global começou a enfrentar as consequências causadas pelo surto de infecção da COVID-19. De acordo com o *site* Worldmeters (2022), é estimado que o Coronavírus já infectou mais de 500 milhões de pessoas ao redor do mundo, com mais de 6 milhões de mortes causadas pela doença.

3.4. Regressão logística

A determinação do tamanho de amostra e a potência estatística são fundamentais para modelos de ensaios clínicos e estudos epidemiológicos (DEMIDENKO, 2007). No contexto de pandemia causada pela COVID-19, ferramentas estatísticas foram amplamente usadas para diagnóstico e prognóstico da doença, com o objetivo de controlar seu avanço.

A regressão logística é um método estatístico muito parecido com a regressão linear, pois as duas buscam uma equação que consiga prever o resultado de uma variável binária, Y, a partir de uma ou mais variáveis de resposta, X. O grande diferencial da regressão logística é que as

variáveis de resposta, X, podem ser categóricas ou contínuas, o que dá mais liberdade para o pesquisador que está lidando com dados que não seguem uma distribuição normal (DIGANGI; HEFNER, 2013).

A regressão logística é uma técnica estatística utilizada para estipular a probabilidade de que um evento aconteça (TIBCO SOFTWARE INC., 2022). Além disso, é um modelo de análise de resultado de variáveis binárias bastante aceito e utilizado (HOSMER *et al.*, 1997). Para isso, Gonzales (2018) determina que para desenvolver esse algoritmo de predição de resultados é necessário realizar uma investigação de quais serão as variáveis independentes utilizadas no modelo.

Neste trabalho, será abordada a regressão logística binária. De acordo com TIBCO Software Inc. (2022), na regressão logística binária, apenas dois resultados serão possíveis para a variável de resposta, sendo ele 0 ou 1. Por exemplo, um aluno ser aprovado ou não em uma matéria é o resultado categórico de uma regressão logística utilizada para prever se o aluno será aprovado ou não naquela matéria.

Além disso, a regressão logística também é utilizada para solucionar questões de classificação. Segundo Urso *et al.* (2018), a classificação através da regressão logística pode utilizar uma combinação linear de mais de uma variável explicativa como argumento da função sigmoide, em que a saída é um valor entre 0 e 1. Com isso, o valor médio é considerado como fronteira para definir o que for referente à classe 0 e à classe 1. Desta forma, um resultado menor que 0,5 é categorizado como classe 0 e um resultado maior que 0,5 é definido como classe 1.

3.5. Modelo Crisp-DM

Para analisar grandes quantidades de dados, é necessária a utilização de ferramentas computacionais adequadas. Com esse propósito, surge a metodologia de mineração de dados, Crisp-DM (CRoss-Industry Standard Process for Data Mining), a partir de um consórcio formado, inicialmente, por DaimlerChrysler, SPSS e NCR. Segundo Moro, Laureano e Cortez (2011), o Crisp-DM determina um projeto como processo cíclico, de forma que os objetivos do negócio possam ser atingidos de forma mais ajustada, a partir de várias iterações durante o ciclo. A metodologia representa o ciclo de vida de um projeto de mineração de dados e é composta por seis etapas que seguem uma sequência não rígida: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* e *deployment* (CHAPMAN *et al.*, 2000), como demonstrado na figura a seguir:

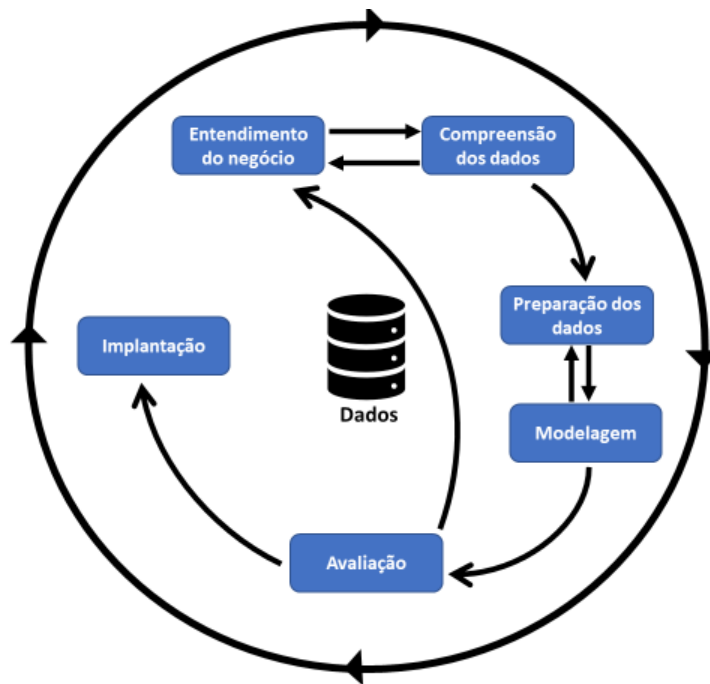


Figura 14: Etapas do Crisp-DM
 Fonte: Chapman *et al.* (2000)

Cada etapa do ciclo é composta por várias tarefas essenciais, que são definidas por Chapman *et al.* (2000). Na primeira etapa, *business understanding*, são determinados os objetivos do negócio, avalia-se a situação, determinam-se as metas de mineração de dados e é criado um plano de projeto. A segunda etapa, *data understanding*, é onde os dados são coletados, descritos, explorados e examinados de forma qualitativa. Na terceira etapa, *data preparation*, os dados são selecionados, polidos, integrados e formatados. A quarta etapa, *modeling*, é composta pela seleção da técnica de modelagem que será utilizada no projeto, seguida da criação de um *design* de teste, da construção do modelo e da sua avaliação. Na quinta etapa, *evaluation*, os resultados são avaliados para mensurar o nível de aderência do modelo com os objetivos do negócio, além disso, todo o processo é revisado e os próximos passos são definidos. Por último, na etapa de *deployment*, cria-se o plano de implantação dos resultados do projeto, planeja-se uma estratégia de monitoramento e manutenção dos resultados, gera-se o relatório final do projeto e é realizada a revisão de todo o projeto.

4. METODOLOGIA

4.1. Tipo de pesquisa

De acordo com a taxonomia proposta por Vergara (1990), este trabalho é classificado como exploratório, visto que seu tema ainda é pouco estudado e considerando que o seu surgimento se deu há pouco tempo. Por isso, o estudo tem o objetivo de tornar este tema mais visível, aperfeiçoando ideias já propostas (GIL, 2002). Quanto à sua abordagem, este estudo possui natureza quantitativa, em que os resultados retratam uma representação de toda a população, por meio da análise de dados de uma amostra (FONSECA, 2002).

Em relação às estratégias de pesquisa, foram aplicadas duas técnicas. A primeira foi uma revisão sistemática da literatura utilizando o método TEMAC de Mariano e Rocha (2017), que traz uma base mais sólida para as análises desenvolvidas, assegurando maior relevância para o trabalho. A segunda foi um levantamento de dados através de um *survey* (questionário).

Além disso, com o intuito de atingir os objetivos estabelecidos, foi empregada a técnica de regressão logística, a qual é uma das ferramentas mais indicadas para trabalhar com variáveis dependentes dicotômicas, isto é, quando a variável Y possui apenas dois resultados possíveis: infectado ou não por COVID-19 (FERNANDES *et al.*, 2021).

4.2. Local do estudo

O local para a realização da pesquisa foi a região sul do estado da Bahia, situada no Brasil. Segundo dados do IBGE de 2021, a região possui quase 2 milhões de habitantes.

4.3. Objeto de estudo

O objeto de estudo deste trabalho foram os resultados de exames de COVID-19, no período entre 2020 e 2022, de pacientes da região sul do estado da Bahia.

4.4. Instrumento de coleta de dados

Base de dados de exames de COVID-19 de um observatório de dados X.

4.5. Amostra e critérios de inclusão e exclusão

Obteve-se uma amostra inicial de 26.822 respostas. O critério de inclusão foram

pacientes que realizaram o exame de detecção de COVID-19 e os dados foram cadastrados. Já o critério de exclusão foram pacientes sem sintomas, o que gerou uma amostra final de 7.960 registros.

4.6. Ferramenta de análise e tratamento dos dados

Os dados secundários adquiridos foram analisados através da ferramenta RapidMiner, plataforma de Ciência de Dados oferecem funcionalidades de engenharia de dados, construção de modelos preditivos e criação de modelos com base em *machine learning* e inteligência artificial.

O *software* RapidMiner foi escolhido devido à sua capacidade de abordar serviços de Ciência de Dados de forma mais flexível, possibilitando a utilização de suas funcionalidades e a criação de modelos sem a necessidade de programação (RAPIDMINER, 2022).

O processo de tratamento de dados foi realizado através do Crisp-DM, metodologia de mineração e tratamento de dados estabelecida por Chapman *et al.* (2000) em um ciclo de seis etapas: entendimento do negócio, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação. A metodologia Crisp-DM será abordada com mais detalhes no próximo capítulo, aplicada ao tratamento de dados do modelo proposto.

4.7. Estrutura da metodologia

O desenho dessa pesquisa pode ser visto na figura 15.

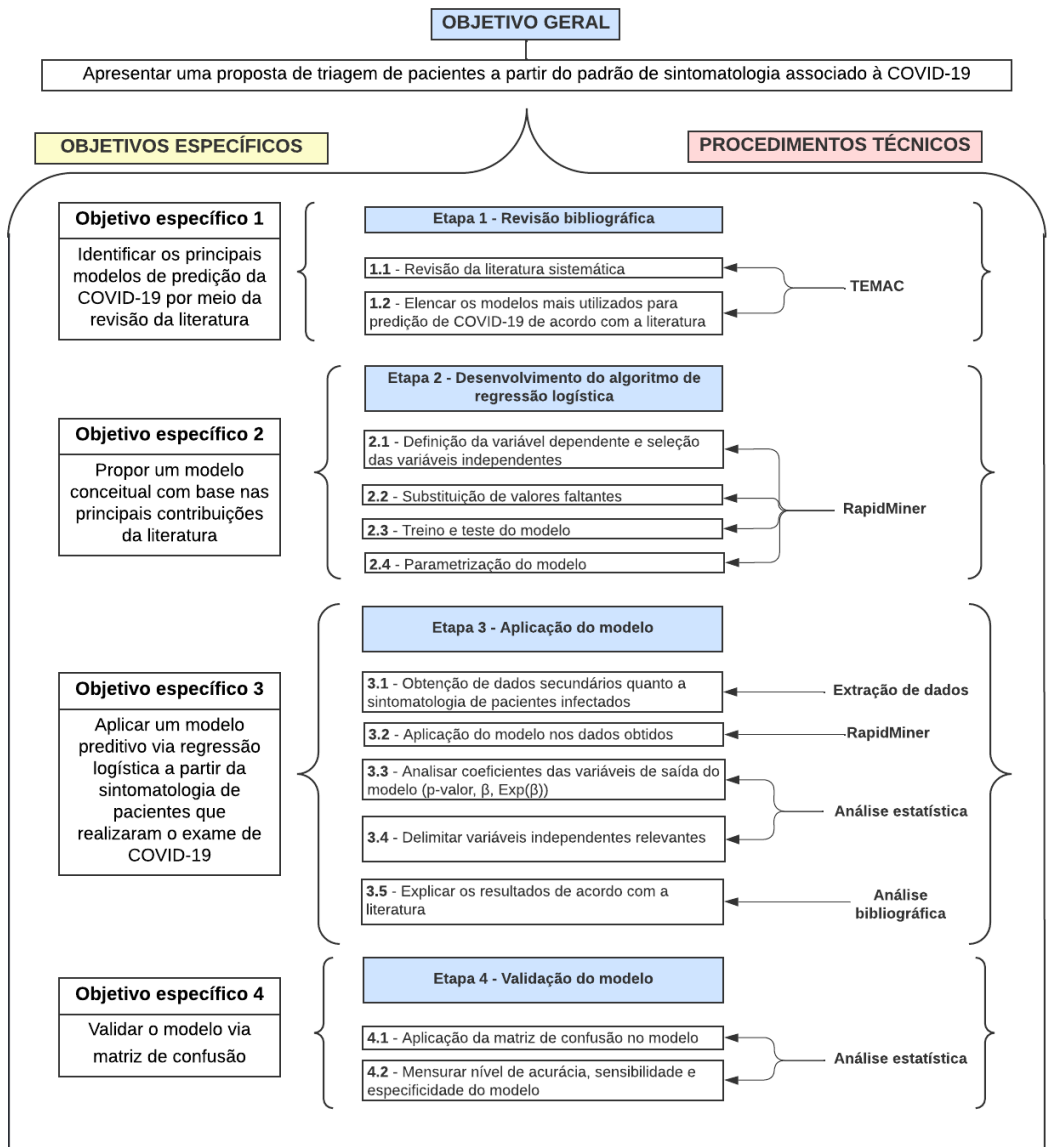


Figura 15: Estrutura da metodologia
 Fonte: O autor

5. RESULTADOS E ANÁLISES

5.1. Entendimento do negócio

Pacientes infectados pelo Coronavírus podem apresentar sintomas (CIOTTI *et al.*, 2020), que podem variar desde sintomas leves até insuficiência respiratória grave com falência múltipla de órgãos (YUKI; FUJIOGI; KOUTSOGIANNAKI, 2020). Nesse contexto, este trabalho buscou estudar a relação entre os principais sintomas (variáveis independentes) para prever quais pacientes tiveram diagnóstico positivo para a COVID-19 (variável dependente). Além de entender as definições da COVID-19, faz-se muito importante conhecer dados quantitativos dos impactos causados pela doença no Brasil e no estado da Bahia, como pode ser visto na Figura 16.

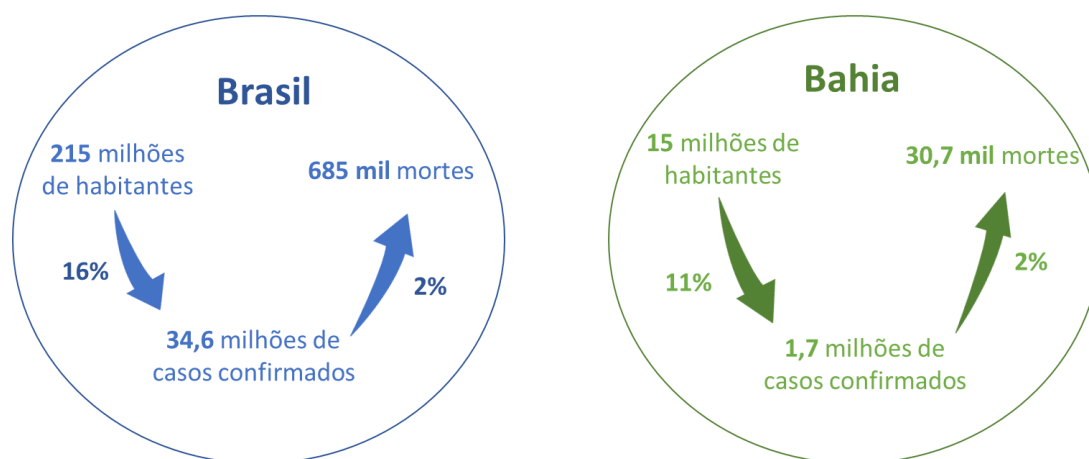


Figura 16: Impactos da COVID-19 no Brasil e na Bahia
Fonte: O autor. Dados: Ministério da Saúde (2022) e IBGE (2022)

Segundo dados do Ministério da Saúde (2022) e do IBGE (2022), o Brasil possui, neste momento, aproximadamente 34 milhões e 600 mil casos confirmados de COVID-19, o que representa quase 16% da população do país, de mais de 215 milhões de habitantes. Além disso, o número de mortes causadas pela doença já ultrapassa os 685 mil, ou seja, quase 2% dos pacientes confirmados vieram a óbito. Nesse contexto, a Bahia já possui quase 1 milhão e 700 mil casos confirmados, representando mais de 11% da população atual do estado, de pouco mais de 15 milhões de habitantes. Igualmente, quase 2% dos pacientes infectados vieram a óbito, ou seja, aproximadamente 30 mil e 700 mortes.

5.2. Compreensão dos dados

A partir do levantamento realizado via *survey*, foram obtidos os sintomas apresentados pelos pacientes. Esses dados possuem ocorrência binária e são caracterizados da seguinte forma:

confusão – desorientação e queda de funções cognitivas (concentração, atenção, memórias etc.); saturação de O₂<95% – nível de oxigenação do sangue; falta de olfato – perda parcial ou total da capacidade de sentir cheiro; falta de paladar – perda parcial ou total da capacidade de sentir sabor; expectoração/catarro no peito – acúmulo de secreções nos pulmões, traqueia e brônquios, que são expelidos através da tosse; tosse seca – mecanismo de defesa dos pulmões que busca expelir secreções acumuladas; fadiga – desgaste/cansaço excessivo; dispneia/falta de ar – respiração difícil e desagradável; rinorreia/coriza – excesso de corrimento nasal; cefaleia/dor de cabeça – sensação de dor em qualquer região da cabeça; mialgia/dor muscular – dores e desconforto nos músculos; vômito – forte eliminação do conteúdo estomacal pela boca; febre – aumento da temperatura corporal acima do limite natural, por volta de 37° C; anorexia – transtorno alimentar caracterizado pelo peso abaixo do normal e perspectiva distorcida do próprio corpo; diarreia – aumento da frequência de evacuações e fezes mais aquosas; dor no corpo – dor generalizada em qualquer região do corpo; manchas vermelhas no corpo – lesão na pele de cor avermelhada; calafrio – sensação de frio pelo corpo e contração involuntária dos músculos; náusea – sensação de enjoo e ânsia de vômito; dor no peito – dor generalizada ou pressão no peito; dor abdominal – dor da parede muscular do abdômen; dificuldade para deglutir – dor ou desconforto ao engolir; gânglios linfáticos aumentados – linfonodos inchados; taquicardia – aumento dos batimentos cardíacos; tontura – alteração do equilíbrio e posicionamento; tiragem intercostal – retração da musculatura entre as costelas durante a inspiração; faringalgia/dor de garganta – dor ou irritação na garganta; batimentos das asas do nariz – alargamento da entrada das narinas durante a respiração.

Além dos sintomas, uma variável de saída do modelo é fornecida, determinada pelo resultado do diagnóstico de infecção por COVID-19: “detectado” ou “não detectado”. Essa caracterização pode ser observada no Quadro 3.

Quadro 3: Definição dos dados

Sintomas	Ocorrência		Tipo
	Sim	Não	
Confusão	Sim	Não	Binário
Saturação de O ₂ <95%	Sim	Não	Binário
Falta de olfato	Sim	Não	Binário
Falta de paladar	Sim	Não	Binário
Expectoração/catarro no peito	Sim	Não	Binário
Tosse seca	Sim	Não	Binário
Fadiga	Sim	Não	Binário
Dispneia/falta de ar	Sim	Não	Binário
Rinorreia/coriza	Sim	Não	Binário
Cefaleia/dor de cabeça	Sim	Não	Binário
Mialgia/dor muscular	Sim	Não	Binário
Vômito	Sim	Não	Binário
Febre	Sim	Não	Binário
Anorexia	Sim	Não	Binário
Diarreia	Sim	Não	Binário
Dor no corpo	Sim	Não	Binário
Manchas vermelhas no corpo	Sim	Não	Binário
Calafrios	Sim	Não	Binário
Náusea	Sim	Não	Binário
Dor no peito	Sim	Não	Binário
Dor abdominal	Sim	Não	Binário
Dificuldade para deglutir	Sim	Não	Binário
Gânglios linfáticos aumentados	Sim	Não	Binário
Taquicardia	Sim	Não	Binário
Tontura	Sim	Não	Binário
Tiragem intercostal	Sim	Não	Binário
Faringalgia/dor de garganta	Sim	Não	Binário
Batimentos das asas do nariz	Sim	Não	Binário
Resultado do diagnóstico	Detectado	Não detectado	Binário

Fonte: O autor

Após compreender os dados, a próxima etapa é prepará-los para o desenvolvimento do modelo de predição.

5.3. Preparação dos dados

5.3.1. Buscar valores faltantes

Para tratamento dos dados da pesquisa, primeiramente foram identificados aqueles valores faltantes, ou seja, perguntas do *survey* que, por algum motivo desconhecido, não foram respondidas.

5.3.2. Substituir dados faltantes

O segundo passo é recodificar essas variáveis ausentes, substituindo os valores faltantes pela média de valores daquela variável, como sugere a literatura.

5.3.3. Variável dependente

A variável dependente é definida pelo resultado do exame, em que só há duas opções: detectado e não detectado.

5.3.4. Variáveis independentes

As variáveis independentes são definidas pelos 28 sintomas, apresentados ou não pelos pacientes, previamente listados no Quadro 3.

5.4. Modelagem

Nesta etapa, o modelo foi desenvolvido. Todos os processos, desde a entrada dos dados até a saída do resultado, estão ilustrados na Figura 17.

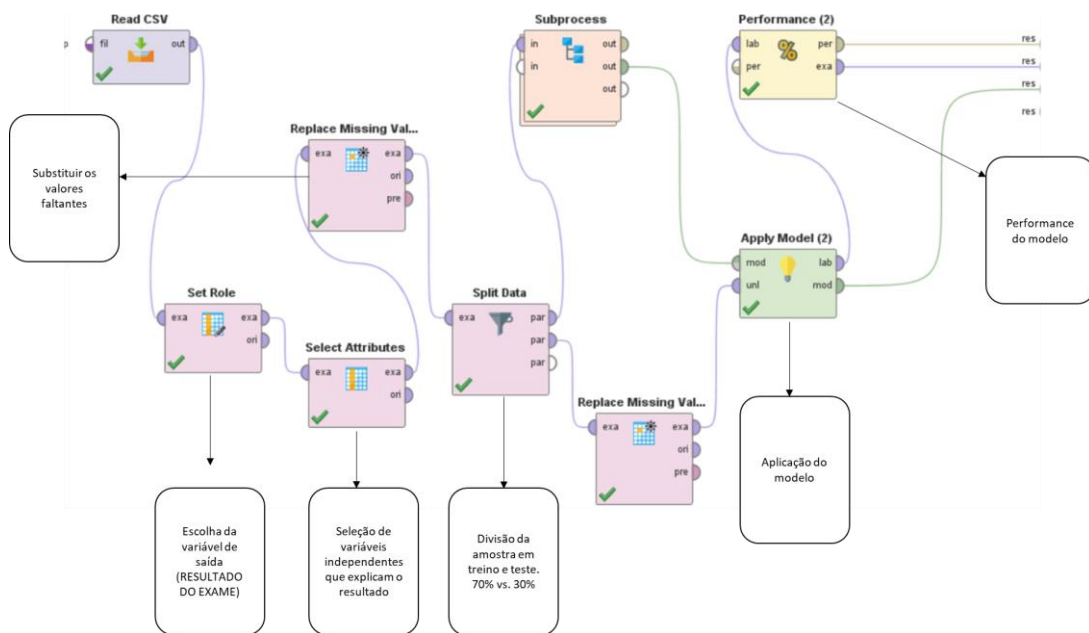


Figura 17: Etapas do desenvolvimento do modelo de regressão logística
 Fonte: O autor, utilizando o *software* RapidMiner

Inicialmente, foram inseridos os dados em formato CSV. Em seguida, a variável dependente foi definida e, em sequência, as variáveis independentes foram selecionadas. Logo depois, os valores faltantes foram substituídos, como mencionado na etapa de preparação dos dados. Seguidamente, um total de 30% da amostra foi selecionado para treinar o modelo de regressão logística e os 70% restantes foram utilizados para testar o modelo. Em seguida, foram substituídos os valores faltantes (30% utilizados para treinar o algoritmo). Em paralelo, uma série de subprocessos foi desenvolvida para parametrizar o modelo, como pode ser observado na Figura 18.

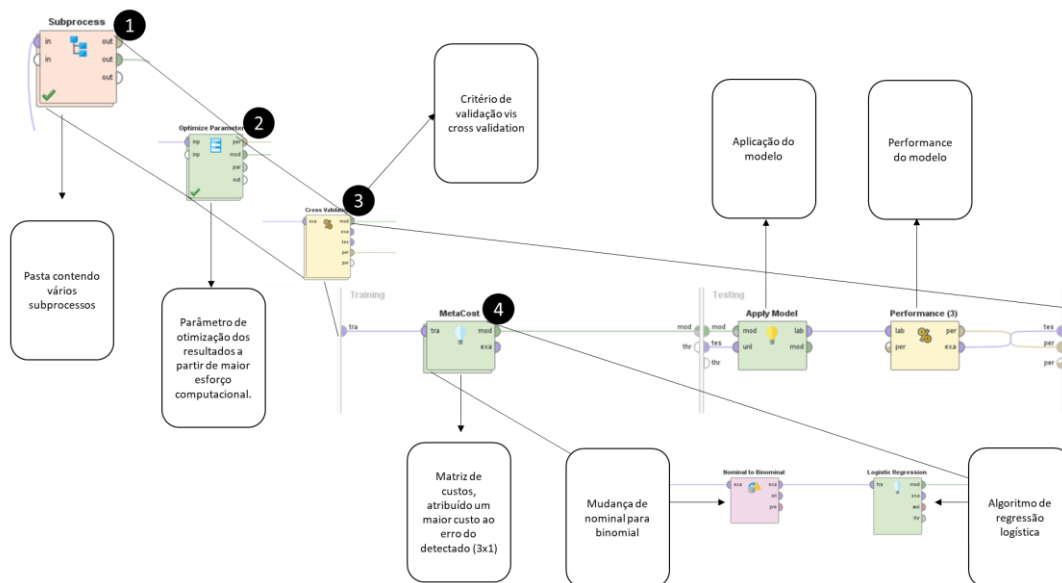


Figura 18: Subprocessos do modelo de regressão logística
 Fonte: O autor, utilizando o *software* RapidMiner

Com a finalidade de obter alto nível de sensibilidade no modelo, foram definidos parâmetros de otimização para, assim, conseguir uma probabilidade maior de acertar o resultado “detectado” para COVID-19, no maior número possível de pacientes. Esses parâmetros foram estabelecidos por critérios de validação que, por sua vez, foram determinados por uma matriz de custos, em que o modelo teve um custo três vezes maior para cada vez que errasse um “falso detectado” do que ao errar um “falso não detectado”. Com isso, o modelo prioriza o acerto de um resultado “detectado” por um *trade-off* de baixa especificidade para determinar um resultado “não detectado”. Esse *trade-off* fica mais perceptível na próxima etapa do Crisp-DM: avaliação. Em sequência, o modelo foi parametrizado como binomial, pois os valores retornados só podem ser binários. E, por fim, o algoritmo de regressão logística foi definido.

Voltando para a visão geral do modelo, ilustrado na Figura 17, o algoritmo foi aplicado e tanto os resultados quanto a sua performance foram obtidos logo depois, apresentados nas Tabelas 1 e 2, respectivamente.

Após a aplicação do modelo de regressão logística binária, os resultados mostraram os sintomas dos pacientes infectados e o quanto esses sintomas afetam a probabilidade de um paciente estar contaminado por COVID-19. O resultado pode ser verificado na Tabela 1 a seguir:

Tabela 1: Resultado da regressão logística

Atributo	β	Exponencial (β)	Aumento percentual	Aumento percentual (%)	z-valor	p-valor
Confusão	1,477	4,3801	3,380	338%	2,521	0,012
Saturação de O ₂ <95%	1,202	3,3258	2,326	233%	6,384	0,000
Falta de olfato	0,970	2,6391	1,639	164%	9,800	0,000
Falta de paladar	0,810	2,2478	1,248	125%	8,184	0,000
Expectoração/catarro no peito	0,597	1,8174	0,817	82%	3,009	0,003
Tosse seca	0,591	1,8052	0,805	81%	16,050	0,000
Fadiga	0,571	1,7705	0,771	77%	7,225	0,000
Dispneia/falta de ar	0,424	1,5275	0,527	53%	10,116	0,000
Rinorreia/coriza	0,294	1,3412	0,341	34%	6,813	0,000

Fonte: O autor

A partir do resultado encontrado, primeiramente foram descartadas as variáveis que não são significativas para o modelo, ou seja, aquelas em que o p-valor é maior que 0,05. Dentre as variáveis restantes, foram escolhidas apenas aquelas em que o coeficiente β é positivo, já que um coeficiente β negativo pode confundir a interpretação do resultado, pois nem sempre isso significa um efeito negativo sobre a variável dependente. O resultado incluindo todas as variáveis encontra-se no Apêndice A deste trabalho.

Como resultado, obtêm-se nove variáveis significantes para o modelo. Dentre elas, o sintoma de confusão é o que tem maior impacto no diagnóstico do paciente, pois um paciente com esse sintoma tem 4,3 vezes mais chances de estar infectado. Além disso, segundo o modelo, aqueles pacientes que apresentam saturação de O₂<95% possuem 3,3 vezes mais chances de estarem com COVID-19, sendo estes os dois principais sintomas que o potencial paciente infectado pode sentir. Entre os sintomas menos impactantes, o sintoma de rinorreia/coriza aumenta em 34% a probabilidade de um paciente estar infectado. Todas as nove variáveis estão elencadas de forma mais visível na Figura 19.

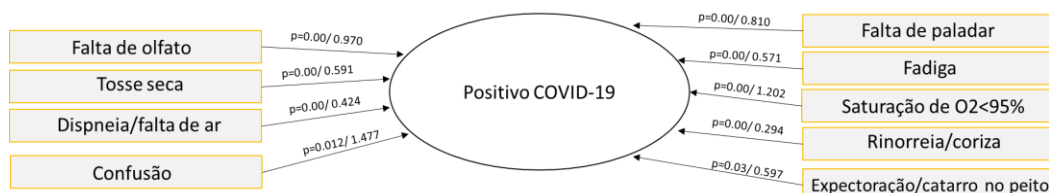


Figura 19: Sintomas mais impactantes para o resultado positivo de COVID-19

Fonte: O autor

A importância da sintomatologia como forma de predição da COVID-19 é respaldada

por Zhou *et al.* (2020), Chen *et al.* (2020) e Huang *et al.* (2020), que buscaram esclarecer as características clínicas e epidemiológicas de pacientes infectados. Dentre os principais sintomas identificados, muitos possuem correspondência com os resultados obtidos neste trabalho, sendo eles: tosse, expectoração/produção de escarro, fadiga, confusão, rinorreia/coriza e dispneia/falta de ar. Além disso, Santos *et al.* (2021) e Vaira *et al.* (2020) mostram em seus estudos que disfunções olfativas e gustativas são sintomas relativamente comuns em pacientes contaminados.

Em outras abordagens, Jehi *et al.* (2020) também utiliza diversos sintomas como variáveis para o modelo de predição de COVID-19 proposto em seu estudo. Já Iwendi *et al.* (2020) propõe um modelo para mensurar a gravidade dos casos e prever a recuperação ou morte do paciente. O trabalho de Knight *et al.* (2020) usou o nível de saturação de oxigênio no sangue como uma das variáveis para desenvolver uma pontuação de risco pragmática para estratificar pacientes internados em diferentes grupos.

5.5. Avaliação

Para medir os níveis de acurácia, sensibilidade e especificidade do modelo, foi desenvolvida uma matriz de confusão, como exposto na Tabela 2.

Tabela 2: Matriz confusão

Acurácia: 33,99%			
	<i>True</i> "Detectado"	<i>True</i> "Não Detectado"	<i>Class precision</i>
Predição "Detectado"	2701	5252	33,96%
Predição "Não Detectado"	2	5	71,43%
<i>Class recall</i>	99,93%	0,10%	

Fonte: O autor

Apesar de apresentar uma acurácia de apenas 33,99%, o objetivo do modelo é prever quais pacientes têm maiores chances de estarem infectados por COVID-19. Por esse motivo, a sensibilidade de 99,93% traz muito mais relevância para atingir o objetivo de identificar potenciais pacientes infectados. Em contrapartida, a especificidade foi de apenas 0,1%, o que já era esperado, visto que existe um *trade-off* entre a sensibilidade e a especificidade (FERNANDES *et al.*, 2021).

A sensibilidade de 99,93% do modelo proporciona significativa relevância para ele, o que na prática poderia fazer muita diferença para identificar potenciais pacientes infectados, garantindo mais eficiência em processos de triagem prévia.

5.6. Implantação

Uma quantidade crescente de pacientes que buscam serviços de emergência de assistência médica pode levar à superlotação, prejudicando a organização da instituição (AZEREDO *et al.*, 2015). Nesse contexto, Duro e Lima (2010) destacam a importância de métodos de triagem hospitalar para organizar o fluxo de atendimento dos pacientes. Uma escala de triagem bastante utilizada para melhorar essa organização é o protocolo de Manchester, forma de classificar o risco associado aos pacientes em cinco categorias representadas por cores: emergência (vermelho), muito urgente (laranja), urgente (amarelo), pouco urgente (verde) e não urgente (azul) (AZEREDO *et al.*, 2015).

Harrell, Lee e Mark (1996) e Copas (1983) citam que ferramentas de regressão multivariável são frequentemente utilizadas para prever resultados clínicos em estudos da área. Essas ferramentas estão sendo cada vez mais aplicadas para prever o diagnóstico e o prognóstico de pacientes com uma diversidade de doenças (HARRELL *et al.*, 1984).

Dessa forma, propõe-se possível implantação prática do modelo de predição deste trabalho aplicando-o ao protocolo de Manchester, como mostrado na Figura 20.

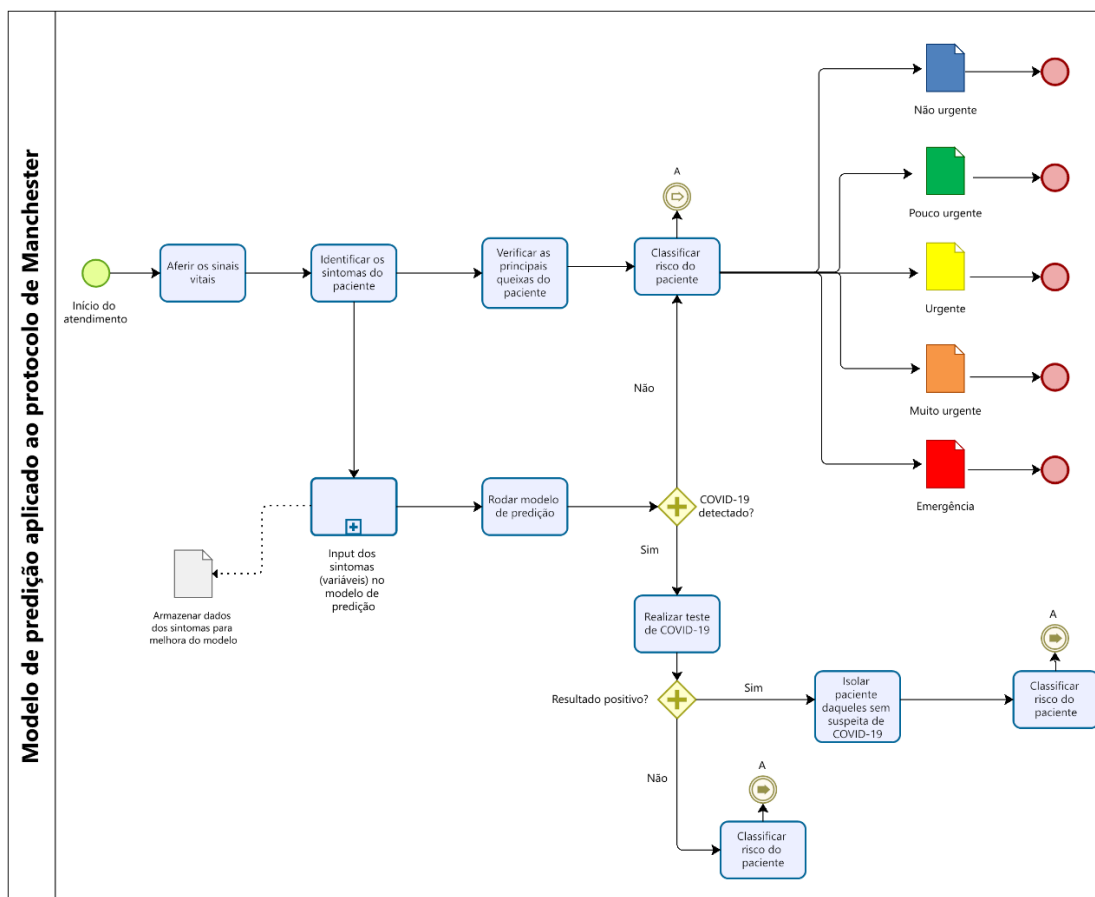


Figura 20: Modelo de predição aplicado ao protocolo de Manchester
 Fonte: O autor, utilizando o *software* Bizagi Modeler

O modelo de predição atua em paralelo ao protocolo de Manchester, completando o processo de triagem em um cenário de pandemia. Logo no início do atendimento, após a aferição dos sinais vitais do paciente, realiza-se a identificação dos sintomas apresentados. Com a identificação dos sintomas, verificam-se as principais queixas do paciente ao mesmo tempo em que o modelo é aplicado aos dados sintomatológicos. Se o algoritmo apontar resultado negativo, o paciente continua o processo de classificação de riscos normalmente, seguindo o protocolo de Manchester. Caso contrário, o paciente é direcionado para a realização do teste de COVID-19 imediatamente para confirmar o resultado. Se o resultado for negativo, o paciente retorna ao protocolo novamente. Caso contrário, o paciente é isolado em ambiente adequado, diminuindo o risco de contaminação de outras pessoas. Só então o paciente passa pelo processo de classificação de risco para verificar possíveis complicações em seu quadro e determinar a prioridade do seu atendimento.

Desta forma, a instituição teria maior agilidade organizacional no atendimento dos

pacientes nos setores específicos, utilizando os recursos necessários de forma mais efetiva. Essa otimização possibilita capacidade maior de atendimentos, ganhos de eficiência na gestão de tempo, força de trabalho e custos, visto que os profissionais da saúde terão mais objetividade na hora de prestar seus serviços de assistência médica, empregando menos recursos para isso. Além disso, a instituição oferece maior segurança para o ambiente e para as pessoas sem suspeita de infecção. Vale frisar que o modelo não substitui o diagnóstico médico, de forma que a ferramenta serve apenas para agilizar o processo, possibilitando um atendimento mais organizado e eficiente.

6. CONSIDERAÇÕES FINAIS, LIMITAÇÕES E FUTURAS LINHAS DE PESQUISA

O problema definido neste estudo era descobrir como a sintomatologia está associada à infecção de COVID-19. Ele foi resolvido a partir da investigação dos sintomas apresentados por pacientes infectados, realizada por meio da inserção desses dados em um modelo de predição, com base na literatura. Foi adotado o método Crisp-DM para o tratamento dos dados e o desenvolvimento do algoritmo de regressão logística.

O objetivo de apresentar uma proposta de triagem de pacientes a partir do padrão de sintomatologia associado à COVID-19 foi realizado, visto que os objetivos específicos de identificação dos principais modelos de predição da COVID-19 por meio da revisão da literatura, a proposição de um modelo conceitual com base nas principais contribuições da literatura, a aplicação de um modelo preditivo via regressão logística a partir da sintomatologia de pacientes que realizaram o exame de COVID-19 e a validação do modelo via matriz de confusão foram alcançados. Além disso, o algoritmo proposto foi validado pela matriz de confusão, apresentando uma sensibilidade de 99,93%, e, por fim, aplicado com êxito, promovendo resultados claros e comprovando a relevância da revisão bibliográfica acerca dos principais modelos de predição da COVID-19, por meio do método TEMAC. Por fim, a pesquisa mostrou a importância da Engenharia de Produção aplicada à área da saúde, propondo uma ferramenta que não é capaz de substituir o diagnóstico médico, porém, busca cumprir o seu papel de otimizar a organização de instituições de saúde integrando informação, tecnologia e serviços.

Um fator limitante desta pesquisa foi o *trade-off* entre a sensibilidade e a especificidade do modelo que, para conseguir maior probabilidade de prever pacientes infectados, não possui boa especificidade para prever pacientes não infectados. De maneira prática, essa limitação poderia implicar em uma “superlotação” na hora de realizar testes de COVID-19 com o intuito de constatar o resultado da triagem proposta pelo modelo, visto que praticamente todos os pacientes que apresentassem os sintomas citados seriam considerados infectados de acordo com o algoritmo.

Propõe-se, para futuras linhas de pesquisa, encontrar formas de minimizar ou eliminar essa limitação, por intermédio de outros métodos que possam contribuir para melhorar o modelo proposto, buscando uma especificidade melhor sem a necessidade de sacrificar a sensibilidade do modelo.

REFERÊNCIAS BIBLIOGRÁFICAS

ACETO, Giuseppe; PERSICO, Valerio; PESCAPÉ, Antonio. The role of information and communication technologies in healthcare: taxonomies, perspectives, and challenges. **Journal of Network and Computer Applications**, v. 107, p. 125-154, 2018.

AHMAD, Khairol Amali Bin *et al.* Emerging trends and evolutions for smart city healthcare systems. **Sustainable Cities and Society**, v. 80, p. 103695, 2022.

AHMED, Ejaz *et al.* The role of big data analytics in Internet of Things. **Computer Networks**, v. 129, p. 459-471, 2017.

AL-FUQAHA, Ala *et al.* Internet of Things: a survey on enabling technologies, protocols, and applications. **IEEE Communications Surveys & Tutorials**, v. 17, n. 4, p. 2347-2376, 2015.

AMORIM DOS SANTOS, J. *et al.* Oral manifestations in patients with COVID-19: a living systematic review. **Journal of Dental Research**, v. 100, n. 2, p. 141-154, 2021.

ARMBRUST, Michael *et al.* Above the clouds: a Berkeley view of cloud computing. **Technical Report UCB/EECS-2009-28**, EECS Department, University of California, Berkeley, 2009.

ATZORI, Luigi; IERA, Antonio; MORABITO, Giacomo. The Internet of Things: a survey. **Computer Networks**, v. 54, n. 15, p. 2787-2805, 2010.

AZEREDO, Thereza Raquel Machado *et al.* Efficacy of the Manchester Triage System: a systematic review. **International Emergency Nursing**, v. 23, n. 2, p. 47-52, 2015.

BAHRIN, Mohd Aiman Kamarul *et al.* Industry 4.0: a review on industrial automation and robotic. **Jurnal Teknologi**, v. 78, n. 6-13, 2016.

BRASIL. Ministério da Saúde. **Painel Coronavírus**, 2022. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 21 set. 2022.

BUCZAK, Anna L.; GUVEN, Erhan. A survey of data mining and machine learning methods for cyber security intrusion detection. **IEEE Communications Surveys & Tutorials**, v. 18, n. 2, p. 1153-1176, 2015.

CATARINUCCI, Luca *et al.* An IoT-aware architecture for smart healthcare systems. **IEEE Internet of Things Journal**, v. 2, n. 6, p. 515-526, 2015.

CHAPMAN, Pete *et al.* CRISP-DM 1.0: step-by-step data mining guide. **SPSS Inc**, v. 9, n. 13, p. 1-73, 2000.

CHAPMAN, T. *et al.* From 3G to 4G: background and motivation of 3G evolution. **HSPA Evolution**, p. 3-20, 2015.

CHEN, Nanshan *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. **The Lancet**, v. 395, n. 10223, p. 507-513, 2020.

CHOI, Tsan-Ming; WALLACE, Stein W.; WANG, Yulan. Big data analytics in operations management. **Production and Operations Management**, v. 27, n. 10, p. 1868-1883, 2018.

CIOTTI, Marco *et al.* The COVID-19 pandemic. **Critical reviews in clinical laboratory sciences**, v. 57, n. 6, p. 365-388, 2020.

COPAS, John B. Regression, prediction and shrinkage. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 45, n. 3, p. 311-335, 1983.

CUI, Jie; LI, Fang; SHI, Zheng-Li. Origin and evolution of pathogenic coronaviruses. **Nature Reviews Microbiology**, v. 17, n. 3, p. 181-192, 2019.

DA FONSECA, João José Saraiva. **Apostila de Metodologia da Pesquisa Científica**. João José Saraiva da Fonseca, 2002.

DAYAN, Ittai *et al.* Federated learning for predicting clinical outcomes in patients with COVID-19. **Nature medicine**, v. 27, n. 10, p. 1735-1743, 2021.

DEMIDENKO, Eugene. Sample size determination for logistic regression revisited. **Statistics in Medicine**, v. 26, n. 18, p. 3385-3397, 2007.

DHAMA, Kuldeep *et al.* Coronavirus disease 2019 – COVID-19. **Clinical Microbiology Reviews**, v. 33, n. 4, p. e00028-20, 2020.

DIGANGI, Elizabeth A.; HEFNER, Joseph T. Ancestry estimation. **Research Methods in Human Skeletal Biology**, p. 117-149, 2013.

DONG, Ensheng; DU, Hongru; GARDNER, Lauren. An interactive web-based dashboard to track COVID-19 in real time. **The Lancet Infectious Diseases**, v. 20, n. 5, p. 533-534, 2020.

DURO, Carmen Lucia Mottin; LIMA, Maria Alice Dias da Silva. O papel do enfermeiro nos sistemas de triagem em emergências: análise da literatura. **Online Brazilian Journal of Nursing**, Rio de Janeiro, v. 9, n. 3, f. 1-13, 2010.

ECCLESTON, Christopher *et al.* Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. **Pain**, v. 161, n. 5, p. 889, 2020.

ERBOZ, Gizem. How to define Industry 4.0: main pillars of Industry 4.0. **Managerial Trends in the Development of Enterprises in Globalization Era**, v. 761, p. 767, 2017.

FAGHERAZZI, Guy *et al.* Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. **Journal of Medical Internet Research**, v. 22, n. 6, p. e19284, 2020.

FERNANDES, Antônio Alves Tôrres *et al.* Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**, v. 28, 2021.

FRANK, Alejandro Germán; DALENOGARE, Lucas Santos; AYALA, Néstor Fabián. Industry 4.0 technologies: implementation patterns in manufacturing companies. **International Journal of Production Economics**, v. 210, p. 15-26, 2019.

GARTNER, I. T. **Big data**, 2012. Disponível em: <https://www.gartner.com/en/information-technology/glossary/big-data>. Acesso em: 15 abr. 2022.

GIL, Antonio Carlos *et al.* **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GINSBURG, Geoffrey S.; PHILLIPS, Kathryn A. Precision medicine: from science to value. **Health Affairs**, v. 37, n. 5, p. 694-701, 2018.

GONZALEZ, Leandro de Azevedo. **Regressão logística e suas aplicações**, 2018.

HARRELL JR, Frank E. *et al.* Regression modelling strategies for improved prognostic prediction. **Statistics in Medicine**, v. 3, n. 2, p. 143-152, 1984.

HARRELL JR, Frank E.; LEE, Kerry L.; MARK, Daniel B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. **Statistics in Medicine**, v. 15, n. 4, p. 361-387, 1996.

HARRINGTON, Christina N.; KOON, Lyndsie Marie; ROGERS, Wendy A. Design of health information and communication technologies for older adults. *In: Design for Health*. Academic Press, p. 341-363, 2020.

HOSMER, David W. *et al.* A comparison of goodness-of-fit tests for the logistic regression model. **Statistics in Medicine**, v. 16, n. 9, p. 965-980, 1997.

HUANG, Chaolin *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. **The Lancet**, v. 395, n. 10223, p. 497-506, 2020.

IBGE. **Projeção da população do Brasil e das unidades da Federação**, 2022. Disponível em: <https://www.ibge.gov.br/apps/populacao/projecao/index.html>. Acesso em: 21 set. 2022.

ISLAM, SM Riazul *et al.* The Internet of Things for health care: a comprehensive survey. **IEEE Access**, v. 3, p. 678-708, 2015.

IWENDI, Celestine *et al.* COVID-19 patient health prediction using boosted random forest algorithm. **Frontiers in Public Health**, v. 8, p. 357, 2020.

JEHI, Lara *et al.* Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 patients. **Chest**, v. 158, n. 4, p. 1364-1375, 2020.

JORDAN, Michael I.; MITCHELL, Tom M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255-260, 2015.

KNIGHT, Stephen R. *et al.* Risk stratification of patients admitted to hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. **BMJ**, v. 370, 2020.

KUMAR, Sachin; VEER, Karan. Forecasting of COVID-19 cases using machine learning approach. **Current Respiratory Medicine Reviews**, v. 16, n. 4, p. 240-245, 2020.

KUSIAK, Andrew. Smart manufacturing. **International Journal of Production Research**, v. 56, n. 1-2, p. 508-517, 2017. BAHRIN, Mohd Aiman Kamarul *et al.* Industry 4.0: A review on industrial automation and robotic. **Jurnal teknologi**, v. 78, n. 6-13, 2016.

LAI, Shengjie *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. **Nature**, v. 585, n. 7825, p. 410-413, 2020.

LANEY, Doug *et al.* 3D data management: controlling data volume, velocity and variety. **Meta Group Research Note**, v. 6, n. 70, p. 1, 2001.

LASI, Heiner *et al.* Industry 4.0. **Business & Information Systems Engineering**, v. 6, n. 4, p. 239-242, 2014.

MARIANO, Ari Melo *et al.* Desafios da Indústria 4.0: integração da literatura por meio do TEMAC. *In: II Congresso Iberoamericano Ajicede*, 2019.

MARIANO, Ari Melo; ROCHA, Maíra Santos. Revisão da literatura: apresentação de uma abordagem integradora. *In: AEDEM International Conference*, p. 427-442, 2017.

MERTENS, Gaëtan *et al.* Fear of the coronavirus (COVID-19): predictors in an online study conducted in March 2020. **Journal of Anxiety Disorders**, v. 74, p. 102258, 2020.

MINOLI, Daniel; SOHRABY, Kazem; OCCHIOGROSSO, Benedict. IoT considerations, requirements, and architectures for smart buildings – energy optimization and next-generation building management systems. **IEEE Internet of Things Journal**, v. 4, n. 1, p. 269-283, 2017.

MORO, Sergio; LAUREANO, Raul; CORTEZ, Paulo. **Using data mining for bank direct marketing**: an application of the crisp-dm methodology, 2011.

PROVOST, Foster; FAWCETT, Tom. Data science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51-59, 2013.

RAHMANI, Amir M. *et al.* Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: a fog computing approach. **Future Generation Computer Systems**, v. 78, p. 641-658, 2018.

RANZANI, Otavio T. *et al.* Effectiveness of the CoronaVac vaccine in older adults during a gamma variant associated epidemic of covid-19 in Brazil: test negative case-control study. **BMJ**, v. 374, 2021.

RAPIDMINER. **Model building**, 2022. Disponível em: <https://rapidminer.com/platform/model-building/>. Acesso em: 10 set. 2022.

RAYMUNDO, Carlos Eduardo *et al.* Spatial analysis of COVID-19 incidence and the sociodemographic context in Brazil. **PLoS One**, v. 16, n. 3, p. e0247794, 2021.

RÜßMANN, Michael *et al.* Industry 4.0: the future of productivity and growth in manufacturing industries. **Boston Consulting Group**, v. 9, n. 1, p. 54-89, 2015.

SANCHEZ-PINTO, L. Nelson; LUO, Yuan; CHURPEK, Matthew M. Big data and data science in critical care. **Chest**, v. 154, n. 5, p. 1239-1248, 2018.

SCHWAB, Klaus. The fourth industrial revolution. **Currency**, 2017.

SONG, Il-Yeol; ZHU, Yongjun. Big data and data science: opportunities and challenges of iSchools. **Journal of Data and Information Science**, v. 2, n. 3, p. 1-18, 2017.

SONG, Il-Yeol; ZHU, Yongjun. Big data and data science: what should we teach? **Expert Systems**, v. 33, n. 4, p. 364-373, 2015.

TIBCO SOFTWARE INC. **O que é regressão logística?**, 2022. Disponível em: <https://www.tibco.com/pt-br/reference-center/what-is-logistic-regression#:~:text=A%20regress%C3%A3o%20log%C3%ADstica%20%C3%A9%20um,ajudar%20a%20criar%20previs%C3%B5es%20precisas>. Acesso em: 9 maio 2022.

URSO, Alfonso *et al.* Data mining: prediction methods. **Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.**, v. 1, p. 413-430, 2018.

VAID, Akhil *et al.* Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach. **JMIR Medical Informatics**, v. 9, n. 1, p. e24207, 2021.

VAIRA, Luigi Angelo *et al.* Objective evaluation of anosmia and ageusia in COVID-19 patients: single-center experience on 72 cases. **Head & Neck**, v. 42, n. 6, p. 1252-1258, 2020.

VERGARA, Sylvia Constant. **Tipos de pesquisa em administração**. 1990.

WANG, Honggang *et al.* Guest editorial: special issue on Internet of Things for smart and connected health. **IEEE Internet of Things Journal**, v. 2, n. 1, p. 1-4, 2015.

WANG, Huwen *et al.* Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. **Cell Discovery**, v. 6, n. 1, p. 1-8, 2020.

WANG, Shiyong *et al.* Towards smart factory for Industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. **Computer Networks**, v. 101, p. 158-168, 2016.

WORLDOMETER. **COVID-19 coronavírus pandemic**, 2022. Disponível em: <https://www.worldometers.info/coronavirus/>. Acesso em: 20 abr. 2022.

WU, Joseph T.; LEUNG, Kathy; LEUNG, Gabriel M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. **The Lancet**, v. 395, n. 10225, p. 689-697, 2020.

WYNANTS, Laure *et al.* Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. **BMJ**, v. 369, 2020.

YUKI, Koichi; FUJIOGI, Miho; KOUTSOGIANNAKI, Sophia. COVID-19 pathophysiology: a review. **Clinical Immunology**, v. 215, p. 108427, 2020.

ZANELLA, Andrea *et al.* Internet of Things for smart cities. **IEEE Internet of Things Journal**, v. 1, n. 1, p. 22-32, 2014.

ZHANG, Juanjuan *et al.* Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. **Science**, v. 368, n. 6498, p. 1481-1486, 2020.

ZHOU, Fei *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. **The Lancet**, v. 395, n. 10229, p. 1054-1062, 2020.

APÊNDICES

APÊNDICE A

Atributo	β	Exponencial (β)	Aumento percentual	Aumento percentual (%)	z-valor	p-valor
Confusão	1,477	4,3801	3,380	338%	2,521	0,012
Saturação de O ₂ <95%	1,202	3,3258	2,326	233%	6,384	0,000
Falta de olfato	0,970	2,6391	1,639	164%	9,800	0,000
Falta de paladar	0,810	2,2478	1,248	125%	8,184	0,000
Expectoração/catarro no peito	0,597	1,8174	0,817	82%	3,009	0,003
Tosse seca	0,591	1,8052	0,805	81%	16,050	0,000
Fadiga	0,571	1,7705	0,771	77%	7,225	0,000
Dispneia/falta de ar	0,424	1,5275	0,527	53%	10,116	0,000
Rinorreia/coriza	0,294	1,3412	0,341	34%	6,813	0,000
Cefaleia /dor de cabeça	-0,083	0,9207	-0,079	-8%	-2,053	0,040
Mialgia/dor muscular	-0,278	0,7574	-0,243	-24%	-3,553	0,000
Vômito	-0,542	0,5814	-0,419	-42%	-3,423	0,001
Febre	-0,642	0,5264	-0,474	-47%	-	0,000
Anorexia	-1,211	0,2978	-0,702	-70%	17,159	0,000
Diarreia	-0,151	0,8603	-0,140	-14%	-3,853	0,051
Dor no corpo	-0,136	0,8726	-0,127	-13%	-1,954	0,101
Manchas vermelhas no corpo	0,787	2,1960	1,196	120%	-1,641	0,126
Calafrios	-0,152	0,8593	-0,141	-14%	1,532	0,241
Náusea	0,167	1,1814	0,181	18%	-1,172	0,293
Dor no peito	0,229	1,2570	0,257	26%	1,052	0,302
Dor abdominal	-0,198	0,8200	-0,180	-18%	1,033	0,339
Dificuldade para deglutir	0,286	1,3306	0,331	33%	-0,957	0,436
Gânglios linfáticos aumentados	0,335	1,3979	0,398	40%	0,779	0,627
Taquicardia	7,020	1119,0027	1118,003	111800%	0,486	0,754
Tontura	0,079	1,0822	0,082	8%	0,290	0,772
Tiragem intercostal	5,723	305,9597	304,960	30496%	0,145	0,885
Faringalgia/dor de garganta	-0,005	0,9946	-0,005	-1%	-0,134	0,893
Batimentos das asas do nariz	0,133	1,1428	0,143	14%	0,108	0,914