



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA
DA COMPUTAÇÃO



Visualização, interação e análise de dados do COVID-19 nos sistemas prisionais

Trabalho de Dissertação

Lucas dos Santos Nunes

São Cristóvão – Sergipe

2023

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Lucas dos Santos Nunes

**Visualização, interação e análise de dados do COVID-19 nos
sistemas prisionais**

Trabalho de Dissertação submetido ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de Mestrado em Ciência da Computação.

Orientador(a): Edward David Moreno Ordonez
Coorientador(a): Glauco de Figueiredo Carneiro

São Cristóvão – Sergipe

2023

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE**

N972v Nunes, Lucas dos Santos
Visualização, interação e análise de dados do COVID-19 nos sistemas prisionais / Lucas dos Santos Nunes ; orientador Edward David Moreno Ordonez. - São Cristóvão, 2023.
96 f. : il.

Dissertação (mestrado em Ciência da Computação) – Universidade Federal de Sergipe, 2023.

1. Prisões. 2. Algoritmos computacionais. 3. Aprendizado do computador. 4. COVID – 19 (Doença). I. Santos, Naldisson dos orient. II. Título.

CDU 004.8



**UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**Ata da Sessão Solene de Defesa da Dissertação do
Curso de Mestrado em Ciência da Computação-UFS.
Candidato: Lucas dos Santos Nunes**

Em 22 dias do mês de agosto do ano de dois mil e vinte três, com início às 14h00min, realizou-se na Sala de Reunião do PROCC da Universidade Federal de Sergipe, na Cidade Universitária Prof. José Aloísio de Campos, a Sessão Pública de Defesa de Dissertação de Mestrado do candidato **Lucas dos Santos Nunes**, que desenvolveu o trabalho intitulado: **“Visualização. Interação e Análise de dados do COVID-19 no Sistema Prisional prisional”**, sob a orientação do Prof. Dr. **Edward David Moreno Ordonez** e coorientação do Prof. Dr. **Glauco Figueiredo Carneiro**. A Sessão foi presidida pelo Prof. Dr. **Edward David Moreno Ordonez** (PROCC/UFS), que após a apresentação da dissertação passou a palavra aos outros membros da Banca Examinadora, Prof. Dr. **Wanderson Roger Azevedo Dias** (IFRR) e, em seguida, o Prof. Dr. **Renê Pereira de Gusmão** (Procc/UFS) e o Prof. Dr. **Glauco de Figueiredo Carneiro** (Procc/UFS). Após as discussões, a Banca Examinadora reuniu-se e considerou o mestrando (a) APROVADO. Atendidas as exigências da Instrução Normativa 05/2019/PROCC, do Regimento Interno do PROCC (Resolução 67/2014/CONEPE), e da Resolução nº 04/2021/CONEPE que regulamentam a Apresentação e Defesa de Dissertação, e nada mais havendo a tratar, a Banca Examinadora elaborou esta Ata que será assinada pelos seus membros e pelo mestrando.

Cidade Universitária “Prof. José Aloísio de Campos”, 22 de agosto de 2023.

**Prof. Dr. Edward David Moreno Ordonez
(PROCC/UFS)
Presidente**

**Prof. Dr. Renê Pereira de Gusmão
(PROCC/UFS)
Examinador Interno**

**Prof. Dr. Wanderson Roger Azevedo Dias
(IFRR)
Examinador Externo**

**Prof. Dr. Glauco de Figueiredo Carneiro
(Procc/UFS)
Coorientador**

**Lucas dos Santos Nunes
Candidato**

Agradecimentos

A Deus por ter me concedido saúde e esforço para superar os obstáculos, inclusive, tranquilidade quando tentava escrever algo em meio a longos estresses, após assistir jogos do Flamengo.

A minha mãe Angela, o meu irmão Maurício, ou seja, toda a minha família. Muitos que fazem parte dela, mesmo morando tão distantes, ainda hoje me dão apoio durante essa caminhada. Até mesmo as minhas sobrinhas Miriã e Mayra, ainda crianças e com pouca idade, me fizeram sorrir em algum instante nos momentos de lazer.

Devo reconhecer o apoio de todos os meus amigos e colegas da UFS que diretamente ou indiretamente fizeram parte da minha formação acadêmica.

A todos os participantes convidados que dispuseram do seu tempo e decidiram contribuir com o projeto.

Ao meu orientador Edward pela atenção, pelo seu jeito descontraído e os seus incentivos que foram dados para este trabalho.

Ao meu coorientador Glauco pelas correções e sugestões, além de ter me ajudado bastante com vários conceitos que aperfeiçoaram o esboço durante a sua elaboração.

Aos avaliadores da banca por aceitarem o convite e darem contribuições para o projeto.

E por fim, ao PROCC, a Elaine e a todos os professores que colaboraram para que eu chegasse até aqui. Todos eles ensinam mais que uma profissão, ou melhor, educam para a vida. Muito obrigado!

Resumo

Contexto: A pandemia por coronavírus recém-identificada, mais tarde denominada como COVID-19, é altamente transmissível e patogênica. Uma preocupação adicional neste contexto refere-se à entrada e disseminação da doença nas prisões brasileiras, cujas condições de encarceramento no país são celas superlotadas e mal ventiladas, o que torna esses ambientes extremamente suscetíveis à rápida disseminação da doença.

Objetivo: Analisar os dados do COVID-19 no sistema prisional a fim de caracterizar seu uso em termos de fontes, finalidade e disponibilidade de dados.

Métodos: Apresentar a elaboração de um mapeamento sistemático, com o intuito de identificar publicações relacionadas ao COVID-19 no sistema prisional que utilizam registros do Departamento de Informática do Sistema Único de Saúde (DATASUS) e como estão sendo tratados. Em seguida, é feita a coleta de informações do COVID-19 e de outras doenças classificadas como cardíacas, respiratórias, infecciosas e mentais no estado da Bahia, Brasil, onde serão processadas e armazenadas. O intuito é converter estes dados patológicos em um painel visual, denominado como *dashboard*, ou seja, uma representação da pandemia por meio de atributos visuais de fácil compreensão, como gráficos, histogramas e mapas geográficos. O painel será desenvolvido com a colaboração de heurísticas de Nielsen e testes de usabilidade com participantes convidados, para que possam auxiliar pesquisadores e organizações interessadas, principalmente na área da saúde e da computação. E por fim, algoritmos de *Machine Learning* (ML) foram usados para prever a propagação dessas doenças, com o objetivo de ajudar os responsáveis a tomarem ações emergentes.

Resultados: Foram revisadas 125 publicações, das quais 29 foram identificadas como relevantes em relação aos objetivos do mapeamento sistemático realizado. Em seguida, foram encontrados mais 8 trabalhos por meio da técnica de *Forward Snowballing* (FS), totalizando 37 estudos. Enquanto aos experimentos com ML, os modelos de Regressão Polinomial obteve as melhores aferições com os dados do DATASUS. Os resultados apontam para tendências e necessidades de pesquisas sobre o assunto, pois a ideia é auxiliar na redução da taxa de mortalidade por COVID-19 e outras doenças. Já as contribuições feitas pelos participantes dos testes vêm obtendo importância para o desenvolvimento do *dashboard*, para assim compreendermos melhor o seu funcionamento, além da possibilidade de aperfeiçoá-lo ainda mais.

Palavras-chave: Sistema prisional, *Machine Learning*, Dados abertos, COVID-19, Mapeamento sistemático.

Abstract

Context: The newly identified coronavirus pandemic, later named COVID-19, is highly transmissible and pathogenic. An additional concern in this context refers to the entry and spread of the disease in Brazilian prisons, whose conditions of incarceration in the country are overcrowded and poorly ventilated cells, which makes these environments extremely susceptible to the rapid spread of the disease.

Objective: To analyze COVID-19 data in the prison system to characterize its use in terms of sources, purpose, and data availability.

Methods: To present the elaboration of a systematic mapping, to identify publications related to COVID-19 in the prison system that use records from the Department of Informatics of the Unified Health System (DATASUS) and how they are being treated. Next, information is collected on COVID-19 and other diseases classified as cardiac, respiratory, infectious, and mental in the state of Bahia, Brazil, where they will be processed and stored. The intention is to convert these pathological data into a visual panel, called a dashboard, that is, a representation of the pandemic through easily understood visual attributes, such as graphs, histograms, and geographic maps. The panel will be developed with the collaboration of Nielsen heuristics and usability tests with invited participants, so that they can help interested researchers and organizations, mainly in the area of health and computing. Finally, Machine Learning (ML) algorithms were used to predict the spread of these diseases, to help those responsible to take emerging actions.

Results: 125 publications were reviewed, of which 29 were identified as relevant to the objectives of the systematic mapping carried out. Then, 8 more studies were found using the Forward Snowballing (FS) technique, totaling 37 studies. As for the experiments with ML, the Polynomial Regression models obtained the best measurements with DATASUS data. The results point to trends and need for research on the subject, as the idea is to help reduce the mortality rate from COVID-19 and other diseases. The contributions made by test participants have become important for the development of the dashboard, so that we can better understand how it works, in addition to the possibility of improving it even further.

Keywords: Prison system, Machine Learning, Open data, COVID-19, Systematic mapping.

Lista de ilustrações

Figura 1 – Fases do processo de seleção do artigo	22
Figura 2 – Evidências da literatura para responder às Questões de Pesquisa (RQ - <i>Research Questions</i>)	23
Figura 3 – Distribuição dos estudos por Análise de Domínio	24
Figura 4 – Fontes de dados	26
Figura 5 – Finalidade dos estudos do COVID-19 no sistema prisional	27
Figura 6 – Permissões de acesso aos dados do COVID-19	29
Figura 7 – Países de origem das publicações selecionadas	31
Figura 8 – O <i>pipeline</i> de visualização de informações, extraído de Chi e Riedl (CHI; RIEDL, 1998)	40
Figura 9 – Diagrama esquemático do processo de visualização	41
Figura 10 – Tabelas com bases de dados extraídas do DATASUS	41
Figura 11 – Todos os municípios da Bahia que obtiveram mortes por COVID-19 em 2020	42
Figura 12 – Todos os municípios da Bahia que obtiveram mortes por CIDs em 2020	43
Figura 13 – Localização geográfica do município	43
Figura 14 – Filtros e gráficos sobre o número de mortes por CIDs na Bahia em 2020	44
Figura 15 – Filtros e gráficos sobre o número de mortes por CIDs para cada município em 2020	44
Figura 16 – Listando as CIDs de cada município	45
Figura 17 – Descrição das CIDs encontradas na Bahia	45
Figura 18 – <i>Word Cloud</i> (Nuvem de Palavras)	60
Figura 19 – O funcionamento do modelo de previsão através de um fluxograma	67
Figura 20 – Regressão Linear Simples e Polinomial	67
Figura 21 – Importando os dados para a Análise de Regressão	69
Figura 22 – Dividindo os dados em conjuntos de treinamento e teste	70
Figura 23 – Formatando as informações das datas de óbito	70
Figura 24 – Instanciando o modelo de Regressão Polinomial	70
Figura 25 – Gráficos gerados dos modelos de Regressão Polinomial	75

Lista de tabelas

Tabela 1 – O objetivo formalizado usando o modelo GQM (<i>Goal Question Metric</i>) (BASILI; WEISS, 1984)	17
Tabela 2 – Usando os critérios PICO para construir as <i>strings</i> de pesquisa	18
Tabela 3 – Selecionando os termos principais	19
Tabela 4 – Selecionando os termos alternativos	19
Tabela 5 – Fontes selecionadas para este mapeamento	20
Tabela 6 – Critérios para Exclusão, Inclusão e Avaliação de Qualidade	21
Tabela 7 – Total de artigos selecionados com critérios aplicados em cada base de dados	22
Tabela 8 – Artigos encontrados através do <i>Forward Snowballing</i> (FS)	23
Tabela 9 – Estudos em cada Área de Domínio	25
Tabela 10 – Fonte de Dados sobre COVID-19 no Sistema Prisional	26
Tabela 11 – Métodos de estudo de dados do COVID-19	28
Tabela 12 – Permissões de acesso aos dados do COVID-19	29
Tabela 13 – Regiões dos estudos selecionados	32
Tabela 14 – Informações dos usuários selecionados	46
Tabela 15 – Testes com os participantes das Turmas 1 e 2	47
Tabela 16 – Quantidade de erros das Turmas 1 e 2	48
Tabela 17 – Grau de dificuldade das tarefas entre as Turmas 1 e 2	49
Tabela 18 – Princípios gerais de usabilidade e fatores associados	55
Tabela 19 – Informações dos usuários selecionados para a avaliação	58
Tabela 20 – Testes com os participantes da avaliação	58
Tabela 21 – Resultados das escalas por média para cada principio de usabilidade	59
Tabela 22 – Resultados das escalas por mediana para cada principio de usabilidade	59
Tabela 23 – Resultados das escalas por intervalo para cada principio de usabilidade	60
Tabela 24 – Resultados das predições do DATASUS (COVID-19)	71
Tabela 25 – Resultados das predições do DATASUS (outras doenças)	72
Tabela 26 – Resultados das predições de <i>Connecticut</i>	72
Tabela 27 – Resultados das predições de <i>The Marshall Project</i>	72
Tabela 28 – Resultados das predições de <i>Connecticut</i> em 2020	72
Tabela 29 – Resultados das predições de <i>The Marshall Project</i> em 2020	72

Lista de abreviaturas e siglas

BOP	<i>Federal Bureau of Prisons</i>
BS	<i>Backward Snowballing</i>
CAD	<i>Computer Aided Diagnostic</i>
CEP	Código de Endereçamento Postal
CID	Classificação Internacional de Doenças
COVID-19	<i>Coronavirus Disease 2019</i>
CNN	<i>Convolutional Neural Network</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DATASUS	Departamento de Informática do Sistema Único de Saúde
EUA	Estados Unidos da América
FBS	<i>Forward Backward Snowballing</i>
FS	<i>Forward Snowballing</i>
GQM	<i>Goal Question Metric</i>
IA	Inteligência Artificial
IDH	Índice de Desenvolvimento Humano
IHC	Interação Humano-Computador
MS	Mapeamento Sistemático
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Square Error</i>
ML	<i>Machine Learning</i>
OMS	Organização Mundial da Saúde
ONU	Organização das Nações Unidas
PICO	<i>Population, Intervention, Comparison and Outcomes</i>
QP	Questão de Pesquisa

QPE	Questão de Pesquisa Específica
R^2	<i>R-squared</i>
RBF	<i>Radial Base Function</i>
RMSE	<i>Root Mean Square Error</i>
RQ	<i>Research Question</i>
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machine</i>
TCLE	Termo de Consentimento Livre e Esclarecido
UFS	Universidade Federal de Sergipe
UNEB	Universidade do Estado da Bahia
UNIFACS	Universidade Salvador
URL	<i>Uniform Resource Locator</i>
USA	<i>The United States of America</i>

Sumário

1	Introdução	12
1.1	Motivações	13
1.1.1	Dados abertos do sistema prisional	13
1.1.2	Painéis visuais	14
1.1.3	Algoritmos de <i>Machine Learning</i>	14
1.2	Objetivos	15
1.3	Metodologia	15
1.4	Estrutura do documento	16
2	Mapeamento Sistemático dos dados do COVID-19 no sistema prisional	17
2.1	Fases do mapeamento	17
2.1.1	Definição do objetivo	17
2.1.2	Planejamento	18
2.1.3	Execução	21
2.2	Resultados e discussões	23
2.2.1	Respondendo as perguntas de pesquisa	24
2.2.2	Países de estudos selecionados	28
2.2.3	Ameaças à validade	31
2.3	Perspectivas e desafios no estudo dos dados do COVID-19 nos sistemas penitenciários	31
2.4	Considerações finais do Mapeamento Sistemático	35
3	Dashboard: visualização e interação dos dados	37
3.1	A importância dos <i>dashboards</i>	37
3.2	Modelo proposto do <i>dashboard</i>	39
3.3	Coleta dos dados do COVID-19 no sistema prisional	39
3.4	Desenvolvimento inicial do <i>dashboard</i>	42
3.5	Plano de testes do protótipo inicial	43
3.5.1	Proposta	43
3.5.2	Seleção de participantes	44
3.5.3	Metodologia	45
3.5.4	Seleção das tarefas	46
3.5.5	Objetivo do moderador	47
3.5.6	Quantidade de erros e o grau de dificuldade na execução das tarefas	48
3.5.7	Considerações finais sobre o plano de testes	48
3.5.8	Análise, discussão das descobertas e sugestões	50

3.5.8.1	Número total de mortes por COVID-19 nos municípios da Bahia	50
3.5.8.2	A CID com mais mortes dentre os municípios da Bahia	50
3.5.8.3	O mês com mais mortes dentre um dos municípios da Bahia e a CID com maior relevância	50
3.5.8.4	O mês com mais mortes dentre um dos municípios da Bahia e o número geral de óbitos	51
3.5.8.5	Segunda CID com mais mortes no estado da Bahia	51
3.6	Considerações finais sobre o capítulo	51
4	Avaliação heurística de usabilidade	53
4.1	Objetivos	53
4.2	Justificativa	54
4.3	Estrutura da avaliação heurística	54
4.4	Aplicação da avaliação heurística	57
4.4.1	Seleção de participantes	57
4.5	Resultados e considerações finais	58
5	Análise dos dados usando algoritmos de <i>Machine Learning</i>	62
5.1	Motivações	63
5.1.1	<i>Google Datasets Search</i>	63
5.1.2	Dados abertos de <i>Connecticut</i>	64
5.1.3	<i>The Marshall Project</i>	64
5.2	Objetivos	65
5.3	Metodologia	66
5.3.1	Análise de Regressão	66
5.3.2	<i>Support Vector Machine (SVM)</i>	68
5.3.3	<i>Random Forest</i>	68
5.3.4	<i>Decision Tree</i>	68
5.3.5	Implementação	69
5.3.5.1	Pré-processamento dos dados	69
5.3.6	Resultados	70
5.3.6.1	Critérios de avaliação	70
5.3.6.2	Predições dos modelos treinados	71
5.4	Considerações finais da análise dos dados usando <i>Machine Learning</i>	73
6	Conclusões	77
6.1	Trabalhos futuros	78
	Referências	80

Anexos	87
ANEXO A Plano de Testes - Lista de perguntas e tarefas	88
ANEXO B Plano de Testes – Termo de Consentimento Livre e Esclarecido (TCLE)	89
ANEXO C Avaliação Heurística – Termo de Consentimento Livre e Esclarecido – TCLE	90
ANEXO D Avaliação Heurística - Lista de perguntas e tarefas	92
ANEXO E <i>Dashboard</i> - Classificação das CIDs	94

1

Introdução

A recente pandemia pelas doenças provocadas pelo coronavírus (COVID-19), que desde dezembro de 2019, quando surgiu em Wuhan, Hubei, China, tem sido chamada de doença infecciosa, infectou uma abundância de pessoas em todo o mundo e causou mais de 6 milhões de mortes até hoje. Os sinais mais comuns dessa doença são tosse seca, dor de garganta e febre. Os sintomas podem progredir para uma forma grave de pneumonia com complicações críticas, incluindo choque séptico, edema pulmonar, síndrome da dificuldade respiratória aguda e falência de múltiplos órgãos (MAO et al., 2020) provocou e ainda continua a estimular vários estudos tanto na área da saúde quanto na coleta, visualização e análise de dados na área de saúde.

Um grande obstáculo no controle da propagação desta doença é a ineficiência e a escassez de exames médicos. A detecção precoce do COVID-19 é importante para interromper a transmissão comunitária da doença e o atendimento ao paciente. Mesmo com a aprovação emergencial de vacinas, o isolamento e a quarentena dos pacientes suspeitos foram as maneiras mais eficazes de impedir a disseminação da doença.

Um caso crítico relaciona-se com as prisões, uma vez que as prisões podem se tornar verdadeiras incubadoras do COVID-19 e outras doenças devido à superlotação e à notória deficiência de ventilação e mecanismos de saúde (FRANCO-PAREDES et al., 2020b). Com o início da pandemia, as preocupações com o contágio e a morte entre encarcerados, funcionários e visitantes do sistema prisional brasileiro tornaram-se eminentes. Considerando também que as autoridades vêm estabelecendo e adotando medidas de prevenção e combate ao coronavírus no sistema penitenciário.

Atualmente, outro grande risco é o manuseio do prontuário do paciente, um dos documentos mais importantes na atenção domiciliar e o que mais recebe contato físico, já que reúne informações detalhadas sobre o paciente e seu tratamento terapêutico e passa diariamente pelas mãos de diversos profissionais da área médica.

Outro ponto crítico refere-se ao fato que foram diagnosticadas implicações cardiovasculares com desfecho fatal para pacientes com doença do coronavírus (COVID-19), aumentando ainda mais a preocupação com esse diagnóstico, uma vez que as informações sobre o impacto das complicações cardiovasculares no desfecho fatal de COVID-19 ainda são escassos (GUO et al., 2020a).

Isso motiva a necessidade dos pesquisadores saberem como os dados do COVID-19 no sistema prisional foram produzidos e utilizados. Embora os estudos que analisam dados tenham amplamente adotado vários tipos de perspectivas e áreas para focar a análise, há uma falta de compreensão adequada de como esses estudos individuais contribuem para todo o campo de dados do sistema prisional. Até onde sabemos, nenhum estudo secundário investigou e está investigando como os dados do COVID-19 relacionados aos sistemas penitenciários estão sendo usados pela comunidade de pesquisa.

1.1 Motivações

1.1.1 Dados abertos do sistema prisional

No artigo (CARNEIRO et al., 2021), os pesquisadores encontraram artigos publicados entre 2000 e 2019 para apresentar uma visão panorâmica de como a literatura tem lidado com questões relacionadas a dados dos sistemas prisionais, isto é, um Mapeamento Sistemático (MS) como primeira tentativa de fechar a lacuna da ausência de estratégias eficazes de ciência de dados entre a situação atual dos dados carcerários e um novo cenário no qual os dados estariam disponíveis a quem possa interessar, sem o risco de violação do segredo legítimo merecido tanto por indivíduos como por organizações.

As evidências fornecidas pelos estudos selecionados também mostraram poucas discussões públicas sobre tecnologia e metodologias para lidar com dados de sistemas penitenciários. O tipo de uso da maioria dos estudos selecionados estava relacionado à análise dos dados coletados muito mais do que ao armazenamento e atualização dos dados. As fontes de dados são, em sua maioria, produzidas *ad-hoc* por cada grupo de pesquisa, o que significa vários esforços para produzir dados apenas para uso privado e episódico.

Como resultado, no trabalho (CARNEIRO et al., 2021), os pesquisadores obtiveram uma visão de que existe potencial para mais estudos secundários na área de sistemas penitenciários com um total de 531 registros, onde 196 artigos foram selecionados na literatura. A grande maioria dos artigos analisados (75%) utilizou dados restritos. Apenas 18 estudos (9%) forneceram dados, o que dificulta as iniciativas de replicação. Isso indicou a necessidade de analisar o sistema prisional de forma integrada, em que a multidisciplinaridade e a transparência são questões relevantes a serem consideradas nesses estudos.

1.1.2 Painéis visuais

No artigo (DOWDING; MERRILL, 2018), os autores descreveram uma abordagem feita para combinar os princípios das 10 heurísticas de usabilidade originais de Nielsen com heurísticas derivadas especificamente para avaliar a usabilidade de sistemas de visualização de informações, para auxiliar na avaliação heurística de uma forma comum de tecnologia usada em cuidados de saúde como painéis clínicos. A abordagem apresentada produziu um conjunto de heurísticas de usabilidade que podem ser usadas para avaliar sistemas semelhantes.

Como metodologia usaram a técnica de grupo nominal para determinar um conjunto final apropriado. Os princípios de usabilidade existentes combinados e fatores associados foram distribuídos por e-mail para um grupo de 12 especialistas em informática de uma variedade de disciplinas de saúde. Os entrevistados foram solicitados a classificar cada fator em sua importância como uma heurística de avaliação para sistemas de visualização em uma escala de 1 (definitivamente não inclui) a 10 (definitivamente inclui). A distribuição dos escores para cada item foi calculada. Uma pontuação média de 8 representou consenso para inclusão na lista de verificação final.

Dentre os resultados, 10 dos 12 especialistas responderam com classificações e comentários por escrito. A lista de verificação final elaborada pelos autores (DOWDING; MERRILL, 2018) consiste em 10 princípios de usabilidade (7 gerais e 3 específicos para visualização de informações) fundamentados por 49 fatores de usabilidade. Três especialistas em informática de enfermagem usaram a lista de verificação para avaliar um painel de sinais vitais desenvolvido para enfermeiras de cuidados domiciliares, usando uma lista de tarefas projetada para explorar toda a funcionalidade do painel. Os especialistas usaram a lista de verificação sem dificuldade e indicaram que ela abrangia todos os principais problemas de usabilidade encontrados durante a conclusão da tarefa.

1.1.3 Algoritmos de *Machine Learning*

Durante a pandemia foram diagnosticadas implicações cardiovasculares com resultados fatais para pacientes com a doença do coronavírus (COVID-19), aumentando ainda mais a preocupação com este diagnóstico. Uma vez que as informações sobre o impacto da complicação cardiovascular no resultado fatal da doença ainda são escassas (GUO et al., 2020b).

Deste modo, é bastante relevante que surjam alternativas econômicas e precisas para a detecção rápida de doenças cardíacas e respiratórias, como a captura de imagens médicas, que inclui raios-X, tomografias e ultrassons, que tradicionalmente exige a flexibilidade do olho humano para detectar anomalias biológicas. Neste contexto, o uso de técnicas de Aprendizado de Máquina, conhecido como *Machine Learning* (ML), provou ser uma ferramenta eficaz para classificar sinais comuns de doenças pulmonares nos últimos anos, como o COVID-19. Isso está mudando o papel do radiologista de hoje, que agora pode tirar proveito de um diagnóstico

auxiliado por computador (CAD - *Computer Aided Diagnostic*) para imagens médicas (COWAN; MACDONALD; FLOYD, 2013).

Além do mais, a ML tem sido amplamente utilizada no diagnóstico automático (ZHANG et al., 2018), na detecção de nódulos do pulmão (LI; LI, 2017) e classificação (SONG et al., 2017). Como exemplo, a inclusão também de Redes Neurais Convolucionais (CNNs - *Convolutional Neural Networks*) (CHITYALA; PUDIPEDDI, 2020), isto é, classes de redes neurais artificiais do tipo *feed-forward*, que vem sendo incluídas com relevância na análise e no processamento de imagens digitais.

1.2 Objetivos

Este trabalho tem como objetivo geral analisar os dados do COVID-19 no sistema prisional, a fim de caracterizar seu uso em termos de fontes, finalidade e disponibilidade de informações. Além disso, também são definidos os seguintes objetivos específicos:

- Realizar um estudo de Mapeamento Sistemático para fornecer uma visão abrangente dos impactos que podem impedir o uso eficaz dos dados do COVID-19, questões críticas e lacunas na pesquisa acadêmica atual;
- Apoiar esses profissionais e especialistas pro meio de uma plataforma *web*, onde é possível ficarem informados sobre a pandemia do COVID-19. Como estudo de caso, foi selecionado o sistema prisional do estado da Bahia, no Brasil;
- Uso dos algoritmos de *Machine Learning* para identificar o estágio inicial e prever a propagação dessas doenças com uma análise dos registros que foram coletados.

1.3 Metodologia

A ideia inicial deste trabalho foi realizar um mapeamento para identificar e classificar os estudos atuais existentes sobre dados abertos do COVID-19, no contexto de sistemas penitenciários. Um MS fornece uma visão geral de uma área de pesquisa para avaliar a quantidade de evidências existentes sobre um tópico de interesse (BA; CHARTERS, 2007).

No mapeamento é apresentado o seu processo, a definição do objetivo e as questões de pesquisa. Também são discutidas as estratégias de busca, a definição de critérios para selecionar os estudos relevantes da revisão e como decidir pela inclusão ou não de um artigo. Em seguida, é tratada a extração e categorização dos dados no MS e os países que mais contribuem com o tema. Logo depois, são debatidas as ameaças à validade da pesquisa e, por fim, são apresentados os resultados da análise (BA; CHARTERS, 2007).

O próximo passo foi elaborar a coleta de dados abertos de casos do COVID-19 e de outras doenças do estado da Bahia, no Brasil, seguido pelo processamento e o armazenamento

das informações. Por meio destes dados, foi possível o desenvolvimento de um painel visual que converte as informações obtidas em indicadores de qualidade (SILVA et al., 2018), como gráficos e mapas. No entanto, para que essa aplicação se torne eficiente e tenha uma ótima interação humana, foi necessária uma avaliação heurística e testes de usabilidade no painel. Deste modo, foram convidados participantes para a realização de testes, com o intuito de avaliar a interface e julgar sua conformidade com princípios de usabilidade, onde foram reconhecidos como heurísticas, além de corrigir possíveis erros ou obter melhorias para o futuro.

Este estudo projeta e analisa uma aplicação com esboços visuais e geográficos representativos de dados da pandemia do COVID-19, no estado da Bahia. Após uma revisão de estudos e técnicas utilizadas na área acadêmica, todo esse esforço poderá servir de exemplo para vários pesquisadores, instituições e partes interessadas que envolvem a área da saúde e da computação, no auxílio ao tratamento e visualização de dados em caso de pandemias e doenças críticas.

Por fim, este trabalho passou por uma avaliação do Comitê de Ética/CONEP, número CAAE:39951720.0.0000.5033, desenvolvido através do projeto de pesquisa "Impacto epidemiológico do COVID-19 no Sistema Prisional na cidade de Salvador, Bahia, Brasil"(fomento pelo edital MCTIC/CNPQ/FNDCT/MS/SCTIE/DECIT – Nº 4031422020-2).

1.4 Estrutura do documento

Para uma melhor compreensão, este documento está dividido em 6 capítulos. No Capítulo 2, foram elaboradas a classificação e síntese dos estudos selecionados por um MS. O Capítulo 3 analisa e interpreta dados do COVID-19 em um painel visual. Já no Capítulo 4 são revelados os detalhes da avaliação heurística feita no *dashboard*. Logo após, no Capítulo 5 são usados algoritmos de ML na análise dos dados do painel. E finalmente, no Capítulo 6, as conclusões e perspectivas deste trabalho são mencionadas.

2

Mapeamento Sistemático dos dados do COVID-19 no sistema prisional

O Mapeamento Sistemático (MS) realizado está de acordo com as diretrizes propostas por Kitchenham e Charters (BA; CHARTERS, 2007). A revisão iniciou-se com a fase de planejamento, que contemplou a definição do objetivo, observada na Tabela 1, a formulação das questões de pesquisa em busca de publicações em bases de dados de artigos amplamente conhecidas na comunidade acadêmica e a definição dos critérios de inclusão e exclusão, seguidos por revisão de estudos.

2.1 Fases do mapeamento

2.1.1 Definição do objetivo

Tabela 1 – O objetivo formalizado usando o modelo GQM (*Goal Question Metric*) (BASILI; WEISS, 1984)

Analisar	dados do COVID-19
para o propósito de	caracterização
com respeito a	domínio, fonte de dados, uso e nível de acesso
do ponto de vista de	partes interessadas e pesquisadores
no contexto de	sistemas prisionais de países em desenvolvimento

2.1.2 Planejamento

A Questão de Pesquisa (QP) deste estudo é "*Como as partes interessadas e os pesquisadores lidaram com os dados do COVID-19 em ambientes correcionais de países em desenvolvimento?*". A questão de pesquisa é compatível com o objetivo estabelecido neste estudo. Seu uso tenta explorar uma incerteza existente em uma área de pesquisa e aponta para a necessidade de investigação deliberada. Uma boa QP ajuda a sustentar uma tese discutível focada e a construção de um argumento lógico (RATAN; ANAND; RATAN, 2019).

Desenvolvemos por meio de quatro perguntas específicas, conforme apresentado a seguir. Questão de Pesquisa Específica 1, representada como QPE1: "*Quais os domínios de destino usam mais os dados do COVID-19 nos sistemas prisionais?*". Questão de Pesquisa Específica 2, representada como QPE2: "*Quais as fontes primárias no estudo de dados do COVID-19 em sistemas prisionais?*". Questão de Pesquisa Específica 3, representada como QPE3: "*Quais são os principais objetivos dos estudos selecionados em relação ao uso de dados do COVID-19?*". Questão de Pesquisa Específica 4, representada como QPE4: "*Quais são as permissões de acesso aos dados do COVID-19 nos estudos selecionados?*".

A relevância da questão de pesquisa justifica-se pela relevância de entender até que ponto as condições de disponibilidade de dados do COVID-19 em ambientes correcionais de países em desenvolvimento afetam sua análise por pesquisadores e partes interessadas. Por isso, foram selecionados os critérios PICO (STONE, 2002) para construir as *strings* de busca na Tabela 2. Aplicou-se essas regras durante o processo seletivo de trabalhos para este estudo secundário.

Tabela 2 – Usando os critérios PICO para construir as *strings* de pesquisa

(P) População	COVID-19 dados em ambientes correcionais de países em desenvolvimento
(I) Intervenção	uso de dados COVID-19 por pesquisadores e partes interessadas
(C) Comparação	não aplicável
(O) Resultados	domínios-alvo nos quais os dados são usados, fontes primárias de dados, objetivos de análise e permissões de acesso de dados COVID-19

São mostrados nas Tabelas 3 e 4 a seguir como construímos a *string* de busca. A Tabela 3 apresenta os principais termos para os objetivos da pesquisa com base nos critérios PICO. Quando pesquisamos estes bancos de dados, estamos limitados a usar todos os termos possíveis. Portanto, antes de iniciar a pesquisa, é importante ter uma pergunta bem construída. E também foram adotados termos alternativos e sinônimos para esses termos principais em inglês. Como por exemplo, o termo *public data* pode ser um termo alternativo para *open data*. Incluímos os termos alternativos na *string* de busca 4. A *string* de pesquisa resultante inclui os termos principais com o booleano "AND", enquanto os termos alternativos unem os termos principais

por meio do booleano "OR". Com esta iniciativa, pretendemos permitir que as cadeias de busca se concentrem em artigos direcionados às questões de pesquisa deste MS.

Pedimos a especialistas em dados do COVID-19 que avaliassem esses termos e o resultado dessa avaliação indicou a adequação desses termos. Vale ressaltar que, antes da execução de toda a análise de pesquisa, selecionamos um conjunto mínimo de estudos pesquisados para avaliar a eficácia da *string* de pesquisa. A eficácia foi considerada aceitável pelos autores e confirmada posteriormente com os resultados apresentados na Tabela 7.

Tabela 3 – Selecionando os termos principais

Critério	Termos principais
(P) População	AND "prison system"AND "open data"AND "COVID-19"
(I) Intervenção	AND "approach"AND "analyze"AND "purpose"AND "availability"AND "type of data"
(C) Comparação	Não aplicável
(O) Resultados	"analysis results"

Tabela 4 – Selecionando os termos alternativos

Termos principais	Termos alternativos
"open data"	"public data"
"prison"	"correctional institution"OR "penitentiary"OR "jail"
"COVID-19"	"coronavirus"

A Tabela 10 mostra as bases de dados (*Scopus, Web of Science, El Compendex, Science-Direct, PubMed Central, Scielo e Oxford University Press*) de onde obtivemos os estudos. Esta tabela também mostra as *strings* de pesquisa das quais recuperamos os estudos. Foram realizadas as buscas em 28 de julho de 2022. Decidimos utilizar essas bases de dados por serem as mais importantes e bastante utilizadas pela comunidade científica em computação e ciências da saúde.

Incluimos o repositório da *Oxford University Press* devido aos seus registros em saúde pública, criminologia, ciências médicas e, claro, a pandemia do COVID-19. Por exemplo, o *British Journal of Criminology*, um local adequado para publicar estudos com foco em questões criminológicas, está disponível neste repositório. Três (MAYCOCK, 2021) (NOVISKY et al., 2021) (SUHOMLINOVA et al., 2021) dos 37 artigos selecionados vieram desta revista. Os outros repositórios são essenciais e relevantes para a comunidade de ciência da computação.

Decidimos utilizar essas bases de dados listadas na Tabela 10 devido à sua representatividade como as mais utilizadas pela comunidade científica em estudos nas áreas de computação e ciências da saúde. Obviamente, outros bancos de dados poderiam ter sido incluídos na pesquisa, mas os bancos de dados que consideramos podem fornecer uma amostra representativa de

estudos realizados sobre o uso de dados do COVID-19 em instituições correcionais de países em desenvolvimento.

Tabela 5 – Fontes selecionadas para este mapeamento

Banco de dados e URL	String de pesquisa aplicada
<i>Scopus</i> < http://www.scopus.com >	("covid" OR "coronavirus") AND ("prison" OR "inmates" OR "incarcerated") AND ("open" OR "public") AND "data"
<i>Web of Science</i> < http://www.isiknowledge.com >	("covid" OR "coronavirus") AND ("prison" OR "inmates" OR "incarcerated" OR "jail") AND ("open" OR "public") AND "data"
<i>El Compendex</i> < http://www.engineeringvillage.com >	("covid" OR "coronavirus") AND ("prison" OR "inmates" OR "incarcerated" OR "jail") AND ("open" OR "public") AND "data"
<i>ScienceDirect</i> < http://www.sciencedirect.com >	("covid" OR "coronavirus") AND ("prison" OR "inmates" OR "incarcerated" OR "jail") AND "open" AND "data"
<i>PubMed Central</i> < https://www.ncbi.nlm.nih.gov/pmc/ >	("covid" OR "coronavirus") AND ("prison" OR "inmates" OR "incarcerated" OR "jail") AND ("open" OR "public") AND "data"
<i>SciELO</i> < https://scielo.org/ >	("covid" OR "coronavirus") AND ("prison" OR "inmates" OR "incarcerated" OR "jail") AND ("open" OR "public") AND "data"
<i>Oxford University Press</i> < https://global.oup.com/?cc=br >	"covid" AND "prison" AND "inmates" AND "jail" AND "open" AND "data"

A Tabela 6 apresenta os critérios de inclusão/exclusão dos estudos. Os critérios de exclusão são independentes (OR). Isso significa que apenas um único critério é suficiente para tirar o papel. Os critérios de inclusão, através do (AND), por outro lado, deve ser atendido para selecionar o estudo candidato. Também adotamos os critérios de qualidade como questões propostas por Dyba e Dingsoyr (DYBÅ; DINGSØYR, 2008). Adaptamos essas questões conforme apresentadas na Tabela 6. Aplicamos essas questões como um exame crítico de todos os outros artigos que atenderam aos critérios (exclusão e inclusão). A resposta deve ser "SIM" para

selecionar continuamente um estudo. Caso contrário, excluimos o estudo. Usou-se os critérios de inclusão, exclusão e qualidade para escolher os artigos conforme apresentado na Tabela 6 a seguir.

Tabela 6 – Critérios para Exclusão, Inclusão e Avaliação de Qualidade

Tipo	ID	Descrição	Conectivo ou Resposta
Exclusão	E1	Publicações enviadas antes de 2020 (desde que o COVID-19 começou no primeiro trimestre desse ano)	OR
Exclusão	E2	Publicações que não enfocaram sobre o tema	OR
Exclusão	E3	Publicações cujos idiomas não eram português ou inglês	OR
Exclusão	E4	Publicações que não forneceram o texto completo	OR
Exclusão	E5	Duplicar publicações	OR
Inclusão	I1	Publicações cujas fontes foram conferências ou periódicos	AND
Inclusão	I2	Publicações cujos termos como COVID-19, Dados Abertos e Sistema Prisional estão no título, resumo ou palavras-chave	AND
Qualidade	Q1	Os objetivos do estudo foram especificados?	SIM/NÃO
Qualidade	Q2	É claro o contexto do estudo? Foi bem definido?	SIM/NÃO
Qualidade	Q3	A forma como os dados são analisados é compatível com os objetivos do estudo?	SIM/NÃO

2.1.3 Execução

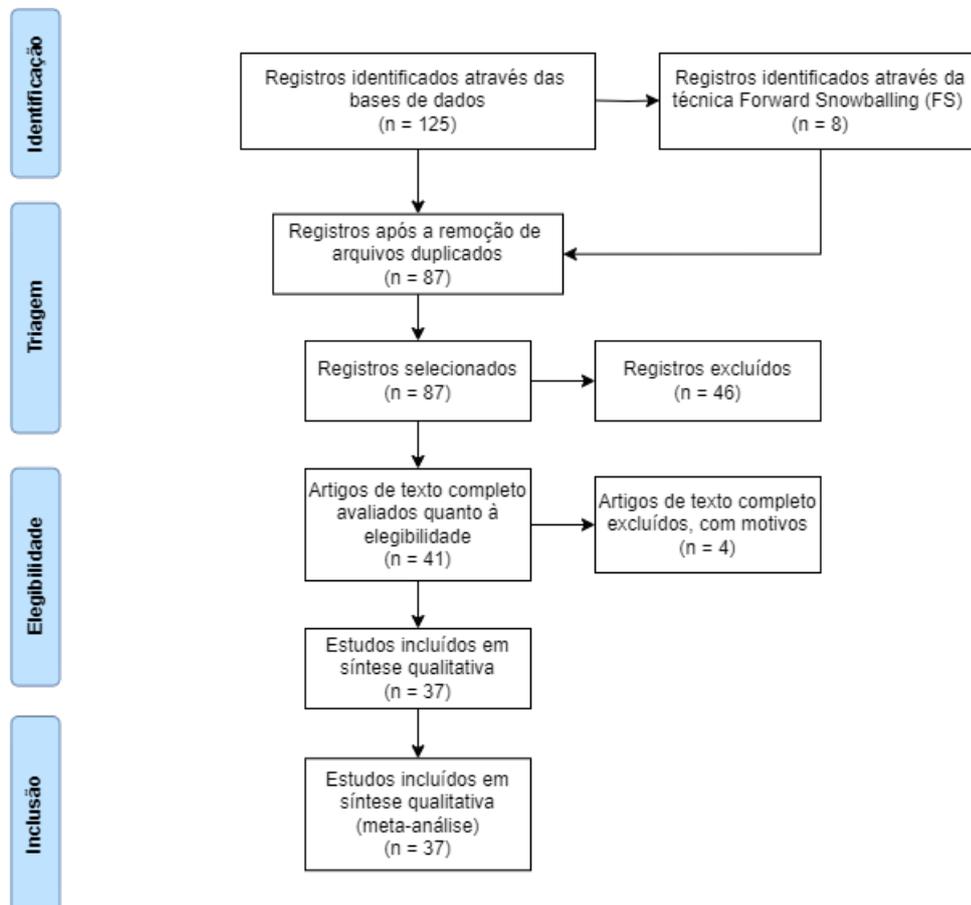
Resumimos as etapas para realizar o processo de seleção na Figura 1. Adotamos o fluxograma PRISMA (MOHER et al., 2009) como referência para o diagrama. Esta figura também apresenta o número de estudos em cada fase desse mapeamento. Foram calculadas a efetividade da busca a partir dos 125 estudos selecionados por MS e apresentados na Tabela 7. Considerou-se um total de 79 estudos após a exclusão de trabalhos duplicados. A *Scopus* contribuiu com 12 artigos, resultando em uma eficácia de busca de 30,7%. Posteriormente, o *PubMed Central* forneceu 8 dos estudos selecionados, correspondendo a uma eficácia de *string* de pesquisa de 42,1%.

A partir desse conjunto inicial de estudos apresentados na Tabela 7, aplicamos a técnica *Forward Snowballing* (FS). Enquanto na busca com *strings*, a maior parte do esforço é dedicada a refinar e adequar os termos às particularidades de cada mecanismo de busca. Por outro lado, a técnica *Snowballing* requer uma parte representativa do esforço para a identificação de referências de destino não encontradas pela *string* de busca original, porém, relacionadas a um tema específico.

Tabela 7 – Total de artigos selecionados com critérios aplicados em cada base de dados

Base	Total	Selecionados	Eficácia de busca
<i>Scopus</i>	39	12	30,7%
<i>Web of Science</i>	35	1	2,8%
<i>El Compendex</i>	1	1	100%
<i>ScienceDirect</i>	8	3	37,5%
<i>PubMed Central</i>	19	8	42,1%
<i>SciELO</i>	3	1	33,4%
<i>Oxford University Press</i>	20	3	15%
<i>Forward Snowballing (FS)</i>	8	8	100%

Figura 1 – Fases do processo de seleção do artigo



A técnica *Snowballing* (WOHLIN, 2014) deve atender a um dos dois critérios a seguir. Um dos critérios refere-se à utilização da lista de referências bibliográficas dos artigos originais selecionados. Isso representa o *Backward Snowballing* (BS). O outro critério refere-se à utilização de artigos que citem os trabalhos originais selecionados. Isso caracteriza o *Forward Backward Snowballing* (FBS).

Ilustra-se na Tabela 8 o resultado da técnica FS em que selecionamos as referências

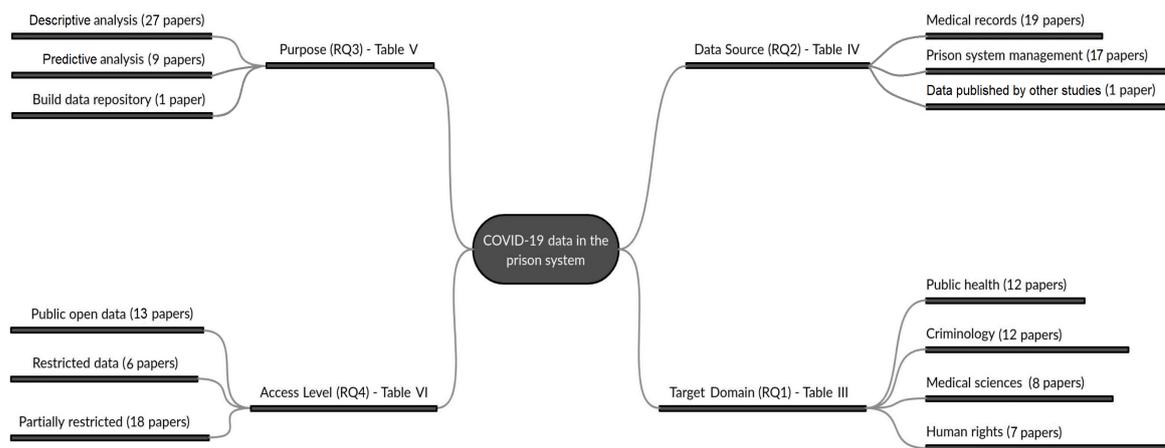


Figura 2 – Evidências da literatura para responder às Questões de Pesquisa (RQ - *Research Questions*)

originais (BYRNE et al., 2021), (CARVALHO; SANTOS; SANTOS, 2020) e (MEKONNEN et al., 2021) para chegar às referências (EZEIBE et al., 2022) e (GALLEGUILLLOS et al., 2022), (RIBEIRO; DINIZ, 2020), além de (MARTIN-HOWARD, 2022), (JUMBE et al., 2022), (MANDERSON et al., 2022), (HOUT; BIGLAND; MARINIELLO, 2022) e (FERDIK; FROGGE; DOGGETT, 2022), respectivamente. Essas novas referências incluídas via FS forneceram evidências para responder às questões de pesquisa declaradas e à discussão correspondente.

Tabela 8 – Artigos encontrados através do *Forward Snowballing* (FS)

Referência original	Artigos encontrados através do FS
(BYRNE et al., 2021)	(EZEIBE et al., 2022; GALLEGUILLLOS et al., 2022)
(CARVALHO; SANTOS; SANTOS, 2020)	(RIBEIRO; DINIZ, 2020)
(MEKONNEN et al., 2021)	(MARTIN-HOWARD, 2022; JUMBE et al., 2022; MANDERSON et al., 2022; HOUT; BIGLAND; MARINIELLO, 2022; FERDIK; FROGGE; DOGGETT, 2022)

2.2 Resultados e discussões

Esta seção apresenta a análise dos resultados obtidos a partir da extração de informações dos 37 estudos relevantes que ajudam a responder às questões de pesquisa deste estudo. Neste estudo, incluímos os países de origem daqueles papéis que contribuíram para a análise do sistema prisional com dados do COVID-19, suas perspectivas e seus desafios, além de possíveis ameaças à validade.

2.2.1 Respondendo as perguntas de pesquisa

Os resultados dos estudos relevantes são apresentados a seguir, relacionando-os com as questões de pesquisa apresentadas na seção anterior.

QPE1) Quais os domínios de destino usam mais os dados do COVID-19 nos sistemas prisionais?

Esta questão visa analisar as áreas identificadas usando os estudos selecionados a partir dos dados do COVID-19. A resposta desta pergunta de pesquisa pode ser visualizada na Figura 2, especificamente no canto inferior direito, onde são encontrados os dados relacionados à Domínio de Destino (*Target Domain*). Organizamos os 37 artigos em quatro assuntos (ver Figura 3): (i) Saúde Pública (*Public Health*) - 12 artigos, (ii) Criminologia (*Criminology*) - 12 artigos, (iii) Ciências Médicas (*Medical Sciences*) - 8 artigos e (iv) Direitos Humanos (*Human Rights*) - 7 artigos. Mostramos as respectivas referências em cada assunto nesta Análise de Domínio na Tabela 9.

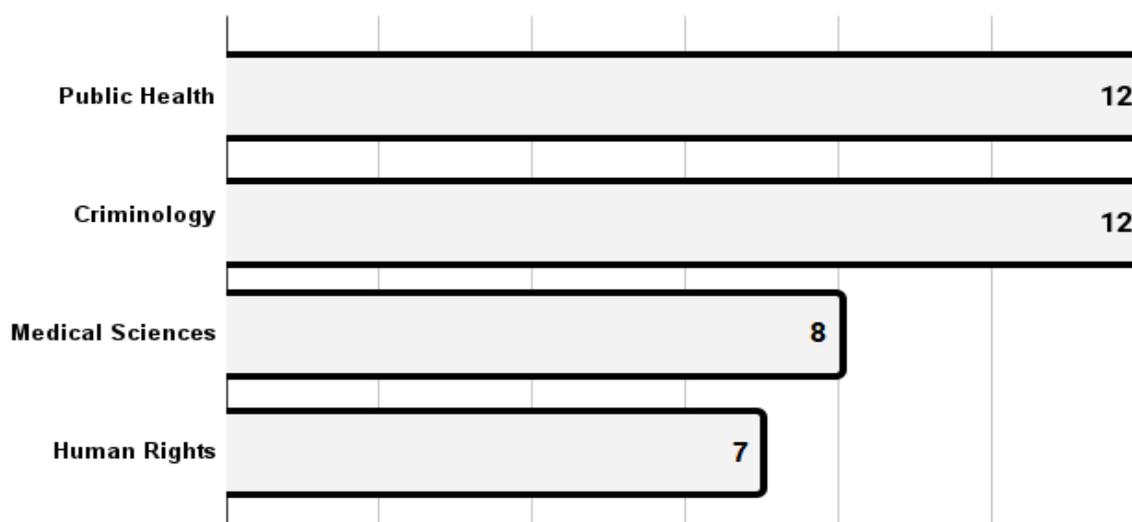


Figura 3 – Distribuição dos estudos por Análise de Domínio

Como pode ser observado na Figura 3, dos quatro domínios mencionados, Saúde Coletiva foi a área mais citada entre as publicações selecionadas, seguida pelas áreas de Criminologia, Ciências Médicas e Direitos Humanos. A relevância das áreas de Saúde Pública e Ciências Médicas pode ser devida a financiamentos e bolsas de estudo quando comparadas a outras. Outra possível razão é que os ambientes prisionais apresentam condições especiais para um conjunto específico de vírus e doenças que chamam a atenção de pesquisadores da área da saúde e da medicina, em especial a pandemia do COVID-19.

O interesse pelas áreas de Criminologia e Direitos Humanos se deve à desigualdade social e à dificuldade de acesso à saúde, segurança e educação de qualidade para a população carcerária. Além disso, vale ressaltar que o crescimento da população carcerária não diminui os

sinais de violência, o que nos leva a crer que as prisões têm sido utilizadas como instituições públicas de incentivo ao crime. Embora vários artigos tenham sido encontrados nas duas áreas, a maioria foi descartada após o processo de condução, devido a textos incompletos ou acesso restrito às publicações. A Tabela 9 identifica as publicações contidas na Figura 3.

Tabela 9 – Estudos em cada Área de Domínio

Áreas	Referências
Saúde Pública (12)	(CARVALHO; SANTOS; SANTOS, 2020; REINHART; CHEN, 2021; LEMASTERS et al., 2020; STRODEL et al., 2021; STUFANO et al., 2021; GRUBESIC et al., 2021; CRISPIM et al., 2021; NOSRATI et al., 2021; WALLACE et al., 2021a; IRVINE et al., 2020; RYCKMAN et al., 2021; NOWOTNY; SEIDE; BRINKLEY-RUBINSTEIN, 2021)
Criminologia (12)	(HAIDUWA; HASHIYANA, 2020; BYRNE et al., 2021; PITTS; INKPEN, 2021; MAYCOCK, 2021; NOVISKY et al., 2021; NOVISKY; NARVEY; SEMENZA, 2021; SUHOMLINOVA et al., 2021; HAGAN et al., 2020; TUDELA, 2021; WALLACE et al., 2021b; GALLEGUILLOS et al., 2022; MARTIN-HOWARD, 2022)
Ciências Médicas (8)	(WARD et al., 2021; GONCALVES et al., 2021; WILBURN et al., 2021; FRANCO-PAREDES et al., 2020a; MEKONNEN et al., 2021; VEST et al., 2020; JUMBE et al., 2022; FERDIK; FROGGE; DOGGETT, 2022)
Direitos Humanos (7)	(MUNTINGH, 2020; EZEIBE et al., 2022; GALLEGUILLOS et al., 2022; RIBEIRO; DINIZ, 2020; MARTIN-HOWARD, 2022; MANDERSON et al., 2022; HOUT; BIGLAND; MARINIELLO, 2022)

QPE2) Quais as fontes primárias no estudo de dados do COVID-19 em sistemas prisionais?

Na Figura 2 traça no canto superior direito os dados relacionados ao ramo Fonte de Dados (*Data Source*) - QPE2. Enquanto na Figura 4 apresenta um gráfico que esclarece os tipos de fontes que dão origem aos dados do COVID-19 no sistema prisional.

A Tabela 10 mostra o resultado destacando os tipos de documentos. Observa-se que, das 37 publicações selecionadas, 19 são de Prontuários Médicos (*Medical Records*), pois uma das dificuldades do sistema prisional hoje é identificar quantos detentos em cada instituição apresentam determinada doença. E uma das soluções foi a adoção do prontuário eletrônico também no serviço de saúde prisional. Por causa da pandemia, eles nos obrigaram a fazer ações muito mais integradas e, ao mesmo tempo, mostrar mais solidariedade e elaborar ações mais articuladas. Os outros 17 são sobre Registros de Gestão de Sistemas Penitenciários (*Prison Systems Management Records*), ou seja, dados em que pesquisadores autorizavam o acesso a

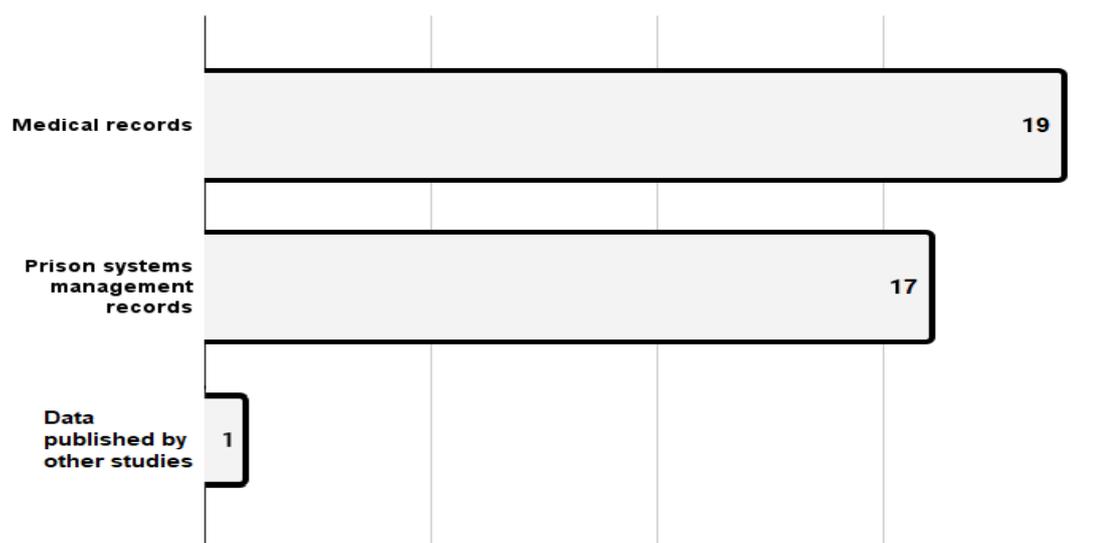


Figura 4 – Fontes de dados

grupos específicos de detentos. Apenas 1 trabalho era referente a Dados Publicados por Outros Estudos (*Data Published by Other Studies*).

Tabela 10 – Fonte de Dados sobre COVID-19 no Sistema Prisional

Tipos	Referências
Registros Médicos (19)	(WARD et al., 2021; GONCALVES et al., 2021; WILBURN et al., 2021; FRANCO-PAREDES et al., 2020a; MEKONNEN et al., 2021; VEST et al., 2020; CARVALHO; SANTOS; SANTOS, 2020; REINHART; CHEN, 2021; LEMASTERS et al., 2020; STRODEL et al., 2021; STUFANO et al., 2021; GRUBESIC et al., 2021; CRISPIM et al., 2021; NOSRATI et al., 2021; WALLACE et al., 2021a; IRVINE et al., 2020; RYCKMAN et al., 2021; NOWOTNY; SEIDE; BRINKLEY-RUBINSTEIN, 2021; FERDIK; FROGGE; DOGGETT, 2022)
Registros de gerenciamento de sistemas prisionais (17)	(MUNTINGH, 2020; HAIDUWA; HASHIYANA, 2020; BYRNE et al., 2021; PITTS; INKPEN, 2021; MAYCOCK, 2021; NOVISKY et al., 2021; NOVISKY; NARVEY; SEMENZA, 2021; SUHOMLIANOVA et al., 2021; HAGAN et al., 2020; TUDELA, 2021; WALLACE et al., 2021b; GALLEGUILLOS et al., 2022; RIBEIRO; DINIZ, 2020; MARTINHOWARD, 2022; JUMBE et al., 2022; MANDERSON et al., 2022; HOUT; BIGLAND; MARINIELLO, 2022)
Dados publicados por outros estudos (1)	(EZEIBE et al., 2022)

QPE3) Quais são os principais objetivos dos estudos selecionados em relação ao uso de dados do COVID-19?

Na Figura 2 é traçado no canto superior esquerdo os dados relativos ao objetivo dos estudos selecionados (*Purpose*). Enquanto na Figura 5, identifica a finalidade dos trabalhos indicados para lidar com os dados do COVID-19 no sistema prisional.

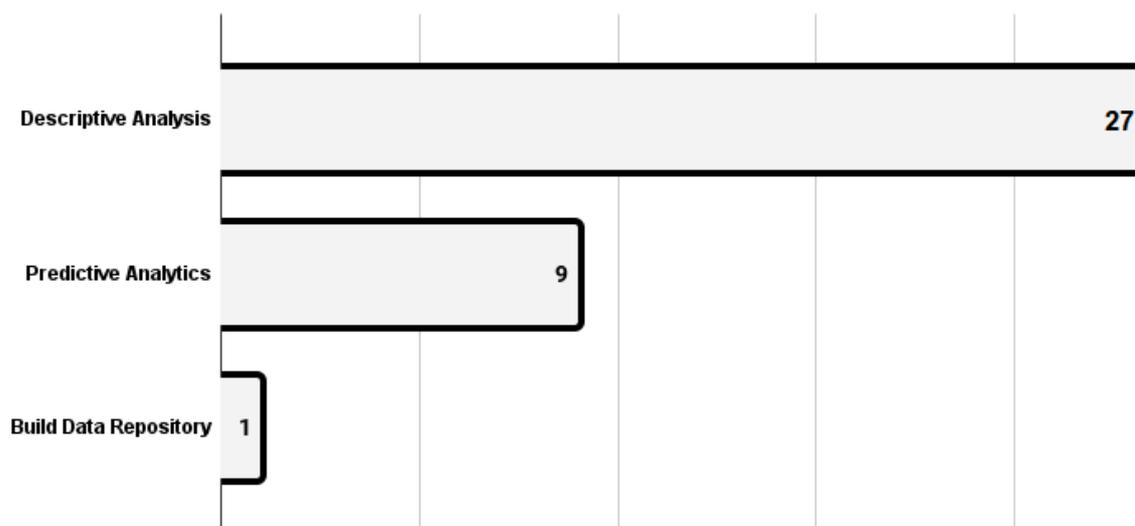


Figura 5 – Finalidade dos estudos do COVID-19 no sistema prisional

As evidências encontradas nos estudos selecionados, também representadas na Tabela 11, mostram a preferência pela análise descritiva (27 estudos). Nesse caso, a maioria dos estudos busca a causalidade em eventos já ocorridos nos sistemas prisionais. Os resultados mostram que apenas 9 estudos focam na Análise Preditiva. Este cenário indica oportunidades para pesquisadores e partes interessadas dedicarem mais esforços ao uso de análises preditivas em vários cenários. Por exemplo, identificar oportunidades de melhoria nas condições de saúde de internos com base no uso de aprendizado de máquina para especificar indivíduos com maior probabilidade de doença para diagnóstico e prognóstico (CALSTER et al., 2019). Gostaríamos de destacar o único caso da África (foi visto que apenas um dos 37 artigos visava coletar informações sobre prisões na Namíbia, na África). A intenção dos autores (HAIDUWA; HASHIYANA, 2020) era desenvolver um aplicativo Android para apoiar o distanciamento social, ajudando residentes a localizar seus parentes e amigos em centros de reabilitação no país africano. Os dados foram coletados por meio de questionários, observações diretas na prisão e entrevistas em grupo, por meio da técnica de amostragem em um público-alvo de 30 participantes.

QPE4) Quais são as permissões de acesso aos dados do COVID-19 utilizadas na realização nos estudos selecionados?

Na Figura 2 traça no canto inferior esquerdo os dados relacionados ao objetivo dos estudos selecionados (*Access Level*). Enquanto na Figura 6, mostramos as permissões de acesso

Tabela 11 – Métodos de estudo de dados do COVID-19

Metodologia	Referências
Análise Descritiva (27)	(MUNTINGH, 2020; WARD et al., 2021; WILBURN et al., 2021; FRANCO-PAREDES et al., 2020a; MEKONNEN et al., 2021; BYRNE et al., 2021; MAYCOCK, 2021; NOVISKY et al., 2021; NOVISKY; NARVEY; SEMENZA, 2021; SUHOMLINOVA et al., 2021; HAGAN et al., 2020; TUDELA, 2021; WALLACE et al., 2021b; CARVALHO; SANTOS; SANTOS, 2020; LEMASTERS et al., 2020; STRODEL et al., 2021; NOSRATI et al., 2021; RYCKMAN et al., 2021; NOWOTNY; SEIDE; BRINKLEY-RUBINSTEIN, 2021; EZEIBE et al., 2022; GALLEGUILLOS et al., 2022; RIBEIRO; DINIZ, 2020; MARTIN-HOWARD, 2022; JUMBE et al., 2022; MANDERSON et al., 2022; HOUT; BIGLAND; MARINIELLO, 2022; FERDIK; FROGGE; DOGGETT, 2022)
Análise Preditiva (9)	(GONCALVES et al., 2021; VEST et al., 2020; PITTS; INKPEN, 2021; REINHART; CHEN, 2021; STUFANO et al., 2021; GRUBESIC et al., 2021; CRISPIM et al., 2021; WALLACE et al., 2021a; IRVINE et al., 2020)
Construir repositório de dados (1)	(HAIDUWA; HASHIYANA, 2020)

aos dados obtidas nas 37 publicações selecionadas.

Como pode ser visto na Tabela 12, a maioria dos estudos enfocou dados Parcialmente Restritos (*Partially Restricted*) - 18 artigos, seguidos de Dados Abertos (*Open Data*) - 13 artigos e Dados Restritos (*Restricted Data*) - 6 artigos do COVID-19 nos sistemas penitenciários. Uma possível explicação é a necessidade de cumprir restrições legais e de privacidade no contexto de dados pessoais, como prontuários de pacientes e registros. Desta forma, os interessados nos dados aplicam as salvaguardas necessárias para evitar o uso indevido da informação.

No entanto, percebe-se que as iniciativas de estudos estimulam o uso de dados abertos, pois oferecem vários benefícios para pesquisadores e profissionais. Também promove condições para melhorar a visibilidade dos resultados pela comunidade e fomenta a reutilização de dados para fins de replicação e novas questões de pesquisa.

2.2.2 Países de estudos selecionados

Na Figura 7 a seguir, apresenta os países nos quais os 37 estudos selecionados realizaram suas pesquisas. Os países desenvolvidos se destacam dos demais por características específicas,

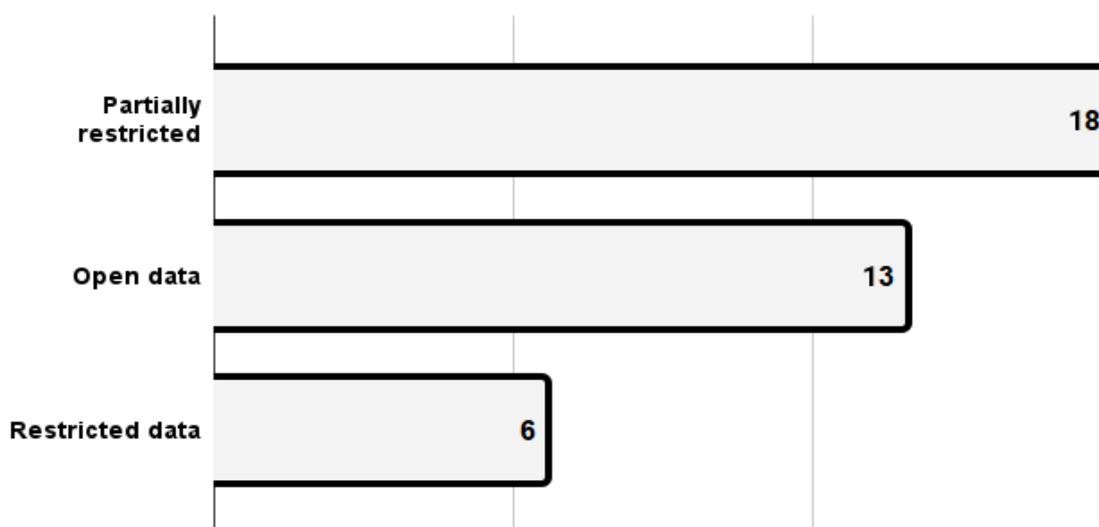


Figura 6 – Permissões de acesso aos dados do COVID-19

Tabela 12 – Permissões de acesso aos dados do COVID-19

Nível de acesso	Referências
Parcialmente restrito (18)	(WARD et al., 2021; GONCALVES et al., 2021; WILBURN et al., 2021; FRANCO-PAREDES et al., 2020a; HAIDUWA; HASHIYANA, 2020; NOVISKY; NARVEY; SEMENZA, 2021; HAGAN et al., 2020; TUDELA, 2021; WALLACE et al., 2021b; CARVALHO; SANTOS; SANTOS, 2020; STRODEL et al., 2021; STUFANO et al., 2021; NOSRATI et al., 2021; IRVINE et al., 2020; RYCKMAN et al., 2021; GALLEGUILLOS et al., 2022; JUMBE et al., 2022; HOUT; BIGLAND; MARINIELLO, 2022)
Dados abertos (13)	(MEKONNEN et al., 2021; VEST et al., 2020; PITTS; INKPEN, 2021; REINHART; CHEN, 2021; LEMASTERS et al., 2020; GRUBESIC et al., 2021; CRISPIM et al., 2021; WALLACE et al., 2021a; NOWOTNY; SEIDE; BRINKLEY-RUBINSTEIN, 2021; EZEIBE et al., 2022; RIBEIRO; DINIZ, 2020; MANDERSON et al., 2022; FERDIK; FROGGE; DOGGETT, 2022)
Dados restritos (6)	(MUNTINGH, 2020; BYRNE et al., 2021; MAYCOCK, 2021; NOVISKY et al., 2021; SUHOMLINOVA et al., 2021; MARTIN-HOWARD, 2022)

como resultados econômicos e sociais. Os parâmetros para classificar um país como desenvolvido ou não se concentram em questões relacionadas à indústria, saúde, qualidade dos produtos e serviços aos cidadãos. A distribuição de renda, por exemplo, mede até que ponto a renda varia entre as populações. Nos países em desenvolvimento, há uma notável disparidade nessa

variação. Reino Unido, Estados Unidos, Suíça, Espanha, Itália, Alemanha e Austrália forneceram evidências de pesquisas realizadas e promoção de estratégias para lidar com o COVID-19 no contexto dos sistemas prisionais.

Os classificados como países em desenvolvimento, geralmente têm um baixo nível de renda per capita, juntamente com um baixo Índice de Desenvolvimento Humano (IDH) como medida de variação de bem-estar. Este termo surgiu após a Segunda Guerra Mundial quando a Organização das Nações Unidas (ONU) publicou uma série de dados sobre a situação socioeconômica de diferentes países (ERENDOR; SUNGURAY, 2019). A partir daí, as diferenças econômicas e de qualidade de vida passaram a ser quantificadas e comparadas. Neste artigo, foram encontrados estudos do Brasil e Chile (América do Sul), Nigéria, Zimbábue, África do Sul e Etiópia (África), Honduras, Guatemala e El Salvador (América Central).

Os dados utilizados nos estudos selecionados vieram da imprensa, governo, autoridades penitenciárias e pesquisas qualitativas com gestores e funcionários penitenciários. Os autores discutiram as consequências do COVID-19 nas prisões e como as partes interessadas lidaram com esse problema. Os autores realizaram estudos baseados em análises descritivas, estatísticas de varredura e análise da evolução dos casos confirmados nos sistemas prisionais. A combinação de análises quantitativas e qualitativas de dados forneceu uma visão abrangente dos impactos da disseminação do vírus.

Os resultados discutidos neste mapeamento são uma primeira tentativa de fechar a lacuna na ausência de estratégias eficazes de ciência de dados entre o estado atual dos dados prisionais e um novo cenário em que os dados do COVID-19 estariam disponíveis para pessoas interessadas, sem o risco de violar o segredo legítimo merecido tanto dos indivíduos como das organizações.

Na subseção anterior, foram apresentadas evidências de estudos que relataram resultados de cenários do COVID-19 no sistema prisional em países desenvolvidos. Por exemplo, existem 19 estudos dos EUA (*USA - The United States of America*), 3 do Reino Unido (*United Kingdom*) e 1 estudo da Espanha (*Spanish*), Itália (*Italy*) e Suíça (*Switzerland*). O conjunto desses 25 estudos de países desenvolvidos mostra uma diferença substancial em termos qualitativos quando comparado aos 12 estudos de países em desenvolvimento. Desses 12 trabalhos, o Brasil (*Brazil*) e a Etiópia (*Ethiopia*) contribuíram com 3 estudos, África do Sul (*South Africa*) somente 2, enquanto Honduras, Guatemala, El Salvador, Namíbia, Zimbábue (*Zimbabwe*), Nigéria e Chile com apenas 1 estudo cada (Tabela 13).

Os resultados mostram uma diferença perceptível, assim como ao número de estudos produzidos em países desenvolvidos, quando comparados aos trabalhos realizados nos países em desenvolvimento. Isso indica a necessidade de apoiar e motivar pesquisadores a dedicarem esforços para analisar, replicar dados e realizar estudos em ambientes prisionais de países emergentes.

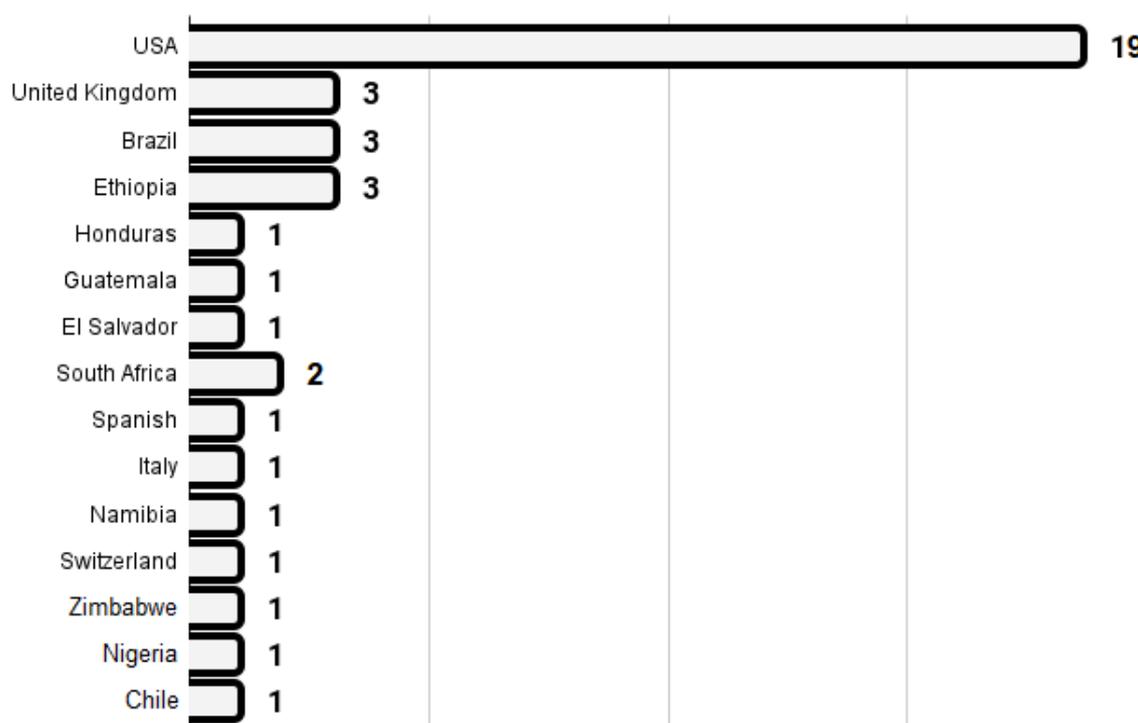


Figura 7 – Países de origem das publicações selecionadas

2.2.3 Ameaças à validade

Seguimos as orientações propostas por Kitchenham e Charters (BA; CHARTERS, 2007) para realizar esse MS e identificar ameaças à validade. A incompletude dos estudos incluídos é uma das ameaças, devido às restrições nas palavras-chave que aplicamos na composição da *string* de busca definida anteriormente. Por esse motivo, foram utilizados sinônimos ou termos semelhantes para construir as sequências de busca e, assim, eliminar estudos não relevantes. Além disso, foi possível eliminar o viés da pesquisa na medida do possível, aderindo às diretrizes e estabelecendo os critérios de seleção e exclusão. Além disso, estudos publicados na literatura cinza não foram incluídos neste mapeamento.

2.3 Perspectivas e desafios no estudo dos dados do COVID-19 nos sistemas penitenciários

Com as respostas fornecidas pelas perguntas de pesquisa específicas da seção anterior e os artigos obtidos pela técnica de *snowballing*, pretendemos apresentar uma visão ampla para apoiar a resposta à principal pergunta de pesquisa. Esta seção destaca desafios e perspectivas na caracterização dos dados do COVID-19 no sistema prisional, em relação à análise de domínio, finalidade dos dados, permissões de acesso e fontes de dados.

Desafios relacionados à Análise de Domínio: O interesse ilustrado na Figura 2 pela prefe-

Tabela 13 – Regiões dos estudos selecionados

Países	Referências
Estados Unidos da América - EUA (19)	(WARD et al., 2021; FRANCO-PAREDES et al., 2020a; VEST et al., 2020; BYRNE et al., 2021; NOVISKY et al., 2021; NOVISKY; NARVEY; SEMENZA, 2021; HAGAN et al., 2020; WALLACE et al., 2021b; REINHART; CHEN, 2021; LEMASTERS et al., 2020; STRODEL et al., 2021; GRUBESIC et al., 2021; NOSRATI et al., 2021; WALLACE et al., 2021a; IRVINE et al., 2020; RYCKMAN et al., 2021; NOWOTNY; SEIDE; BRINKLEY-RUBINSTEIN, 2021; MARTIN-HOWARD, 2022; FERDIK; FROGGE; DOGGETT, 2022)
Reino Unido (3)	(WILBURN et al., 2021; MAYCOCK, 2021; SUHOM-LINOVA et al., 2021)
Brasil (3)	(CARVALHO; SANTOS; SANTOS, 2020; CRISPIM et al., 2021; RIBEIRO; DINIZ, 2020)
Etiópia (3)	(MEKONNEN et al., 2021; JUMBE et al., 2022; MANDERSON et al., 2022)
Honduras (1)	(PITTS; INKPEN, 2021)
Guatemala (1)	(PITTS; INKPEN, 2021)
El Salvador (1)	(PITTS; INKPEN, 2021)
África do Sul (2)	(MUNTINGH, 2020; MANDERSON et al., 2022)
Espanha (1)	(TUDELA, 2021)
Itália (1)	(STUFANO et al., 2021)
Namíbia (1)	(HAIDUWA; HASHIYANA, 2020)
Suíça (1)	(GONCALVES et al., 2021)
Zimbábue (1)	(HOUT; BIGLAND; MARINIELLO, 2022)
Nigéria (1)	(EZEIBE et al., 2022)
Chile (1)	(GALLEGUILLOS et al., 2022)

rência dos estudos selecionados, os temas vão desde saúde, antecedentes médicos e criminais, até dados carcerários e do ponto de vista dos direitos humanos sobre o COVID-19. Como pode ser visto na Figura 3, um dos possíveis motivos para a atenção da comunidade científica à pandemia é a persistência estrutural de sintomas de doenças respiratórias e circulatórias que saem das prisões, com impacto direto na saúde pública. O principal desafio é que essas investigações exigem a participação de funcionários e presidiários para fornecer dados para o estudo.

Perspectivas relacionadas à Análise de Domínio: os dados do COVID-19 podem ser usados em conjunto com dados de saúde pública e criminais e/ou penitenciários. Por exemplo, observar e caracterizar as atividades carcerárias, oferece condições para o aprendizado de

situações estressantes e limitantes. Infelizmente, além do COVID-19, outras doenças não podem ser contidas por paredes.

Assim, existe uma motivação para realizar estudos multidisciplinares combinando diferentes fontes de informação para obter uma compreensão abrangente de uma situação e/ou condição específica. Eles também podem realizar estudos comparativos ou longitudinais de instalações prisionais. Os domínios registrados de análise dos artigos selecionados nos dados do COVID-19, indicam poucos estudos tanto na área da criminologia quanto nos resultados dos estudos de gestão, onde estão relacionados à abordagem de recursos humanos sobre os agentes penitenciários.

Percebe-se que são poucas as iniciativas de vinculação de populações profissionais e de confiança. Ainda que o principal problema na gestão dos sistemas penitenciários seja a relação entre essas duas populações, visto que existem abordagens de estudos penitenciários que consideram presidiários não consideram agentes penitenciários. Presumimos que vasculhar a estrutura de dados da prisão mostrará como agilizar o acesso público a mais dados do que os disponíveis atualmente, dando novas oportunidades para as partes interessadas pesquisarem seus assuntos de estudo com mais facilidade e eficiência.

Desafios relacionados ao acesso aberto aos dados e granularidade dos dados: Em relação às formas de armazenamento dos dados ilustradas na Figura 2, os resultados indicam que a maioria dos pesquisadores opta por manter as informações seguras e não atender a proposta de disponibilizá-los de forma totalmente pública. Quase todos os estudos selecionados exploram dados disponíveis em repositórios de terceiros ou obtêm dados de prontuários médicos e criminais, de acordo com as respostas obtidas na QPE2. Nos resultados obtidos na QPE4, a maioria dos pesquisadores optou por restringir parcialmente os dados do COVID-19 no sistema prisional. A razão provável para esse comportamento, embora a abertura de dados tenha benefícios importantes, é que o compartilhamento de dados ainda traz riscos inerentes à privacidade e à segurança das informações.

Perspectivas relacionadas ao Acesso Aberto aos Dados e Granularidade dos Dados: Embora a maioria dos estudos encontrados não divulgue totalmente os dados, entre os 37 artigos selecionados, 10 utilizaram dados abertos de sistemas prisionais. Os estudos usaram dados com granularidade diferente. Por exemplo, conjuntos de dados contendo campos com nome da cidade, número total de presos e funcionários, sexo, idade, casos confirmados de COVID-19 e número de óbitos, além de outras doenças e detalhes de cada unidade prisional, região ou país. A granularidade apropriada dos dados permite que os pesquisadores realizem análises eficazes e, portanto, alcancem seus objetivos de pesquisa.

Desafios relacionados às Fontes de Dados: As fontes de dados preferidas pelos pesquisadores, em alguns casos, são para fins médicos, onde é desenvolvido um conjunto de amostras biológicas coletadas de presidiários. Enquanto isso, outras informações sobre o COVID-19 são usadas, como estatísticas supranacionais ou nacionais, além dos dados de instalações prisionais,

onde são consideradas oficiais. Inclusive, as unidades prisionais também podem conter registros administrativos para atender aos objetivos da pesquisa.

Por outro lado, os pesquisadores usam dados para rastrear diagnósticos de doenças diferentes do COVID-19. A intenção é planejar amostras representativas de detentos e funcionários. Infelizmente, esses dados não são organizados e estruturados para uso. Mesmo assim, o acesso a essas informações depende de autorização e de muita burocracia. Outro entrave é o formato dos dados, que podem ser digitais ou impressos. Os dados digitais exigem muito menos esforço para serem processados e usados por diferentes partes interessadas do que no formato impresso.

Perspectivas Relacionadas às Fontes de Dados: Disseminar a melhor informação colaborativa dos dados prisionais do COVID-19 é dar às partes interessadas a confiança de que os dados serão usados para fins sociais e médicos. Uma das maneiras de encorajar a colaboração para os processos de produção e uso de dados de doenças seria projetar sistemas abertos e colaborativos. Serve como uma iniciativa transparente e propensa a produzir tanto os dados disponíveis quanto a confiança do público (VASUDEVAN et al., 2020). Por esse motivo, as áreas de imensa preocupação nos estudos penitenciários deve ser a produção, o armazenamento, a acessibilidade e o gerenciamento de dados do COVID-19 e até de outras doenças.

Perspectivas e Desafios sobre Open Data Frameworks: Por outro lado, existem vários benefícios no uso de estruturas de dados abertos, especialmente durante a pandemia do COVID-19. Aqui estão alguns dos principais benefícios: (1) Transparência: as estruturas de dados abertos promovem a transparência na coleta, processamento e análise de dados, permitindo que o público acesse e compreenda os dados que informam as políticas de saúde pública; (2) Colaboração: As estruturas de dados abertos permitem a colaboração entre governos, agências de saúde pública, organizações de pesquisa e outras partes interessadas, promovendo uma resposta mais abrangente e coordenada à pandemia; (3) Tomada de decisão baseada em dados: estruturas de dados abertos fornecem aos tomadores de decisão informações atualizadas e em tempo real sobre o COVID-19, permitindo que tomem decisões informadas sobre alocação de recursos, políticas de saúde pública e respostas de emergência; (4) Acessibilidade: As estruturas de dados abertos tornam os dados do COVID-19 acessíveis a uma ampla gama de usuários, incluindo pesquisadores, formuladores de políticas, profissionais de saúde e o público, promovendo uma abordagem mais inclusiva e participativa para a resposta à pandemia; (5) Inovação: As estruturas de dados abertos permitem o desenvolvimento de novas ferramentas e técnicas de análise de dados, promovendo a inovação e a descoberta na pesquisa em saúde pública; (6) Responsabilidade: Estruturas de dados abertos promovem a responsabilidade no governo e nas agências de saúde pública, permitindo que o público monitore e avalie sua resposta à pandemia; (7) Eficiência: as estruturas de dados abertos promovem a eficiência na alocação de recursos, permitindo que governos e organizações direcionem sua resposta ao COVID-19 de maneira mais eficaz; e (8) Confiança pública:

Por fim, gostaríamos de apresentar algumas lições sobre as estruturas de dados abertos do COVID-19: (i) as estruturas de dados abertos podem ajudar a rastrear e analisar a disseminação

do COVID-19, identificar pontos críticos e monitorar mudanças no impacto do vírus ao longo do tempo; (ii) Padronização da coleta e compartilhamento de dados as práticas são essenciais para estruturas de dados abertos confiáveis e precisas; (iii) A comunicação transparente e clara das fontes, metodologia e limitações dos dados pode gerar confiança entre os usuários e promover a qualidade dos dados; (iv) Governos, agências de saúde pública e organizações de pesquisa devem colaborar para desenvolver e manter estruturas de dados abertos; (v) Medidas de garantia de qualidade, incluindo validação, verificação e limpeza de dados, devem ser implementadas para garantir a exatidão e integridade dos dados; (vi) Medidas de privacidade e segurança devem ser implementadas para proteger dados individuais e informações pessoais; (vii) Atualizações regulares dos dados da COVID-19 devem ser fornecidas para garantir que o público tenha acesso às informações mais atuais e relevantes; (viii) Ferramentas de visualização de dados amigáveis devem ser desenvolvidas para tornar os dados acessíveis e compreensíveis para diferentes grupos de usuários; (ix) Governos e organizações devem priorizar a acessibilidade de dados para populações marginalizadas, incluindo pessoas com deficiência e acesso limitado à internet; (x) Estruturas de dados abertos podem ser usadas para analisar o impacto do COVID-19 em diferentes setores, como saúde, educação e economia; (xi) Acordos e protocolos de compartilhamento de dados devem ser estabelecidos para facilitar o compartilhamento de dados e, ao mesmo tempo, proteger a privacidade individual; (xii) Os governos e as organizações devem oferecer oportunidades de *feedback* e colaboração com os usuários para melhorar o *design* e a funcionalidade das estruturas de dados abertos; (xiii) Estruturas de dados abertos podem ajudar a identificar populações vulneráveis, como idosos, pessoas de baixa renda, e trabalhadores essenciais, que podem estar em maior risco do COVID-19; (xiv) Os governos devem investir em infraestrutura de dados e recursos tecnológicos para apoiar estruturas de dados abertos e (xv) Estruturas de dados abertos podem servir como um recurso valioso para pesquisa, formulação de políticas e tomada de decisões durante a pandemia do COVID-19 e além.

Assim, as estruturas de dados abertos podem ajudar a garantir que o COVID-19 nos sistemas prisionais seja tratada como uma questão de saúde pública, promovendo uma resposta mais holística e abrangente à pandemia. No geral, o uso de estruturas de dados abertos pode fornecer informações valiosas sobre o impacto do COVID-19 nos sistemas prisionais e informar respostas baseadas em evidências à pandemia.

2.4 Considerações finais do Mapeamento Sistemático

Os resultados discutidos neste mapeamento notaram uma ausência de métodos eficientes de ciência de dados entre a situação atual dos dados do COVID-19 no sistema prisional e uma nova perspectiva. No qual as informações estariam disponíveis tanto para pesquisadores quanto para instituições, sem o risco de violação, como a confidencialidade, a disponibilidade ou a integridade dos dados.

As evidências fornecidas pelos estudos selecionados também mostraram poucas discussões públicas sobre tecnologia e métodos para lidar com os dados da doença nos sistemas prisionais. A maioria dos estudos selecionados estavam muito mais relacionados à análise das informações coletadas do que no foco no armazenamento, na visualização e na atualização dos dados.

Nesses estudos, observa-se que a maioria das instalações não implementou atividades de avaliação e gerenciamento de riscos aconselháveis para o COVID-19. Além disso, nenhuma das unidades informou a disponibilidade de exames laboratoriais para casos suspeitos e um plano de contingência para a pandemia, incluindo treinamento de equipes e profissionais para lidar com a prevenção do COVID-19, por meio do uso intensivo de distanciamento físico, uso de banheiros, uso de máscaras e higiene respiratória. Infelizmente, a situação era mais crítica nos países emergentes do que nos países desenvolvidos.

Foram apresentados os resultados para indicar o estado atual da pesquisa, com base em registros do ambiente carcerário e como os pesquisadores conduziram estudos no contexto do COVID-19. Dessa forma, acredita-se que este MS traz uma contribuição relevante para a sociedade, dando suporte a pesquisadores da área de Computação e Saúde, especificamente, em estratégias para dados associados a doença em sistemas prisionais. Pelas consequências apresentadas, nota-se ainda o pouco processamento necessário para a implantação de tais técnicas e seu custo, garantindo que é uma área em ascensão. Mesmo com a ausência de produção, armazenamento, acessibilidade e gerenciamento de dados do COVID-19 e de outras doenças.

Por fim, a maior contribuição deste estudo foi a aprovação da publicação do artigo "A Review of COVID-19 Data Usage in Correctional Institutions from Developing Countries" (NUNES; MORENO; CARNEIRO, 2023) na revista *IEEE Transactions on Computational Social Systems* (Qualis A1), onde contém a estrutura desse mapeamento, identificando e classificando os estudos existentes sobre o uso de dados do COVID-19 em ambientes correcionais, com foco particular nos países em desenvolvimento.

3

Dashboard: visualização e interação dos dados

Na pandemia do COVID-19, os painéis que resumem grandes quantidades de dados se tornaram um tópico de interesse público. Em particular, um painel interativo pode ser útil no monitoramento do surto da pandemia e em manter bilhões de pessoas informadas sobre os desenvolvimentos atuais da doença. Esses painéis visuais, conhecidos como *dashboards*, ajudam a ilustrar figuras-chave em uma tela cheia de dados e oferecem recursos analíticos visuais, funcionais e/ou interativos para lidar com blocos de informações de maneira eficiente.

Da mesma forma, essas ferramentas também podem ajudar a garantir que os tomadores de decisão operacionais nas organizações não sejam sobrecarregados pela ampla quantidade de dados ao tentar usar as informações de forma eficiente. O desafio para as organizações contemporâneas e pesquisas relacionadas é, portanto, tornar os dados coletados mais valiosos para os tomadores de decisão operacionais (SILVA et al., 2018).

3.1 A importância dos *dashboards*

Os *dashboards* e seus recursos subjacentes tornaram-se essenciais para enfrentar este desafio. O objetivo é criar painéis interativos que envolvam os usuários durante a análise dos dados, já que alguns artigos científicos neste fluxo sugerem que os indivíduos devem pesquisar, acumular ou filtrar dados. Esse envolvimento parece apoiar a compreensão dos usuários sobre a natureza complexa dos dados, entretanto, a obtenção desse entendimento envolve o esforço cognitivo e o tempo necessário para processar os dados manualmente, o que pode levar a atrasos nas decisões ou erros.

Hernandez (HERNANDEZ; ROJAS, 2021) descreve um *framework* de desenvolvimento

orientado a modelos para visualização de dados, no contexto de Sistemas de Monitoramento Estrutural. De acordo com os autores, a implantação de um volume de dados na ordem de dezenas de *gigabytes*, desafia os recursos de computação disponíveis para usuários comuns.

Para obter a exibição de informações mais compreensível e precisa, deve ser feito um estudo sobre as técnicas disponíveis de visualização de dados para informações em tempo real. A customização de plataformas existentes e o desenho de quadros específicos, estão entre as tarefas importantes para realizar uma visualização precisa das informações. Toasa (G et al., 2018), realizou uma revisão da literatura sobre visualização de dados, suas técnicas e plataformas de *dashboards* existentes. Foi implementado um painel genérico e dinâmico baseado em informação em tempo real, com o objetivo de avaliar o impacto das técnicas de visualização de dados.

Nesta pesquisa foi realizada uma revisão do estado da arte sobre visualização de dados e identificou um conjunto de técnicas distintas nessa área. Um painel dinâmico para análise de resultados de pesquisas foi desenvolvido usando o *framework Laravel* como base para o desenvolvimento. A proposta foi elaborar um painel visual onde os usuários pudessem interagir com as informações, a partir de um conjunto inicial de dicas, gráficos, tabelas e relatórios, produzidos pelo *dashboard*.

Os painéis podem fornecer um meio único e poderoso para apresentar informações. Por maior que seja a tecnologia, o sucesso de um painel como meio de comunicação é um produto de *design*, resultado de uma aplicação que fala de forma clara e imediata.

Esses sistemas de informação têm desempenhado um papel importante na gestão e controle de doenças epidêmicas, especialmente o COVID-19. Vahedi (VAHEDI et al., 2022) criou um estudo que teve como objetivo identificar os aplicativos, recursos e principais indicadores de *dashboards* avançados na doença. Foi feito um estudo de revisão sistemática que pesquisou em várias bases de dados em 2021, combinando palavras-chave relevantes. Após aplicação dos critérios de inclusão e exclusão e seleção dos artigos, a escolha foi elaborada por meio de um formulário de coleta de dados, onde o estudo foi realizado pelo método de análise de conteúdo.

Considerando os avanços tecnológicos em nível mundial e a enorme quantidade de dados produzidos, uma das soluções eficazes para gerenciar e controlar condições de doenças epidêmicas e pandêmicas é o rápido desenvolvimento de painéis interativos. Assim, sugere-se que os responsáveis de saúde e decisores políticos, além de desenvolverem e atualizarem os *dashboards* existentes no domínio do COVID-19, desenvolvam o painel imediatamente em caso de condições semelhantes.

Os resultados do estudo mostraram que as aplicações mais importantes dos *dashboards* do COVID-19 são seu uso para gerenciamento de leitos hospitalares, para aloca-los a pacientes com prioridade e necessidade de internação, monitoramento da comunidade com base em dados epidemiológicos para decisões macro em nível nacional e regional, revisão do desempenho dos hospitais e do número de serviços prestados para alocação de recursos proporcionais ao seu

desempenho, e acompanhamento de estudos e descobertas científicas recentes nesta área.

Uma das características mais importantes dos *dashboards* estudados foi o foco no nível nacional e a interação com recursos de informação vistos em outros painéis semelhantes. Os principais tópicos de interesse para os *designers* de painéis do COVID-19 incluíram diagnóstico e monitoramento, gestão de recursos, estudos de monitoramento, monitoramento de doenças na comunidade, e informações em vários níveis especializados e gerais.

Considerando os resultados do estudo e o papel dos *dashboards* no rastreamento da prevalência da doença em diferentes comunidades e regiões, sugere-se a utilização mais ampla deste programa no campo do rastreamento da doença, e situações semelhantes de epidemia e pandemia. Dado que o presente estudo examinou todos os aspectos importantes dos painéis do COVID-19, ao longo de um ano de sua prevalência, os resultados do estudo podem ser usados para atualizar os painéis existentes. Isso motiva o desenvolvimento de novos *dashboards* no campo das epidemias e possíveis pandemias no futuro, onde serão usados por formuladores de políticas e *designers* de painéis da área da saúde.

3.2 Modelo proposto do *dashboard*

Chi e Riedl (CHI; RIEDL, 1998) apresentam uma classificação taxonômica de técnicas de visualização que revela, de forma evidente e estruturada, interações de dados para configurar novas visualizações de informações. Diante disso, o *Data State Reference Model* foi desenvolvido em 1998, com o intuito de auxiliar os pesquisadores a visualizar e aprofundar a compreensão do *design* espacial. Da mesma forma que configurar uma estrutura na qual as pessoas que realizam a implementação podem aplicar uma técnica mais eficiente de visualização dos dados.

Mesmo que a modelagem vista na Figura 8 seja do final da década de 1990, é interessante observar como sua configuração serve para elaborar ferramentas e métodos posteriores de visão interativa de dados. Como também, encontram-se hoje aplicações de modelos de visualização de dados em áreas tão diversas, como no desenvolvimento de *frameworks* cognitivos para exibição de informações (JAHNS, 2014). Assim como trabalhos com foco no *design* de painéis visuais de visualização usando tecnologias atuais (exemplificando, um protótipo de medição de sinais vitais) (DOWDING; MERRILL, 2018).

Na Figura 9, o modelo descreve o processo do *dashboard* que vai desde a fonte de dados, as tabelas de dados, o desenho visual em conjunto com aos testes de usabilidade e a representação, num processo de retroalimentação que implica a melhoria destas fases do processo.

3.3 Coleta dos dados do COVID-19 no sistema prisional

O estudo foi feito no site do Departamento de Informática do Sistema Único de Saúde (DATASUS), onde são processadas as informações do Ministério da Saúde do Brasil. Foi lá que

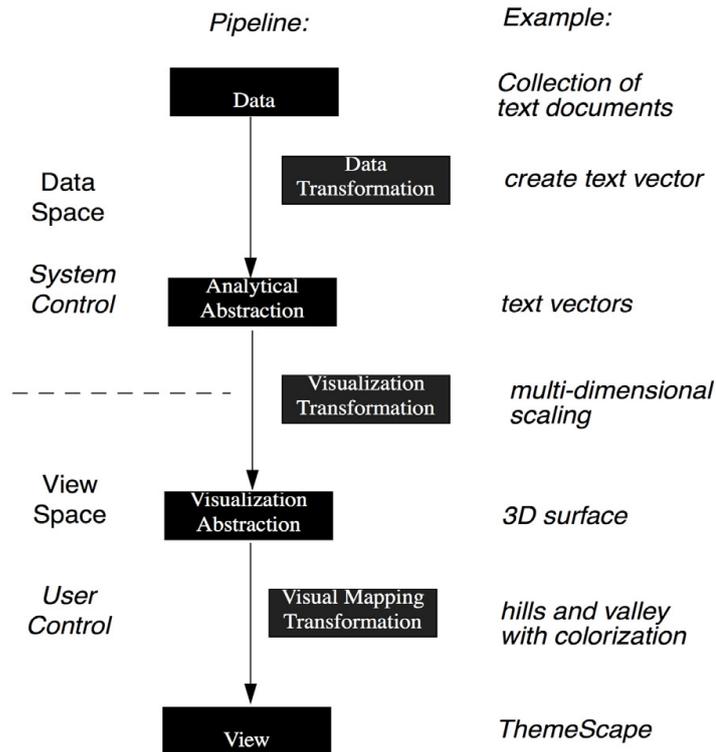


Figura 8 – O *pipeline* de visualização de informações, extraído de Chi e Riedl (CHI; RIEDL, 1998)

obtemos os dados de todos os pacientes que necessitavam de atendimento hospitalar e internados em hospital público, coletando e auditando números referentes a quantidade de procedimentos cirúrgicos, taxas de mortalidade e custos.

De acordo com o diagrama na Figura 9, os dados de casos do COVID-19 são exportados do site do DATASUS e salvos em tabelas de um banco de dados privado. Este processo é feito inicialmente na plataforma TABNET (TABNET/DATASUS, 2022) para ter acesso as informações referentes ao Sistema Único de Saúde (SUS). Apesar de obter uma interface consideravelmente amigável para pessoas leigas, um dos pontos críticos é que os resultados são apresentados frequentemente de forma limitada e fragmentada, sendo difícil saber quais são de fato os dados disponíveis. Também são apontados instabilidades no acesso, sendo um dos motivos para a exportação das bases integrais dos dados.

Logo depois, é usada a plataforma TABWIN (TABWIN/DATASUS, 2022) para descompactar as informações, permitindo dessa forma trabalhar com o gerenciamento de um novo banco de dados, executado em um servidor *web* de hospedagem. Na Figura 10 é exibido o acesso as tabelas construídas com a base de registros do DATASUS, sendo administrados na ferramenta *DBeaver* (DBEAVER..., 2022).

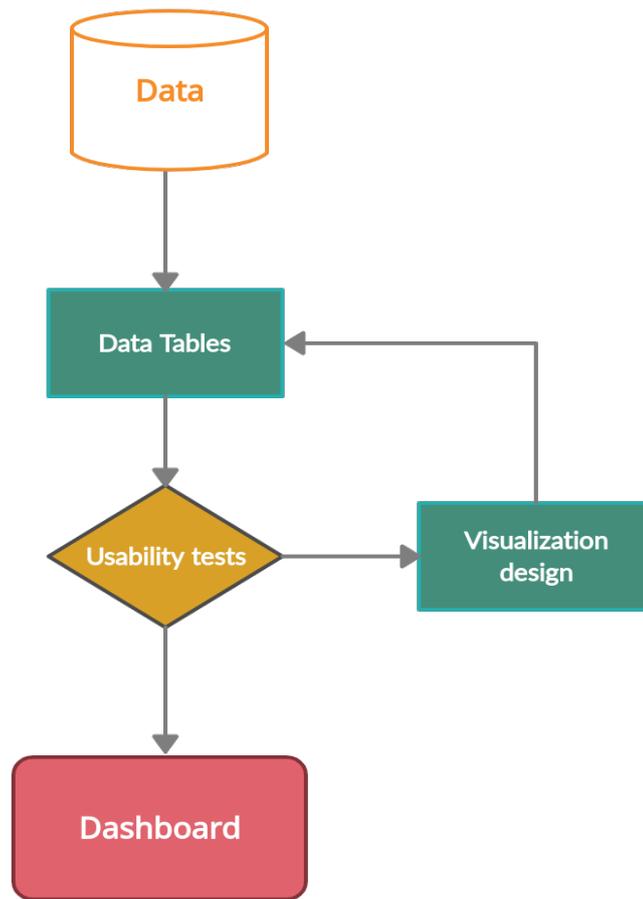


Figura 9 – Diagrama esquemático do processo de visualização

Figura 10 – Tabelas com bases de dados extraídas do DATASUS

DBeaver 21.3.2 - <postgres> Script

SQL: `select * from map_municipios`

	codigo_ibge	sigl nome	latitude	longitude	capital	codigo_uf
1	5.200.050	Abadia de Goiás	-16,757299423	-49,441200256	[]	52
2	3.100.104	Abadia dos Dourados	-18,483100891	-47,391601562	[]	31
3	5.200.100	Abadânia	-16,197000504	-48,705699921	[]	52
4	3.100.203	Abaeté	-19,155099869	-45,444400787	[]	31
5	1.500.107	Abaetetuba	-1,72183001	-48,878799438	[]	15
6	2.300.101	Abaiara	-7,345880032	-39,041599274	[]	23
7	2.900.108	Abaira	-13,248800278	-41,661899567	[]	29

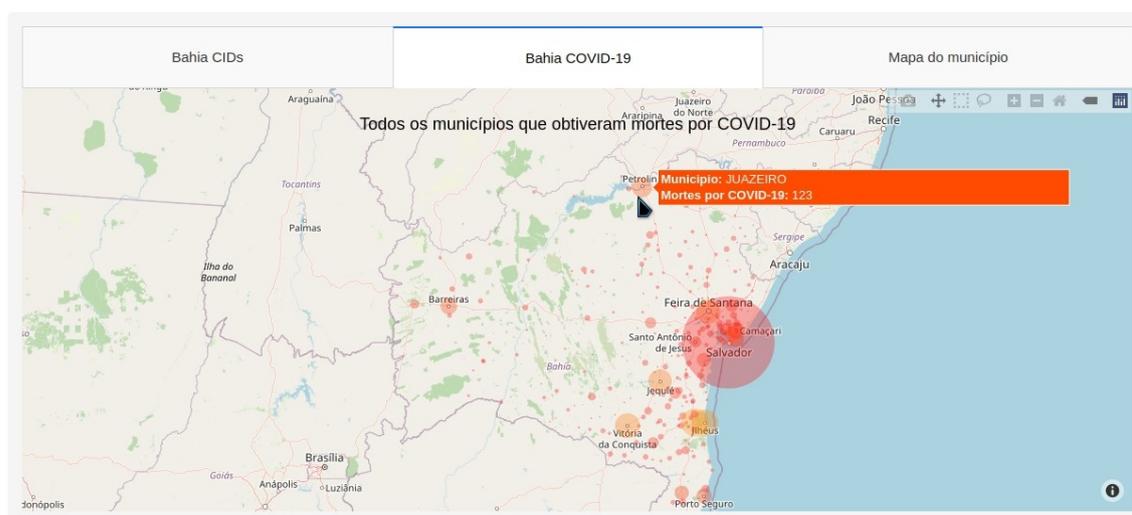
200 row(s) fetched - 238ms (+6ms)

3.4 Desenvolvimento inicial do *dashboard*

Está sendo implementado um protótipo de um painel visual que contém dados do COVID-19 e de outras doenças, além de mapas geográficos por meio de informações exportadas do DATASUS. A doença é classificada como CID (Classificação Internacional de Doenças), criada pela Organização Mundial da Saúde (OMS) em 1893 (ARRAIS et al., 2021). É uma das principais ferramentas epidemiológicas no cotidiano de um médico. A sua vantagem é a diminuição de problemas de tradução e falhas no diagnóstico, facilitando a comunicação entre os diferentes profissionais da saúde do mundo. Além disso, a CID é dividida em 22 capítulos, onde cada um deles reúne grupos de doenças classificadas por letras e números, visto no Anexo E.

Nesse caso, estão sendo usados como testes de desenvolvimento e para uma análise comparativa, os dados de casos confirmados da pandemia do COVID-19 e de outras CIDs nos municípios do estado da Bahia, no ano de 2020. Visto nas Figuras 11, 12, 13, 14, 15, 16 e 17 a seguir.

Figura 11 – Todos os municípios da Bahia que obtiveram mortes por COVID-19 em 2020



As linguagens de programação utilizadas no desenvolvimento do *dashboard* foram o *Python*, o *CSS (Cascading Style Sheets)* e o *HTML (HyperText Markup Language)*, com o auxílio da biblioteca *Plotly* (HASIJA; CHAKRABORTY, 2021). Ela representa os dados complexos em uma forma gráfica usando tabelas e gráficos, além do *framework Flask* (AGGARWAL, 2019) que obtém um *design* modular que permite criar aplicações *web* de forma escalável.

O uso de gráficos como representações visuais para exibir dados, sejam eles, sobre determinada informação, ou valores numéricos de modo geral, são utilizados para demonstrar padrões, tendências e ainda, comparar dados qualitativos e quantitativos num estipulado espaço de tempo. São ferramentas usadas em diversas áreas de estudo matemático, estatístico e geográfico para facilitar a visualização de alguns registros, tal como para tornar os dados mais transparentes e informativos. Dessa forma, a aplicação de gráficos torna a análise mais ágil e objetiva.

Figura 14 – Filtros e gráficos sobre o número de mortes por CIDs na Bahia em 2020

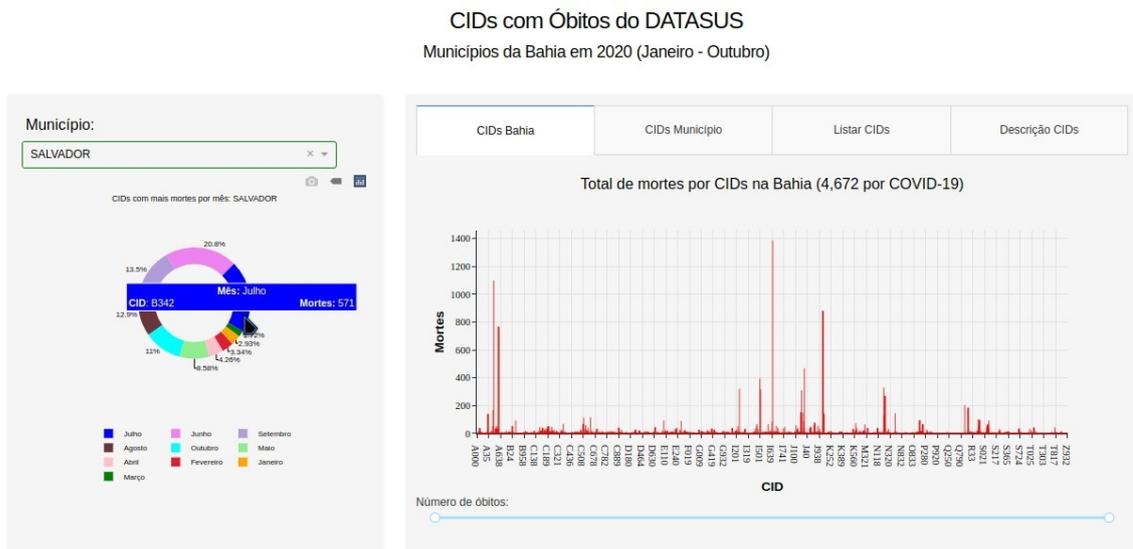
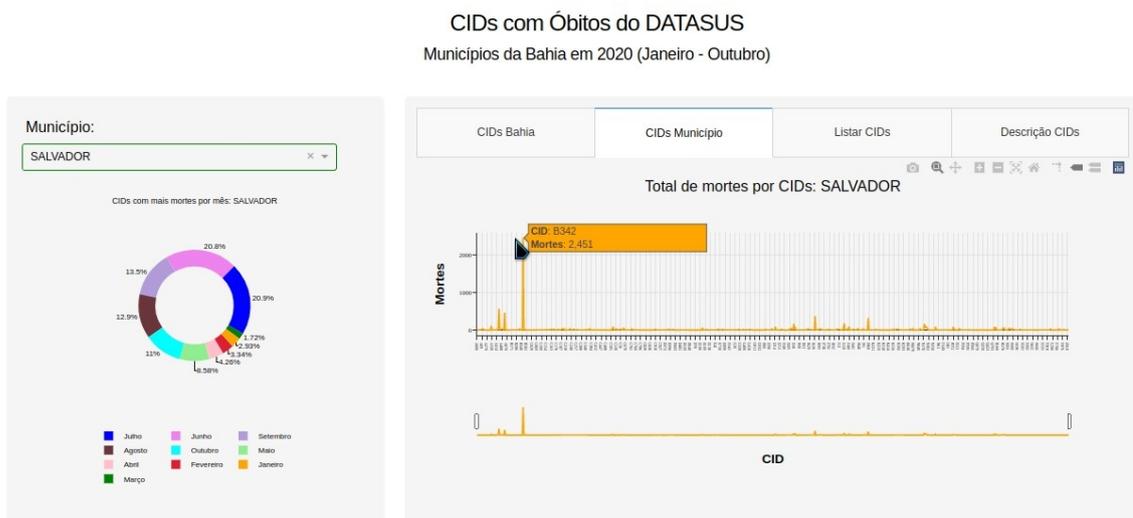


Figura 15 – Filtros e gráficos sobre o número de mortes por CIDs para cada município em 2020



ao levantamento de dados sobre a eficiência, eficácia e satisfação da ferramenta e as demandas relacionadas aos esforços cognitivos. Assim serão realizados ajustes na ferramenta a partir das opiniões, das observações e das dificuldades expressadas pelos mesmos.

Não haveria risco financeiro aos participantes nesse estudo e o sistema estaria disponibilizado gratuitamente pelos pesquisadores. A colaboração dos usuários seria importante para o desenvolvimento do *dashboard*, para assim compreendermos melhor o seu desempenho, além da possibilidade de aperfeiçoá-lo ainda mais.

3.5.2 Seleção de participantes

Foram selecionados alunos da Universidade Federal de Sergipe (UFS), que tinham no mínimo algum conhecimento sobre usabilidade em geral e alguma experiência com *dashboards*.

Figura 16 – Listando as CIDs de cada município

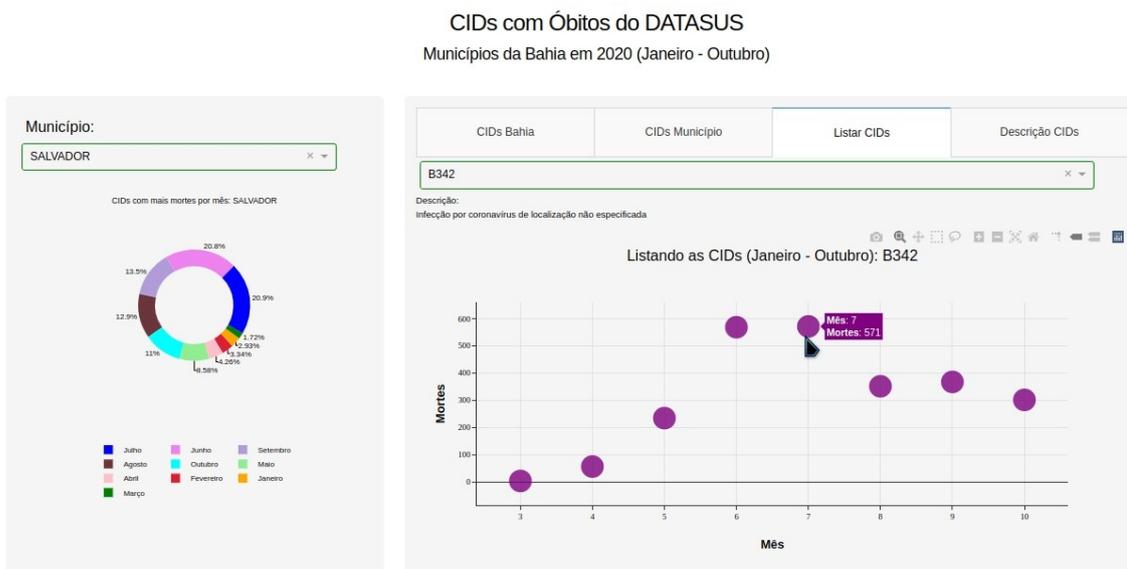
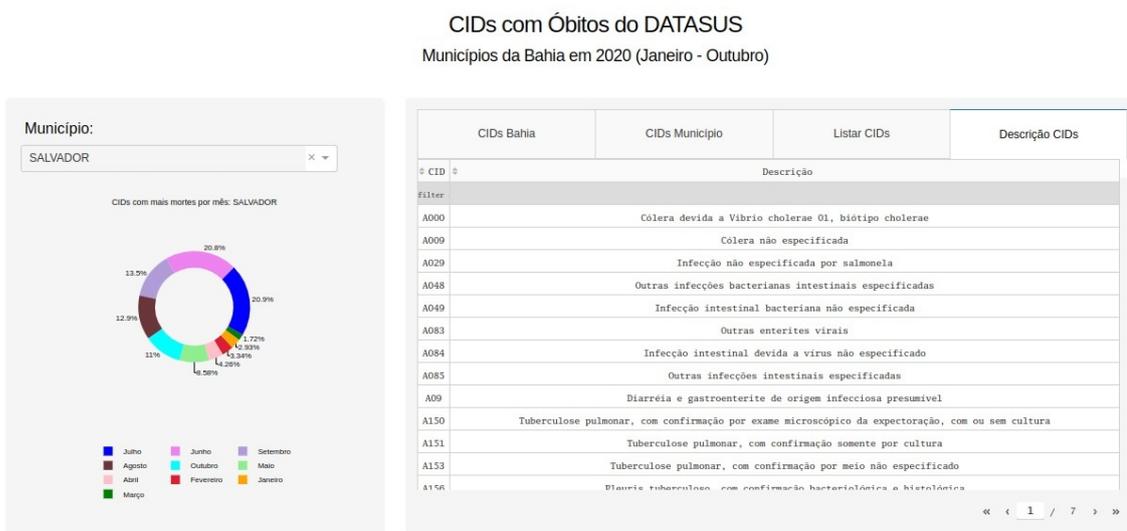


Figura 17 – Descrição das CIDs encontradas na Bahia



Para o teste de usabilidade, foram selecionados 13 participantes com as características descritas na Tabela 14 a seguir.

3.5.3 Metodologia

Para conduzir a sessão de testes deste estudo foi utilizada como ambiente controlado uma sala virtual e o auxílio de 13 participantes, onde foi calculada uma média de tempo aproximada de 27 minutos com cada indivíduo. O teste foi dividido em 2 grupos: uma turma de 5 usuários que fizeram os testes pelo turno da tarde, e uma segunda de turma de 8 participantes que fizeram o teste no turno da noite. Durante esse período, os usuários assinaram a TCLE, que desejavam participar voluntariamente da pesquisa, uma entrevista inicial com informações fornecidas anteriormente na Tabela 14, e durante o tempo restante, os participantes realizaram as tarefas

Tabela 14 – Informações dos usuários selecionados

Características	Número de participantes
Número total de participantes	13
Curso	Graduação: 13
Tempo de experiência em usabilidade em geral	Menos de 1 ano: 9 Mais de 4 anos: 2 De 1 a 2 anos: 1 De 3 a 4 anos: 1
Tempo de experiência em utilizar <i>dashboards</i>	Menos de 1 ano: 10 De 1 a 2 anos: 2 De 3 a 4 anos: 1
Não possuem conhecimentos sobre requisitos de usabilidade aplicados a <i>dashboards</i>	13
Precisaram de algum requisito de acessibilidade digital para o uso de sites	0

propostas no Anexo A e o tempo gasto por cada usuário na Tabela 15 a seguir.

3.5.4 Seleção das tarefas

Com a proposta definida e o método identificado, foi desenvolvida a lista de tarefas a serem realizadas no teste. O desafio está em apresentar essas tarefas aos participantes de forma que o avaliador possa aprender das respostas e consequentes ações dos usuários. No caso de testes de usabilidade, o avaliador deve definir também as medidas a serem observadas para cada postura que se deseja apreciar. Por exemplo, para se avaliar o critério de produtividade, possivelmente será desejável medir o tempo gasto no desempenho de cada tarefa, e o número de erros cometidos por cada questão.

O número de itens a serem desenvolvidas foi limitado a 10 (dez), sendo 5 perguntas e 5 tarefas para que o teste não se tornasse muito cansativo para a avaliação. Teve-se o cuidado de elaborar questões que representassem tarefas reais que os usuários poderiam encontrar na utilização do *dashboard*, entre elas, pesquisas na base de dados utilizando como parâmetros, CID, óbitos, município, período mensal e a localização geográfica. Para reduzir a parcialidade do teste, a seleção de tarefas foi designada para mais de uma pessoa.

O desenvolvimento das questões foi um desafio, pois tem-se que elaborá-las de forma que a interface do sistema fosse testada, e não a habilidade do usuário. Para não dificultar a tentativa de mensurar e identificar os problemas de usabilidade, então, questões com muitas respostas potencialmente “corretas” foram descartadas, dando prioridade às questões com respostas mais objetivas ou de resolução única.

3.5.5 Objetivo do moderador

O intuito do moderador ou administrador do teste é o mais crítico de todos os membros da equipe de teste. Na verdade, o moderador é o membro da equipe que não pode faltar para conduzir o teste. Ele é o responsável pela preparação do material, interagir com os participantes e coordenar da equipe para a realização do teste (RUBIN et al., 2008). Durante a aferição, o avaliador é responsável por todos os aspectos administrativos, incluindo o atendimento aos participantes, coleta dos dados, auxiliando e questionando os participantes. Para a avaliação do teste de usabilidade do painel, foram utilizados dois avaliadores, sendo que somente um, o moderador, interagiria com os participantes. O outro apenas manteria um relatório com o tempo de execução das tarefas. Para manter consistência do teste, os avaliadores, trabalharam com um *checklist* que conteria todos os passos para a condução do teste.

Tabela 15 – Testes com os participantes das Turmas 1 e 2

Turma 1 - Início dos testes	15h33
Participantes	Duração do teste
Usuário 1	15h51
Usuário 2	15h54
Usuário 3	15h57
Usuário 4	16h05
Usuário 5	16h05
Média total	00:25
Turma 2 - Início dos testes	17h25
Participantes	Duração do teste
Usuário 1	17h41
Usuário 2	17h50
Usuário 3	17h50
Usuário 4	17h51
Usuário 5	17h52
Usuário 6	18h03
Usuário 7	18h04
Usuário 8	18h05
Média total	00:29
Média total das Turmas 1 e 2	00:27

3.5.6 Quantidade de erros e o grau de dificuldade na execução das tarefas

Nas Tabelas 16 e 17 mostram o número de erros e o nível de adversidade das tarefas realizadas nos testes de usabilidade com as Turmas 1 e 2.

Teve-se a cautela ao definir o que significaria erro, pois existiam maneiras para se realizar com sucesso as tarefas definidas. No entanto, para deixar a aferição mais objetiva possível, foram considerados incorreções quando em determinadas tarefas mescladas eram cumpridas parcialmente.

Enquanto aos níveis de dificuldade, foram definidos como: muito fácil (grau 1), fácil (grau 2), moderado (grau 3), difícil (grau 4) e muito difícil (grau 5) para a efetuação das tarefas.

Tabela 16 – Quantidade de erros das Turmas 1 e 2

Turma 1					
Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5
Usuário 1	Errou	Acertou	Parcialmente	Parcialmente	Acertou
Usuário 2	Acertou	Acertou	Acertou	Acertou	Acertou
Usuário 3	Acertou	Errou	Parcialmente	Parcialmente	Errou
Usuário 4	Acertou	Errou	Acertou	Acertou	Acertou
Usuário 5	Errou	Errou	Parcialmente	Parcialmente	Errou
Turma 2					
Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5
Usuário 1	Acertou	Acertou	Acertou	Acertou	Errou
Usuário 2	Acertou	Acertou	Parcialmente	Parcialmente	Acertou
Usuário 3	Acertou	Acertou	Parcialmente	Parcialmente	Errou
Usuário 4	Acertou	Acertou	Parcialmente	Parcialmente	Errou
Usuário 5	Acertou	Acertou	Parcialmente	Acertou	Errou
Usuário 6	Acertou	Acertou	Errou	Parcialmente	Errou
Usuário 7	Acertou	Acertou	Acertou	Acertou	Acertou
Usuário 8	Acertou	Acertou	Errou	Errou	Errou

3.5.7 Considerações finais sobre o plano de testes

A média de idade dos participantes do teste de usabilidade eram de 22 anos, onde todos eram alunos de graduação. Nenhum deles possuíam conhecimentos sobre requisitos de usabilidade aplicados a *dashboards*, como também não necessitaram de requisitos de acessibilidade digital para a utilização do painel. Dos 13 participantes, 7 deles abrangiam uma visão do *dashboard* com a função de "monitorar dados em tempo real e representá-los com fácil compreensão

Tabela 17 – Grau de dificuldade das tarefas entre as Turmas 1 e 2

Turma 1					
Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5
Usuário 1	Muito fácil	Fácil	Fácil	Fácil	Moderado
Usuário 2	Fácil	Fácil	Fácil	Fácil	Fácil
Usuário 3	Moderado	Fácil	Difícil	Moderado	Fácil
Usuário 4	Fácil	Muito fácil	Moderado	Muito fácil	Difícil
Usuário 5	Difícil	Fácil	Moderado	Moderado	Muito fácil

Turma 2					
Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5
Usuário 1	Fácil	Fácil	Moderado	Moderado	Muito fácil
Usuário 2	Muito fácil	Fácil	Fácil	Fácil	Difícil
Usuário 3	Moderado	Moderado	Difícil	Moderado	Fácil
Usuário 4	Moderado	Fácil	Fácil	Moderado	Fácil
Usuário 5	Moderado	Difícil	Difícil	Moderado	Muito difícil
Usuário 6	Moderado	Fácil	Moderado	Fácil	Difícil
Usuário 7	Muito fácil	Fácil	Fácil	Muito fácil	Fácil
Usuário 8	Moderado	Fácil	Muito fácil	Muito fácil	Moderado

visual". Enquanto 5 deles traziam como perspectiva "um painel visual de informações que contém métricas e indicadores". Apenas um deles concordava que essas aplicações têm como foco "representar números relevantes para criar estratégias e alcançar os objetivos".

Foi perguntado também aos usuários um levantamento de vantagens e desvantagens de um *dashboard*, onde a maioria descartou qualquer inconveniência com esse tipo de aplicação. Apenas cautelas na hora da conversão de uma enorme quantidade de dados e deixá-los organizados, tornando uma maneira fácil de compreender. Um outro obstáculo detalhado é que certos usuários requerem alguma prática no uso de dispositivos eletrônicos, logo isso acaba excluindo uma parcela da população que não consegue lidar com essas aplicações e sua interatividade.

Enquanto aos problemas relatados pelos participantes durante o teste, eles obtiveram dificuldades em encontrar o número total de mortes de uma determinada CID, como também em buscar a segunda CID com mais mortes no município, principalmente nas Tarefas 3, 4 e 5.

Por fim, a maior dificuldade enfrentada no estudo foi encontrar estudantes dispostos a gastar entre 30 minutos à uma hora de seu tempo para a realização dos testes. Também foi necessário elaborar mais alguns questionários para a identificação do perfil do usuário, como forma de verificar a área que ele cursa e se possuía alguma experiência com painéis visuais.

A vantagem de ter a contribuição dos participantes nesta pesquisa, estão alusivos ao levantamento de dados sobre a efetividade e satisfação da ferramenta. Assim poderão ser feitos novos ajustes na aplicação, de acordo com a perspectiva, das observações e dos obstáculos apresentados por eles.

3.5.8 Análise, discussão das descobertas e sugestões

A análise compreende a descrição dos problemas e possíveis propostas de solução. A partir dos resultados obtidos nos testes de usabilidade nas Turmas 1 e 2, foi feito um estudo para detectar problemas de usabilidade e propor as respectivas mudanças.

3.5.8.1 Número total de mortes por COVID-19 nos municípios da Bahia

Na Tarefa 1, dois participantes da Turma 1 acabaram errando na hora de buscar o número de mortes do COVID-19, nesse caso, filtrando para o município de Jequié. Diferente da Turma 2, nos quais obtiveram 100% de acertos desta tarefa, a maioria dos usuários classificaram que o grau de dificuldade da tarefa foi moderada. Provavelmente devido a dificuldade de encontrar a nomenclatura da doença na aba de descrição das CIDs. O filtro acabava limitando a palavra procurada com o texto da descrição de cada doença. A solução foi a aplicação reconhecer o padrão condizer com alguma parte do texto, mesmo que a palavra estivesse no formato maiúsculo ou minúsculo.

3.5.8.2 A CID com mais mortes dentre os municípios da Bahia

Enquanto na Tarefa 2, dessa vez três participantes da Turma 1 obtiveram erros na hora de buscar a CID com o maior número de mortes, nesse caso, selecionando o município de Feira de Santana. Diferente da Turma 2, nos quais obtiveram 100% de acertos desta tarefa, a maioria dos usuários classificaram que o grau de dificuldade da tarefa foi fácil. Os usuários que erraram, provavelmente se confundiram com o histograma de total de mortes por CIDs, com o gráfico de pizza de CIDs com mais mortes por mês. A solução foi destacar melhor o que cada gráfico traz, adicionando uma descrição em cada aba ilustrando mais detalhes do seu funcionamento.

3.5.8.3 O mês com mais mortes dentre um dos municípios da Bahia e a CID com maior relevância

Na Tarefa 3, apenas dois usuários da Turma 2 não obtiveram êxito na conclusão da atividade. E apesar da maioria dos usuários classificarem que o grau de dificuldade da tarefa foi fácil, ambas as turmas obtiveram erros de forma parcial, devido a quantidade de ofícios em uma única tarefa. Foi notável que a solicitação do número de mortes no município de Ilhéus ficou confusa para a maioria dos participantes, que acabaram confundindo com a quantidade de óbitos de uma única CID, sendo que foi requisitado o número geral de todas as doenças na

região. A solução foi evidenciar o que cada gráfico representa, inserindo uma descrição em cada aba mostrando mais particularidades de sua execução.

3.5.8.4 O mês com mais mortes dentre um dos municípios da Bahia e o número geral de óbitos

Na Tarefa 4, somente um convidado da Turma 2 não conseguiu concluir a atividade. A maioria dos participantes categorizaram que o grau de dificuldade da tarefa foi moderada. Ambas as turmas também obtiveram erros de forma parcial, devido a quantidade de atividades em uma única tarefa. Foi perceptível que a requisição do número de mortes no município de Salvador ficou desconexo para a maioria dos usuários que, por fim, se atrapalharam com a quantidade de mortes, sendo que foi requisitado o número geral de naquela região e não somente da CID com mais óbitos. A solução foi evidenciar o que cada gráfico representa, inserindo uma descrição em cada aba oferecendo mais especificidades do seu funcionamento.

3.5.8.5 Segunda CID com mais mortes no estado da Bahia

Por fim, na Tarefa 5 foram obtidos vários erros dos participantes nas duas turmas durante os testes, sendo que a maioria dos convidados categorizaram o grau de dificuldade da tarefa como fácil. Foi eminente que a maioria se equivocou qual era a CID correta com mais óbitos na Bahia, depois do COVID-19. A melhor saída foi destacar o que cada gráfico representa, adicionando mais detalhes em cada aba, além das suas ações.

3.6 Considerações finais sobre o capítulo

A maior contribuição para decidirmos implementar esse painel visual, foi durante a participação do Projeto de Pesquisa "Impacto epidemiológico do COVID-19 no Sistema Prisional na cidade de Salvador, Bahia, Brasil". Sendo que esse estudo foi aprovado na Chamada "MC-TIC/CNPq/FNDCT/MS/SCTIE/Decit Nº 07/2020 - Pesquisas para enfrentamento da COVID-19, suas consequências e outras síndromes respiratórias agudas graves" do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Este projeto surgiu após uma parceria concreta organizada entre a Universidade do Estado da Bahia (UNEB) e Universidade Salvador (UNIFACS), através do auxílio de professores, doutores, pesquisadores das duas instituições que obtiveram interesse comum e condizente no que refere-se ao tema COVID-19 em panorama multidisciplinar. A pesquisa tem como objetivo, descrever o impacto da pandemia nos serviços de atenção à saúde no sistema prisional de Salvador. A sua execução se iniciou em março de 2021, concluindo-se em julho de 2022, onde nos adequamos as atividades realizadas através de relatórios e reuniões com a equipe de pesquisa.

Além de implementar um painel visual que contém informações do COVID-19 e mapas geográficos em conjunto com os dados do DATASUS, durante o projeto foi desenvolvido também

uma aplicação *web* de um prontuário médico, onde reuniu dados detalhados sobre o encarcerado e o seu tratamento terapêutico. A aplicação serviu como cadastro de informações dos pacientes, incluindo uma opção de consulta dos dados inseridos pelos usuários autenticados (integrantes da equipe). O usuário poderia também emitir um arquivo CSV (*Comma-Separated Values*) com dados totais, sendo ele o curador dos dados.

Seguindo a mesma ideia do *dashboard*, o prontuário foi desenvolvido através das linguagens *Python*, *CSS* e *HTML*, com o auxílio do *Flask*, e o gerenciador de banco de dados (*DBeaver*) para ter acesso as informações hospedadas em um servidor privado.

Conforme os resultados obtidos, pôde-se notar que a pesquisa serviu para colaborar de alguma forma a automação dos dados limitados e fragmentados do COVID-19 obtidas pelo DATASUS, por meio de aplicações que contém pouco processamento necessário para a implantação de técnicas de programação e o seu baixo custo, garantindo que se trate de uma área bastante promissora.

4

Avaliação heurística de usabilidade

Os programas oferecidos pelo computador estão cada vez mais notáveis no cotidiano das pessoas, além de estimular e favorecer a ação de tarefas originalmente manuais. Entretanto, de acordo com estudos, as interfaces desses programas são projetadas de modo incorreto, tornando a execução de uma tarefa mais difícil e de maior tempo (MEDEIROS; CYBIS, 2000).

A usabilidade permite avaliar a funcionalidade de um sistema. Além dos testes, ela pode ser aplicada por meio de uma avaliação heurística, que determina quais recursos do programa não foram projetados de maneira efetiva aos usuários. Essa análise é elaborada normalmente por especialistas da área de Interação Humano-Computador (IHC), onde investiga o *design*, avalia e cria interfaces para que usuários finais possam interagir com sistemas computacionais de uma forma eficiente e intuitiva (HACK; SANTOS, 2010).

Esse tipo de avaliação é uma forma de identificar obstáculos no uso de nossas interfaces e propor soluções. É um procedimento que pode ser tanto uma técnica deliberada de resolução de problemas quanto uma ação de comportamentos automáticos, intuitivos e inconscientes.

4.1 Objetivos

Após os testes de usabilidade, o intuito desse trabalho é também executar uma avaliação heurística, ou seja, uma adaptação no conjunto de heurísticas originalmente proposto por Nielsen (NIELSEN, 1995), para o contexto de *dashboards*. Essa avaliação tem como finalidade identificar problemas na interface de um software, e por meio de seus resultados será reforçado a relevância de avaliar a usabilidade de um sistema.

Os objetivos específicos são:

1. Adaptar as heurísticas propostas por Nielsen para a área de painéis visuais;
2. Aplicar as heurísticas estabelecidas a um software da área por meio de uma avaliação heurística, após os testes de usabilidade;
3. Analisar os problemas de usabilidade encontrados por meio da avaliação heurística e do teste de usabilidade;
4. Comparar os resultados da análise e definir a importância de executar uma avaliação de usabilidade em programas da área, bem como os obstáculos que devem ser melhorados no sistema avaliado.

4.2 Justificativa

A usabilidade de um sistema atua na qualidade notada pelos seus usuários durante sua interação com o *dashboard*. A avaliação da usabilidade é uma ferramenta que permite definir erros na interface. Os instrumentos de avaliação propostos na dissertação são usados por diferentes tipos de avaliadores. Enquanto o teste de usabilidade deve ser feita por usuários finais, a avaliação heurística deve ser ocasionada por especialistas. A execução dessas avaliações devem ser capazes de indicar diferentes tipos de erros e, ao mesmo tempo, indiretamente aferir a qualidade de um painel visual.

4.3 Estrutura da avaliação heurística

A ideia é combinar os princípios das 10 heurísticas de usabilidade originais de Nielsen (NIELSEN, 1995), com heurísticas derivadas especificamente para avaliar a utilização de sistemas de visualização de informação. Além disso, serão realizadas sete etapas no processo de *design* de visualização de dados (JAHNS, 2014). Na maioria dos casos, essas etapas fazem parte de uma metodologia de *design* em espiral, com várias iterações. As etapas são: 1) uma descrição de tarefa cognitiva de alto nível, 2) um inventário de dados, 3) análise de requisitos de tarefa cognitiva, 4) a identificação de tipos de visualização, 5) a identificação e escolha de métodos de interação cognitivamente eficientes, 6) prototipagem e aplicação e 7) avaliação.

Para este estudo, haverá uma lista de princípios heurísticos gerais e fatores de usabilidade específicos derivados dos 10 princípios de usabilidade de Nielsen (NIELSEN, 1995), visto na Tabela 18. Serão usados 7 princípios heurísticos usados para avaliar o painel (visibilidade do status do sistema, correspondência entre o sistema e o mundo real, reconhecimento em vez de lembrança, controle e liberdade do usuário, consistência e padrões, flexibilidade e eficiência de uso, *design* estético e minimalista) em conjunto de fatores de usabilidade específicos.

Tabela 18 – Princípios gerais de usabilidade e fatores associados

Princípio de usabilidade	Fatores de usabilidade
<p>Visibilidade do status do sistema: O sistema deve sempre manter o usuário informado sobre o que é acontecendo por meio de <i>feedback</i> apropriado dentro de um tempo razoável.</p>	<p>Cada tela tem um título ou cabeçalho que descreve seu conteúdo?</p> <p>Existe um esquema de <i>design</i> de ícone consistente e estilístico tratamento em todo o sistema?</p> <p>Há um <i>feedback</i> visual nos menus ou caixas de diálogo sobre quais opções são selecionáveis?</p> <p>A terminologia de nomeação do menu é consistente com o domínio de tarefa dos usuários?</p> <p>O sistema oferece visibilidade: ou seja, olhando, o usuário pode informar o estado do sistema e as alternativas para ação?</p>
<p>Correspondência entre o sistema e o mundo real: O sistema deve falar a linguagem do usuário, com palavras, frases e conceitos familiares ao usuário, em vez de termos orientados ao sistema. Siga as convenções do mundo real, fazendo com que as informações apareçam em uma ordem natural e lógica.</p>	<p>Os ícones são concretos e familiares?</p> <p>Os cabeçalhos e subseções de cada tela estão ordenados da maneira mais lógica?</p> <p>Existe uma sequência natural para as opções de menu para um item de dados?</p> <p>As cores selecionadas correspondem às expectativas comuns sobre códigos de cores?</p> <p>As palavras/conceitos e frases usadas em cada tela são familiares aos usuários?</p>

<p>Controle e liberdade do usuário: Os usuários devem ser livres para selecionar e sequenciar tarefas (quando apropriado), em vez de deixar que o sistema faça isso por eles. Os usuários precisarão de uma “saída de emergência” claramente marcada para sair do estado indesejado sem ter que passar por um diálogo prolongado. Os usuários devem tomar suas próprias decisões sobre os custos de sair do trabalho atual.</p>	<p>Existe uma saída clara em cada tela?</p> <p>Todas as telas estão acessíveis no sistema?</p> <p>Os usuários têm a opção de clicar nos itens de menu com o mouse ou usar um atalho de teclado?</p> <p>Os usuários podem avançar e retroceder facilmente entre as telas?</p>
<p>Consistência e padrões: Os usuários não devem se perguntar se palavras, situações ou ações diferentes significam a mesma coisa.</p>	<p>Os padrões de formatação foram seguidos de forma consistente em todas as telas do sistema?</p> <p>Cada janela tem um título?</p> <p>Os nomes são consistentes, tanto dentro de cada guia quanto em todo o sistema, em estilo gramatical e terminologia?</p>
<p>Reconhecimento em vez de lembrança: Torne objetos, ações e opções visíveis. O usuário não deve ter que lembrar informações de uma parte do diálogo para outra. As instruções para o uso do sistema devem ser visível ou facilmente recuperável sempre que apropriado.</p>	<p>Os avisos, as pistas e as mensagens visíveis são encontradas com facilidade na tela?</p> <p>Os itens foram agrupados em zonas lógicas e os cabeçalhos foram usados para distinguir entre as zonas?</p> <p>O realce de cor é usado para chamar a atenção do usuário?</p> <p>O código de cores é consistente em todo o sistema?</p>
<p>Flexibilidade e eficiência de uso: O sistema deve oferecer aos usuários várias opções na hora de encontrar conteúdo. Os usuários devem ser capazes de atingir seus objetivos de maneira eficiente.</p>	<p>A navegação entre as telas é simples e visível?</p> <p>Se as listas de menu forem curtas, os usuários podem selecionar um item movendo o cursor?</p> <p>Nos menus, os usuários têm a opção de clicar diretamente em um item do menu ou usar um atalho do teclado?</p>

<p>Design estético e minimalista: Os diálogos não devem conter informações irrelevantes ou raramente necessárias. Cada unidade extra de informação em um diálogo compete com as unidades de informação relevantes e diminui sua visibilidade relativa.</p>	<p>Os objetos grandes, as fontes em negrito e as áreas simples foram usadas para distinguir seções?</p> <p>Os rótulos dos campos são breves, familiares e descritivos?</p> <p>O <i>layout</i> visual está bem desenhado?</p>
---	--

4.4 Aplicação da avaliação heurística

Foram convidados participantes para classificar cada fator em sua importância como uma heurística de avaliação para sistemas de visualização em uma escala de 1 (muito insatisfeito) a 10 (muito satisfeito). A distribuição das pontuações para cada item, incluindo médias, medianas e intervalos foram calculadas.

Inicialmente, cada avaliador, separadamente, navegou na interface do programa até buscar um problema de usabilidade. Após isso, o problema encontrado foi classificado de acordo com as heurísticas propostas. Também foram anotados o contexto, a causa, o efeito sobre o usuário e sobre a tarefa, e uma possível correção do problema. Logo, esses dados facilitam a análise de falhas e obtém meios de corrigi-las.

Foi necessário aplicar métodos de avaliação para coletar dados, a fim de identificar a aplicação da ferramenta pelos usuários. Além disso, os convidados responderam um questionário e participaram de uma entrevista. Para a aplicação desses métodos, houve um esclarecimento de cada participante sobre os objetivos e procedimentos da pesquisa, manifestando sua concordância por meio da assinatura de um termo (TCLE), no Anexo C.

4.4.1 Seleção de participantes

Foram selecionados alunos da Universidade Federal de Sergipe (UFS), que tinham no mínimo algum conhecimento sobre usabilidade em geral e alguma experiência com *dashboards*. Para a avaliação foram selecionados 10 participantes com as características descritas na Tabela 19 a seguir.

Durante esse período, os usuários assinaram a TCLE, que desejavam participar voluntariamente da pesquisa, uma entrevista inicial com informações fornecidas anteriormente na Tabela 19, e durante o tempo restante, os participantes realizaram as tarefas propostas no Anexo D e o tempo gasto por cada usuário na Tabela 20 a seguir.

Tabela 19 – Informações dos usuários selecionados para a avaliação

Características	Número de participantes
Número total de participantes	10
Curso	Graduação: 10
Tempo de experiência em usabilidade em geral	Menos de 1 ano: 3 Mais de 4 anos: 2 De 1 a 2 anos: 4 De 3 a 4 anos: 1
Tempo de experiência em utilizar <i>dashboards</i>	Menos de 1 ano: 5 De 1 a 2 anos: 2 De 3 a 4 anos: 3
Não possuem conhecimentos sobre requisitos de usabilidade aplicados a <i>dashboards</i>	9
Precisaram de algum requisito de acessibilidade digital para o uso de sites	0

Tabela 20 – Testes com os participantes da avaliação

Participantes	Duração do teste
Usuário 1	04:00
Usuário 2	12:00
Usuário 3	14:00
Usuário 4	22:00
Usuário 5	11:00
Usuário 6	07:00
Usuário 7	10:00
Usuário 8	05:00
Usuário 9	02:00
Usuário 10	10:00
Média total	09:42

4.5 Resultados e considerações finais

Nesta seção são apresentados os resultados das pontuações geradas por meio da avaliação heurística. Para cada princípio de usabilidade, foram adicionados como tarefas os fatores indicados na Tabela 18. Cada participante selecionou uma escala de 1 a 10 para todas as tarefas vistas no Anexo D, após a análise do *dashboard*. A pontuação para os 7 princípios foi definida pelo auxílio de médias, medianas e intervalos da menor e maior escala, onde o cálculo é feito por intermédio dos resultados de cada fator de usabilidade referente ao princípio sugerido. Esses registros são mostrados nas Tabelas 21, 22 e 23 a seguir.

Tabela 21 – Resultados das escalas por média para cada princípio de usabilidade

Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5	Tarefa 6	Tarefa 7
Usuário 1	7,2	7,2	7,5	7,7	6,8	8,3	6,7
Usuário 2	8,8	9,6	10,0	9,7	8,5	9,7	9,3
Usuário 3	9,8	10,0	9,3	10,0	9,5	10,0	10,0
Usuário 4	7,6	9,2	8,5	8,3	9,3	8,7	8,3
Usuário 5	9,4	7,8	7,3	9,0	7,3	8,3	10,0
Usuário 6	8,0	8,0	7,5	8,0	8,0	7,7	9,0
Usuário 7	10,0	10,0	10,0	10,0	9,8	10,0	10,0
Usuário 8	9,8	9,4	8,5	9,7	9,5	8,7	9,3
Usuário 9	9,8	8,0	9,8	9,7	9,8	9,7	9,3
Usuário 10	9,6	9,4	9,8	10,0	9,8	10,0	10,0

Tabela 22 – Resultados das escalas por mediana para cada princípio de usabilidade

Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5	Tarefa 6	Tarefa 7
Usuário 1	7,0	7,0	7,5	7,7	6,5	8,0	7,0
Usuário 2	9,0	10,0	10,0	9,7	9,0	10,0	10,0
Usuário 3	10,0	10,0	10,0	10,0	10,0	10,0	10,0
Usuário 4	8,0	9,0	8,5	8,3	10,0	9,0	10,0
Usuário 5	10,0	8,0	8,5	9,0	7,5	9,0	10,0
Usuário 6	8,0	8,0	7,5	8,0	8,0	8,0	9,0
Usuário 7	10,0	10,0	10,0	10,0	10,0	10,0	10,0
Usuário 8	10,0	9,0	8,5	9,7	9,5	9,0	9,0
Usuário 9	10,0	8,0	10,0	9,7	10,0	10,0	10,0
Usuário 10	10,0	9,0	10,0	10,0	10,0	10,0	10,0

A média de idade dos participantes do teste de usabilidade eram de 24 anos, onde todos eram alunos de graduação. Apenas um deles possuía conhecimentos sobre requisitos de usabilidade aplicados a *dashboards*, como também não necessitaram de requisitos de acessibilidade digital para a utilização do painel. Dos 10 participantes, 4 deles abrangiam uma visão do *dashboard* com a função de "monitorar dados em tempo real e representá-los com fácil compreensão visual". Enquanto 4 deles traziam como perspectiva "um painel visual de informações que contém métricas e indicadores". Dos dois avaliadores que sobraram, apenas um deles concordava que essas aplicações têm como foco "representar números relevantes para criar estratégias e alcançar os objetivos", entretanto, o outro optou por "um painel que serve para acompanhar o desempenho e embasar decisões".

Tabela 23 – Resultados das escalas por intervalo para cada princípio de usabilidade

Participantes	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5	Tarefa 6	Tarefa 7
Usuário 1	6 - 8	6 - 9	6 - 9	7 - 8	6 - 8	8 - 9	5 - 8
Usuário 2	7 - 10	9 - 10	10 - 10	9 - 10	6 - 10	9 - 10	8 - 10
Usuário 3	9 - 10	10 - 10	7 - 10	10 - 10	8 - 10	10 - 10	10 - 10
Usuário 4	7 - 8	8 - 10	7 - 10	6 - 10	7 - 10	8 - 9	5 - 10
Usuário 5	7 - 10	5 - 10	2 - 10	7 - 10	4 - 10	6 - 10	10 - 10
Usuário 6	8 - 8	8 - 8	7 - 8	8 - 8	8 - 8	7 - 8	9 - 9
Usuário 7	10 - 10	10 - 10	10 - 10	10 - 10	9 - 10	10 - 10	10 - 10
Usuário 8	9 - 10	9 - 10	8 - 9	9 - 10	9 - 10	8 - 9	9 - 10
Usuário 9	9 - 10	6 - 10	9 - 10	9 - 10	9 - 10	9 - 10	8 - 10
Usuário 10	9 - 10	9 - 10	9 - 10	10 - 10	9 - 10	10 - 10	10 - 10

Foi questionado também aos avaliadores um levantamento de vantagens e desvantagens de um *dashboard*, onde a maioria descartou qualquer dificuldade com esse tipo de aplicação. No entanto, alguns problemas e sugestões foram relatados por eles durante o teste, como a ausência de uma opção de busca por CEP do município e, além disso, adicionar uma cor diferente quando o mouse está em cima de uma área clicável no painel.

Figura 18 – *Word Cloud* (Nuvem de Palavras)

Na Figura 18 mostra a *Word Cloud* (Nuvem de Palavras) referente as sugestões dos participantes durante a avaliação. Quanto maior e mais escura a palavra, maior foi a sua menção. E quanto menor e mais clara a palavra, menor foi a sua citação. Sendo assim, esta visualização

visa o comparativo entre as palavras mais relevantes de forma intuitiva, nos problemas relatados pelos usuários.

Além disso, a maior dificuldade enfrentada neste estudo foi buscar participantes dispostos a gastar entre 10 à 15 minutos do seu tempo para a realização dos testes. Também foi necessário adicionar mais questões para identificar o perfil do usuário, bem como obter a área que ele cursa e o grau de experiência com *dashboards*.

De acordo com a análise dos participantes, as Tarefas 5 e 7 foram as que tiveram a menor média de escala, visto na Tabela 21. Já na Tabela 22, as Tarefas 1 e 5 obtiveram a menor mediana de escala. Enquanto na Tabela 23, as Tarefas 3 e 5 foram as que possuíram o menor valor de escala dentre os itens de cada atividade. Logo, como retificação, será aprimorada a visibilidade do *dashboard*, exemplificativamente, as descrições, as mensagens e as cores na tela, além de realçar a interface e torná-la mais familiar aos usuários.

Por fim, a contribuição dos convidados nesta pesquisa facilita em aplicar melhorias futuras no *dashboard*, de acordo com os resultados finais da avaliação heurística, os obstáculos encontrados e o panorama apresentados por eles.

5

Análise dos dados usando algoritmos de *Machine Learning*

Uma vez que a pandemia tem um efeito extremamente negativo no desenvolvimento econômico e no sistema médico, é necessário criar políticas e planos para impedir a propagação do COVID-19 e de outras doenças. Para aumentar a pertinência das políticas governamentais, os pesquisadores devem examinar as condições relevantes relacionadas à propagação de um surto epidêmico. Além disso, se os pesquisadores puderam prever a propagação dessas doenças com uma análise do número de mortes e recuperações, os responsáveis pela tomada de decisões para combater o contágio podem usar o resultado da previsão para decidir ações imediatas.

O COVID-19 surgiu como um grande desafio para a área médica e campo de pesquisa. Os métodos baseados em Inteligência Artificial (IA) provaram ser úteis para prever os riscos de contágio e os efeitos de tais surtos pandêmicos no passado, como por exemplo, a pneumonia e o câncer de pele ([AGARWAL et al., 2021](#)). Portanto, as técnicas avançadas de IA podem ser usadas para prever resultados futuros e prevenir a propagação de tais doenças. Os modelos propostos tendem a alcançar uma alta precisão na previsão, o que ajudará a melhorar a investigação aprofundada ([MALEKI et al., 2020](#)). O intuito é conscientizar e divulgar o número correto de óbitos em todo o planeta.

Em geral, as pandemias podem ocorrer com frequência. Durante tempos de crise de saúde pública, é importante pesquisar a disseminação, as etapas de prevenção e o comportamento no futuro. Constantemente, pesquisadores e especialistas desenvolvem diferentes medicamentos para combater a crise, e a pandemia do COVID-19 não é exceção. A população hoje têm acesso a várias quantidades de dados devido à internet e sua acessibilidade. As informações coletadas documentam o número de pessoas que foram infectadas, curadas e mortas por determinada

doença. O design e a sequência dos dados coletados auxiliam no desenvolvimento de modelos robustos para prever a propagação de uma pandemia no futuro. Atualmente, um conjunto de dados numa série temporal do COVID-19 pode ser usado para fazer previsões, isto é, onde informações sobre seus sintomas podem ser utilizadas para identificar o estágio inicial da doença e ilustrar a probabilidade de previsão do vírus (AGARWAL et al., 2021).

5.1 Motivações

Pesquisadores sobre o COVID-19 testemunham que existem relações positivas e significativas entre o total de casos confirmados, o total de óbitos e o tamanho da população dos países. Por isso propõem uma abordagem inteligente híbrida para prever séries temporais da doença, combinando teoria fractal e lógica difusa com base nos dados de casos confirmados e mortes (CASTILLO; MELIN, 2020). Eles adotam abordagens simples de previsão de séries temporais, usando modelos da família de suavização exponencial (PETROPOULOS; MAKRIDAKIS, 2020). Além disso, eles analisam as notícias consultadas e classificam as notícias em sentimentos negativos e positivos, para entender a influência das notícias no comportamento das pessoas, tanto política quanto economicamente (HAMZAH et al., 2020).

No entanto, alguns desses estudos usam apenas variáveis limitadas para análise, portanto, os fatores potenciais que podem estar relacionados ao surto do COVID-19 e de outras doenças serão ignorados. Nesse caso, a confiabilidade de sua previsão permanecerá obscura. Além disso, alguns estudos usam análise de dados e métodos estatísticos tradicionais, que podem não ser confiáveis para prever a pandemia, pois a enorme quantidade de dados relacionados, como por exemplo, à disseminação do COVID-19 é complexa e varia em diferentes períodos de tempo.

Tendo em vista, EUA e Brasil são os países que alcançaram o maior número de mortes do COVID-19 no mundo até o momento, desde que a OMS decretou o fim da pandemia causada pela doença, três anos depois do surgimento dos primeiros casos (KUPFERSCHMIDT; WADMAN, 2023). Ambos obtêm fatores socioeconômicos, geográficos e estruturais distintos, como o seu Índice de Desenvolvimento Humano (IDH), renda per capita, o nível de escolaridade da população e a distância para um hospital (especialmente em áreas rurais e menos desenvolvidas), sendo razões biológicas na determinação do resultado da pandemia de cada região.

5.1.1 *Google Datasets Search*

Com o propósito de encontrar dados abertos de sistemas prisionais em geral, foi usada a ferramenta de busca desenvolvida pela empresa *Google*, conhecida como *Google Datasets Search*. Anunciada em setembro de 2018, é um mecanismo de busca específico para encontrar registros de pesquisas publicados na internet (CANINO, 2019). Além disso, beneficia cientistas a obterem todos os tipos de informações disponíveis em diferentes portais na internet. A ferramenta traz uma alta capacidade de rastreamento, identificação, indexação, interoperação e disponibilização

de conjuntos de dados disponíveis, utilizando padrões internacionais e apresentando um potencial considerável.

5.1.2 Dados abertos de *Connecticut*

Foram encontrados 988 registros de óbitos do COVID-19 em instalações correcionais do estado de *Connecticut*, nos Estados Unidos. Essas informações podem ser acessadas em um repositório com mais de 700 conjuntos de dados disponíveis, conhecido como *CT Open Data Portal*, hospedado na plataforma *Socrata* ([COVID-19...](#), 2023).

Além dos dados de óbitos do COVID-19 entre detentos, também podem ser detectados:

- Número total de funcionários positivos para o COVID-19;
- Número total de detentos positivos para o COVID-19;
- Número total de detentos positivos para o COVID-19, alojados em Unidades de Isolamento Médico;
- Equipes que obtiveram positivo para o COVID-19 e que voltaram ao trabalho;
- Número total de detentos medicamente liberados.

O intuito do repositório é facilitar a interação com os dados, além da compreensão do contexto e das tendências. Por exemplo, exibir painéis e outros tipos de ativos, inclusive, criar suas próprias visualizações, como um gráfico ou um mapa dentro do portal.

5.1.3 *The Marshall Project*

No início da pandemia do COVID-19 nos EUA, em março de 2020, repórteres do *The Marshall Project* começaram a rastrear encarcerados infectados e mortos pela doença em prisões estaduais e federais em todo o país, bem como aos funcionários dessas instalações. Trabalhando com a *Associated Press*, esses números eram solicitados todas as semanas, aos departamentos estaduais de correções e ao *Federal Bureau of Prisons* (BOP), sendo a agência federal de aplicação da lei dos EUA ([PARK; MEAGHER, 2021](#)).

No entanto, nem todos os departamentos puderam fornecer dados para as datas exatas solicitadas. Logo, esses números foram agrupados pela semana em que os dados foram coletados. Pelo menos 398.627 pessoas presas testaram positivo para o COVID-19. Dentre o número de óbitos, 2.714 presos morreram de causas relacionadas ao coronavírus (onde 2.518 registros foram exportados para os testes de ML). Apenas o estado de *Vermont*, nos EUA, não assinalou mortes de detentos atribuídas à COVID-19. Enquanto isso, até o final de junho de 2021, 114.237 funcionários de presídios testaram positivo para a doença, e os novos casos atingiram um recorde histórico na semana de 22 de dezembro de 2020. Em vista disso, as prisões relataram publicamente 209 mortes entre funcionários.

Em 22 de junho de 2021, à medida em que os casos de COVID-19 estavam diminuindo em relação aos picos e as agências penitenciárias que divulgavam menos informações, o projeto parou de coletar esses dados.

5.2 Objetivos

Em relação à questão acima, métodos de ML foram empregados neste trabalho de dissertação para analisar os dados epidêmicos do estado da Bahia (522 registros). Nos quais foram exportados do site do Departamento de Informática do Sistema Único de Saúde (DATASUS), com o intuito de prever o desenvolvimento futuro do COVID-19 e outras CIDs, onde abrange grupos de doenças respiratórias, infecciosas, sanguíneas e mentais em diferentes municípios da região (visto na Seção 3).

Com o auxílio do *Google Datasets Search*, também foram exportados outros conjuntos de dados abertos do COVID-19, em comparação com as informações exportadas do DATASUS. O acesso a ferramenta foi feito pelo site: <<https://datasetsearch.research.google.com>>. Foi usado o termo "*COVID-19 correctional*" no processo de consulta, onde foram obtidos mais de 100 resultados (quando chegou a “mais de 100”, a ferramenta limitou a quantidade de registros e não mostrou de fato o número real de itens obtidos) indicando vários conjuntos de dados contendo estes termos. No entanto, apenas dois foram selecionados devido a coerência com o tema, disponibilidade, formato e facilidade de acesso aos dados. O primeiro contém o número de óbitos de presidiários de instalações correcionais de *Connecticut*, nos EUA. O segundo é uma redação investigativa sem fins lucrativos dedicada ao sistema de justiça criminal dos EUA e à *Associated Press*, conhecido como *The Marshall Project*, onde seu objetivo é analisar surtos da doença nas prisões estaduais e federais americanas.

Em seguida, foram usados algoritmos de Regressão para descobrir qual modelo se ajusta melhor aos dados de treinamento e de teste. Esses algoritmos são utilizados para prever se os dados são contínuos, como as previsões do tempo, de preços ou tendências de mercado (KUMAR et al., 2020). No nosso caso, a intenção é averiguar o progresso do número de óbitos devido ao COVID-19 e outras doenças nos sistemas prisionais. Após isso, os dados foram divididos e selecionados diferentes modelos para ajustar os dados de períodos distintos para tornar os resultados mais precisos. Considerando que os municípios têm situações diferentes, este trabalho emprega o método de seleção de recursos para descobrir as variáveis mais importantes que influenciam a disseminação das doenças em divergentes regiões.

Finalmente, pode-se concluir que os algoritmos de ML podem ser usados para prever a propagação dessas doenças, o que pode ajudar os responsáveis a tomar ações imediatas. Além disso, explora-se fatores relacionados ao surto do COVID-19 em diferentes regiões para que as políticas no sentido de conter o desenvolvimento da doença possam ser mais direcionadas.

5.3 Metodologia

Iniciou-se definindo, escolhendo e adquirindo os dados com os quais precisávamos trabalhar. No entanto, determinados dados com os quais iniciamos não são estruturados e brutos, e nunca estão no formato correto para processamento. Pode ter dados duplicados, dados ausentes ou várias informações extras que não são necessárias. As informações podem vir de várias fontes e podem acabar sendo duplicadas. Nesse caso, foi necessário pré-processar esses registros para que os algoritmos de ML possam entendê-los, e o lado bom é que a maioria desses métodos possuem módulos de pré-processamento de dados para executar informações brutas ou não estruturadas. No caso dos registros do DATASUS, após exportá-los e salvá-los em tabelas de um banco de dados privado, obtemos alguns bons dados estruturados depois que os dados são pré-processados, e esses registros agora são uma entrada para o processo de aprendizado, visto na Figura 19.

Esse procedimento deve ser iterativo até que os dados estejam disponíveis, isto é, onde a maior parte do tempo do processo de *Machine Learning* é gasto. Logo, é necessário trabalhar os dados para deixá-los mais organizados, seguros, prontos e disponíveis. Além disso, as informações podem ser submetidas aos algoritmos, no qual fornecem não apenas métodos de pré-processamento, mas também uma ampla variedade de métodos de ML. Por fim, um modelo específico é que aplica os dados, mas a questão mesmo é identificar o método final de qual precisamos. Portanto, fizemos a avaliação do modelo e foram considerados os algoritmos com os melhores resultados de acordo com os critérios de avaliação.

5.3.1 Análise de Regressão

Para a previsão dos dados, usou-se a Regressão Polinomial, não sendo um conceito da área de *Machine Learning* em si, e sim, da área de estatística. Tem como finalidade analisar a relação entre duas ou mais variáveis, em que uma depende de outra ou outras.

Em comparação a Regressão Polinomial, a Linear é operada em duas variáveis contínuas para encontrar a relação entre elas. Já a Polinomial modela a relação entre a variável independente X e a variável dependente Y como um polinômio de enésimo grau em X (BHUTIA; PATRA, 2022). Pode ser considerada uma versão estendida de Regressão Linear, implementada em dados relacionados, mas não lineares, como pode ser visto na Figura 20.

A Regressão Polinomial tenta modelar uma relação não linear entre variáveis independentes e dependentes (tecnicamente, entre a variável independente e a média condicional da variável dependente). Uma forma localizada de Regressão Polinomial clássica é usada em algumas dessas abordagens.

Note que a Regressão Polinomial prevê a melhor linha de ajuste que segue o padrão (curva) dos dados. Ela geralmente é usada quando os pontos nos dados não são capturados pelo modelo de Regressão Linear e acaba falhando em descrever claramente o melhor resultado. À

Figura 19 – O funcionamento do modelo de previsão através de um fluxograma

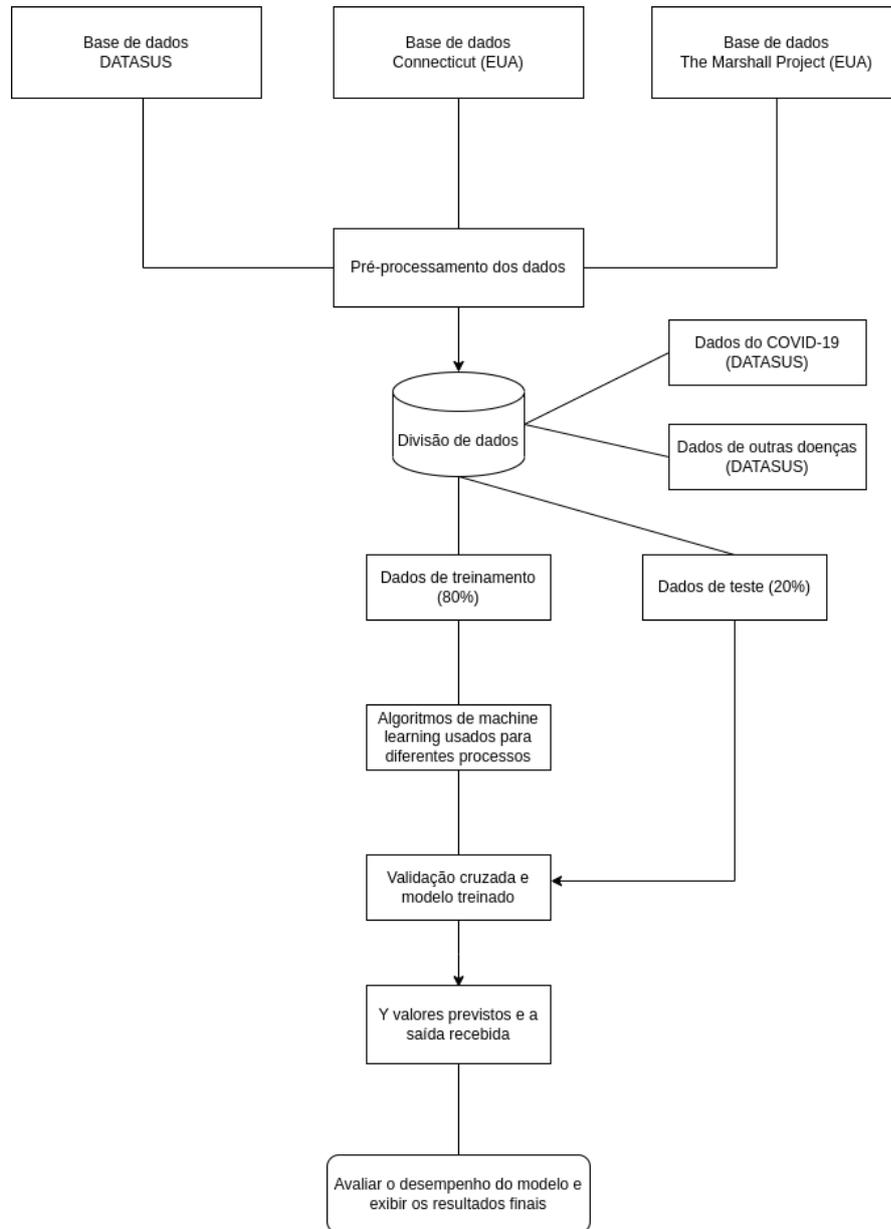
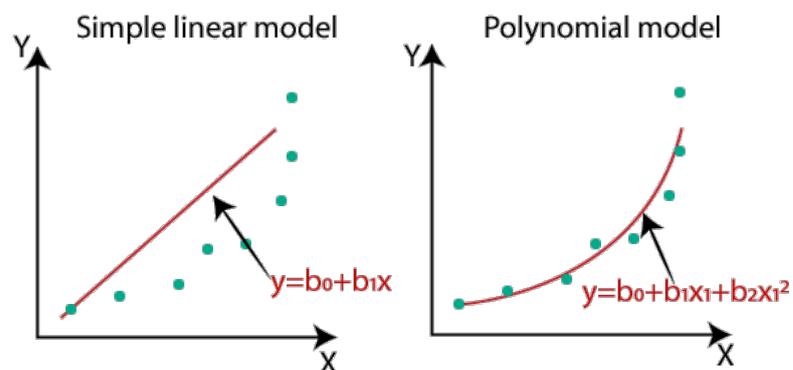


Figura 20 – Regressão Linear Simples e Polinomial



medida que aumentamos o grau no modelo, ele tende a aumentar o desempenho do algoritmo. No entanto, aumentar os graus do modelo também aumenta o risco de superajuste e subajuste dos dados.

5.3.2 *Support Vector Machine (SVM)*

O *Support Vector Machine* também conhecido como SVM, é uma técnica de ML supervisionada empregada tanto para tarefas de regressão, quanto para tarefas de classificação. O objetivo principal deste algoritmo é identificar pontos de dados utilizando um hiperplano em um espaço N-dimensional. O tamanho e a natureza do hiperplano dependem do número de recursos disponíveis. Por exemplo, com duas características de entrada, o hiperplano assume a forma de uma linha simples, enquanto com três características de entrada, torna-se um plano bidimensional. O SVM emprega várias funções de *kernel*, incluindo linear, não linear, polinomial, sigmoide e a Função de Base Radial (RBF - *Radial Base Function*), sendo este último utilizado nos experimentos com os dados (CHOPRA; KHURANA, 2023).

5.3.3 *Random Forest*

Random Forest é um algoritmo que utiliza uma técnica que combina vários classificadores para resolver problemas complexos. Este modelo consiste em uma coleção de Árvores de Decisão (*Decision Trees*). Ou seja, essa "floresta" é treinada por meio de um processo chamado *bagging* ou agregação de *bootstrap*. Este método serve como um meta-algoritmo para conjuntos, aumentando a eficiência dos modelos de ML. O modelo determina seu resultado considerando as previsões das árvores de decisão individuais (SLÁDEK, 2023). A previsão final é normalmente obtida pela média ou cálculo da média das saídas do conjunto de árvores. Sendo que a ideia de aumentar o número de árvores no *Random Forest* melhora a precisão do resultado geral.

5.3.4 *Decision Tree*

Este algoritmo é uma das abordagens práticas mais comumente usadas para aprendizado supervisionado. Mesmo sendo um modelo mais aplicado em tarefas de Classificação, ele também pode ser usado para resolver tarefas de Regressão. É um classificador estruturado em árvore com três tipos de nós. O nó raiz é o nó inicial que representa toda a amostra e pode ser dividido em outros nós. Já os nós interiores representam as características de um conjunto de dados e os ramos representam as regras de decisão. Finalmente, os nós de folha representam o resultado (SHARMA; IQBAL, 2023). Este algoritmo é muito útil para resolver problemas relacionados à decisão.

5.3.5 Implementação

5.3.5.1 Pré-processamento dos dados

No sistema carcerário da Bahia, foram coletados dados diários de mortes do COVID-19 e de outras doenças nos 417 municípios do estado, nos meses de janeiro até outubro de 2020. Para confrontar, também foram exportados outros dois conjuntos de dados públicos de ambientes estaduais e federais dos EUA. Por exemplo, no *The Marshall Project* foi feito um levantamento do surto do COVID-19 no país, entre os meses de março de 2020 até junho de 2021. Já o conjunto de dados do estado de *Connecticut*, para este estudo foram coletados registros dos meses de março de 2020 até junho de 2023. Em vista disso, foram desenvolvidos modelos de predição usando algoritmos de Regressão, com o auxílio da linguagem *Python* e suas bibliotecas:

- *Numpy*: para operações matemáticas e estatísticas;
- *Matplotlib*: em que a classe *pyplot* é usada para visualizar os dados;
- *Pandas*: para criar quadro de dados e manipular ou ajustar dados. Na Figura 21 temos um exemplo de importação do conjunto de dados usando o método *read_csv* da biblioteca. Para os dados referentes ao período diário, foi armazenada a variável independente, isto é, as datas em *X* como variável dependente e o número de óbitos em *y*. Em seguida, os dados são acessados usando o método *iloc* que obtém os registros localizados no índice de linhas e colunas;
- *Datetime*: fornece as classes para manipulação de datas e horas;
- *Math*: obtém acesso às funções matemáticas;

Figura 21 – Importando os dados para a Análise de Regressão

```
import math
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np
from sklearn.model_selection import train_test_split
%matplotlib inline
```

• *Scikit-learn*: sendo a principal biblioteca de ML em *Python*, onde possui a implementação de inúmeros métodos de aprendizado, desde algoritmos de Regressão e Classificação, como também as Redes Neurais. Em seguida, é criado um conjunto de treinamento e o conjunto de teste usando a essa biblioteca como classe da *model_selection*. E por fim, é feita a divisão de 20% do conjunto de teste (*test_size*) do total de dados, com o gerador de números aleatórios (*random_state*) igual a 3.

Para treinar o modelo, usa-se as classes *PolynomialFeatures* (grau 2) do módulo *preprocessing*, *LinearRegression* do módulo *linear_model*, *DecisionTreeRegressor* do módulo *tree*,

RandomForestRegressor do módulo *ensemble* com profundidade máxima de 3 e *SVR* do módulo *svm* incluindo a função de *kernel* RBF.

Com o propósito de esclarecer melhor, nas Figuras 21, 22, 23 e 24 ilustram a implementação das previsões da Análise de Regressão percorrendo um conjunto de teste *x_test* para o método *predict* do objeto *lr*.

Figura 22 – Dividindo os dados em conjuntos de treinamento e teste

```
import datetime as dt

X = df[['data_saida']]
Y = df['mortes']

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=3)
```

Figura 23 – Formatando as informações das datas de óbito

```
import datetime as dt

x_train = X_train['data_saida'].apply(pd.to_datetime, format='%Y/%m/%d')
x_train = x_train.apply(dt.datetime.toordinal)
x_train = np.array(x_train).reshape(-1,1)

x_test = X_test['data_saida'].apply(pd.to_datetime, format='%Y/%m/%d')
x_test = x_test.apply(dt.datetime.toordinal)
x_test = np.array(x_test).reshape(-1,1)
```

Figura 24 – Instanciando o modelo de Regressão Polimomial

```
from sklearn.preprocessing import PolynomialFeatures
poly = PolynomialFeatures(degree=2)
train_x_poly = poly.fit_transform(x_train)
train_x_poly

test_x_poly = poly.transform(x_test)
y_pred = lr.predict(test_x_poly)
```

5.3.6 Resultados

5.3.6.1 Critérios de avaliação

Como parâmetros de avaliação e implementados por meio da biblioteca *Scikit-learn* do *Python* (visto na Figura 21), usaremos as variáveis a seguir para calcular a exatidão de nossas previsões.

- **R-quadrado (*R-squared* - R^2):** o coeficiente de determinação ou R-quadrado representa a proporção da variância na variável dependente que é explicada pelo modelo de regressão. É uma pontuação sem escala, ou seja, independentemente dos valores serem pequenos ou grandes,

o valor de R^2 será menor que um

$$R^2 = 1 - \frac{\sum (y_i - y'_i)^2}{\sum (y_i - \bar{y})^2}$$

• **Erro Quadrático Médio (Mean Square Error - MSE):** representa a diferença quadrática média entre os valores estimados (y_i) e os valores previstos (y'_i)

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

• **Raiz do Erro Quadrático Médio (Root Mean Squared Error - RMSE):** é a raiz quadrada do MSE. Ele mede o desvio padrão dos resíduos

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}$$

• **Erro Médio Absoluto (Mean Absolute Error - MAE):** é a média de todos os erros

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$

5.3.6.2 Predições dos modelos treinados

Os modelos treinados e testados com os modelos de regressão são utilizados para prever o número total de casos de morte por dia. Nas Tabelas 24 e 25 mostram os valores previstos de morte entre os dados do COVID-19 e de outras doenças, respectivamente, onde foram exportados do site do DATASUS. Já nas Tabelas 26 e 27, são apresentadas as predições com os dados gerais de óbitos do COVID-19 de instalações correcionais do estado de *Connecticut*, nos EUA, e dos registros levantados pelo *The Marshall Project*, por essa ordem.

Tabela 24 – Resultados das predições do DATASUS (COVID-19)

Modelo	MAE	MSE	RMSE	R ²
Regressão Polinomial	7	72	8.49	0.78
Decision Tree	8	122	11.03	0.62
Random Forest	9	116	10.77	-0.44
SVM	8	91	9.55	0.72

Para uma melhor comparação e eliminar qualquer incerteza na análise do conjunto de dados do DATASUS, nos quais foram coletados apenas no ano de 2020, época em que surgiram as primeiras mortes do COVID-19, resolvemos calcular outras predições. Somente os registros do sistema carcerário dos EUA, até o mês de dezembro deste mesmo ano, esboçados nas Tabelas 28 e 29.

Tabela 25 – Resultados das previsões do DATASUS (outras doenças)

Modelo	MAE	MSE	RMSE	R²
Regressão Polinomial	17	449	21.19	0.45
Decision Tree	18	457	21.38	0.44
Random Forest	27	904	30.07	-5.67
SVM	17	493	22.2	0.40

Tabela 26 – Resultados das previsões de *Connecticut*

Modelo	MAE	MSE	RMSE	R²
Regressão Polinomial	2	4	1.93	0.77
Decision Tree	2	3	1.82	0.80
Random Forest	2	4	1.88	0.79
SVM	2	4	1.91	0.78

Tabela 27 – Resultados das previsões de *The Marshall Project*

Modelo	MAE	MSE	RMSE	R²
Regressão Polinomial	32	2344	48.42	0.06
<i>Decision Tree</i>	34	2743	52.37	-0.10
<i>Random Forest</i>	32	2363	48.61	0.05
SVM	27	2564	50.63	-0.03

Tabela 28 – Resultados das previsões de *Connecticut* em 2020

Modelo	MAE	MSE	RMSE	R²
Regressão Polinomial	2	9	2.99	-3.80
<i>Decision Tree</i>	3	9	2.94	-3.66
<i>Random Forest</i>	3	9	2.93	-3.62
SVM	3	9	2.93	-3.60

Tabela 29 – Resultados das previsões de *The Marshall Project* em 2020

Modelo	MAE	MSE	RMSE	R²
Regressão Polinomial	28	1793	42.35	0.04
<i>Decision Tree</i>	29	1999	44.71	-0.07
<i>Random Forest</i>	28	1814	42.59	0.03
SVM	24	2049	45.26	-0.10

Os critérios de avaliação vão ajudar a escolher o melhor modelo entre suas possibilidades.

Por exemplo, os resultados de R^2 variam geralmente entre 0 e 1. Ainda sim, caso algum modelo não se ajuste ao algoritmo dado, valores negativos de R^2 serão obtidos. O MAE apresenta uma boa interpretação, principalmente para mencionar os resultados da predição. Já o MSE é afetado por valores discrepantes, o que pode ser relevante quando é preciso ter uma análise mais criteriosa do algoritmo. No entanto, o RMSE é amplamente usado do que o MSE para avaliar o desempenho do algoritmo de regressão com outros modelos aleatórios, pois possui as mesmas unidades da variável dependente (eixo Y).

Como podemos observar, não existe a melhor métrica. Diante disso, tudo depende do cenário e dos objetivos para analisar os critérios e definir a performance do algoritmo, com o intuito de auxiliar a tomar uma decisão. O menor valor de MAE, MSE e RMSE implica a maior acurácia de um modelo de regressão. Enquanto isso, o maior valor de R^2 é considerado desejável, ou seja, quanto mais próximo de 1, melhor é a precisão do modelo (CHICCO; WARRENS; JURMAN, 2021).

Avaliando como exemplo, o R^2 , podemos ter uma falsa sensação de que o modelo seja aceitável, entretanto, quando resolvemos usar uma outra métrica, o algoritmo resulta em um desempenho ruim. Portanto, o melhor é sempre utilizar mais de um critério de avaliação para garantir que não há nenhum problema na predição do nosso modelo.

5.4 Considerações finais da análise dos dados usando *Machine Learning*

Por meio deste projeto, tentou-se implementar modelos de previsão do COVID-19 e de outras CIDs, onde abrange grupos de doenças respiratórias, infecciosas, sanguíneas e mentais. Os seus resultados geraram predições e gráficos que podem nos ajudar a obter a situação atual das causas de mortes patogênicas nos sistemas prisionais. Isso pode auxiliar departamentos governamentais a tomarem um planejamento mais eficaz, pois elas podem comparar facilmente usando uma representação gráfica do que com valores numéricos. A validação de um modelo preciso e a análise dos dados podem contribuir para estratégias de cura e prevenção de doenças em estágio inicial.

Em primeiro lugar, foram pesquisados vários algoritmos de ML para a previsão. Com os melhores resultados, obtemos a Regressão Polinomial usando os dados do DATASUS. Tanto para os óbitos do COVID-19 quanto para os falecimentos pertencentes as outras CIDs, como as doenças infecciosas (A419 - Septicemia não especificada e A499 - Infecção bacteriana não especificada), circulatórias (I64 - Acidente vascular cerebral não especificado como hemorrágico ou isquêmico e I500 - Insuficiência cardíaca congestiva) e respiratórias (J960 - Insuficiência respiratória aguda e J189 - Pneumonia não especificada), onde alcançaram o maior número de óbitos. Já nas predições do estado de *Connecticut*, a *Decision Tree* obteve a melhor aferição (exceto nos registros de 2020, onde ambos a estimativa de R^2 foi negativa), enquanto a Regressão

Polinomial e a *Random Forest* alcançaram as melhores exatidões com os dados do *The Marshall Project*.

Percebe-se que as predições com os dados do DATASUS (com exceção dos modelos do *Random Forest*), referentes aos sistemas prisionais do estado da Bahia, obtiveram um resultado significativo que traduz na época o avanço da pandemia do COVID-19, tanto na região como também no Brasil. Já as previsões dos dados do COVID-19 relativos aos ambientes carcerários dos EUA, apenas os registros gerais de *Connecticut* obteve uma apuração considerável do início até a situação atual do surto da doença, no país.

Logo, com os resultados do estado da Bahia, uma região do Nordeste do Brasil com carência socioeconômica, observamos que na época da pandemia houve uma ausência de um plano de contingência no local. Incluindo o treinamento de equipes e profissionais no sistema prisional para lidar com a prevenção do COVID-19, por meio do uso intensivo de distanciamento físico, uso de banheiros, uso de máscaras e higiene respiratória.

Em 2020, o COVID-19 se espalhou rapidamente por todo o Brasil, apesar do país ter um amplo sistema público de saúde, o SUS, que é administrado em conjunto pelos governos municipais, estaduais e federais. No entanto, aconteceu um intenso conflito entre as autoridades sobre a relevância e a extensão de tais medidas, onde unidades federativas implementaram diferentes medidas de mitigação, o que refletiu disparidades localizadas na gravidade da pandemia.

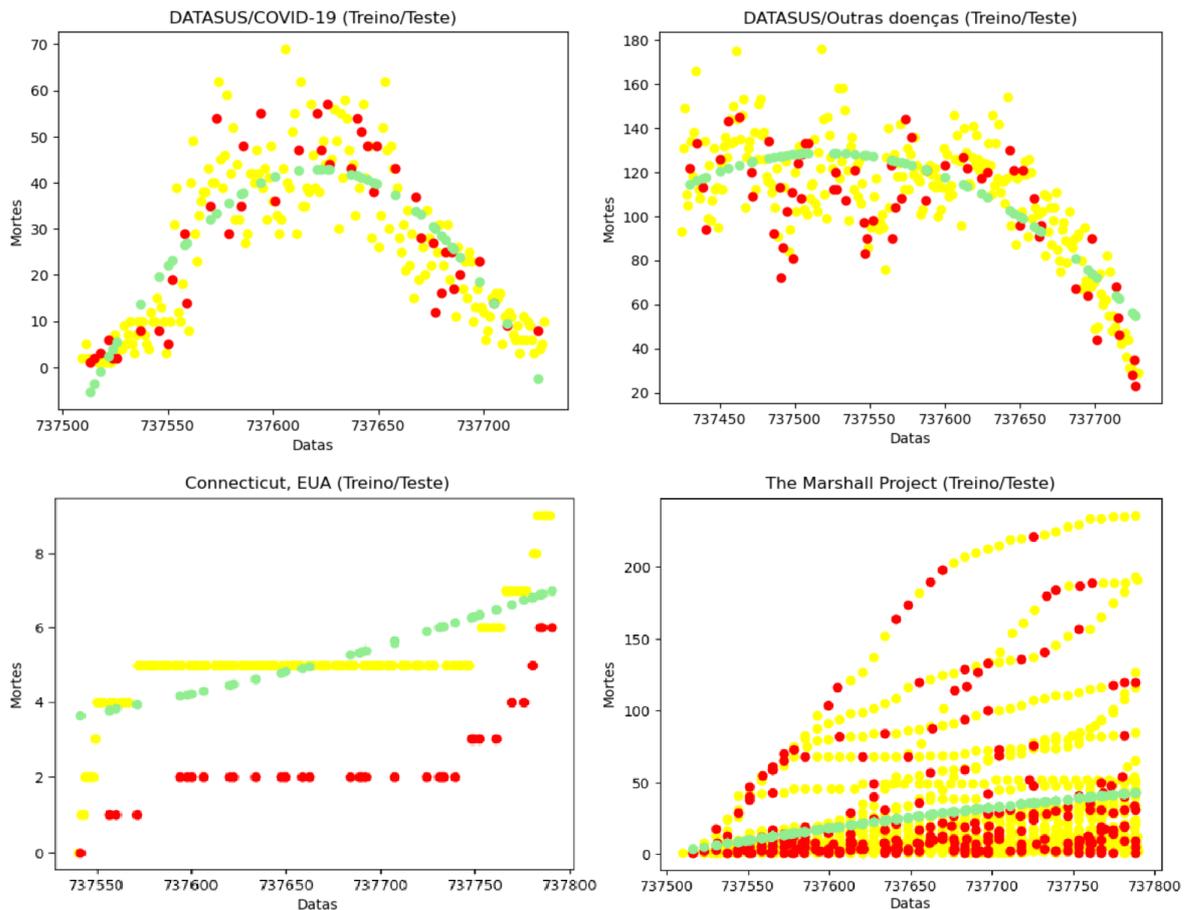
O presidente brasileiro, seus aliados políticos e altos funcionários do governo federal na época, também se esforçaram para sabotar várias intervenções bem estabelecidas, como distanciamento social, uso de máscara e vacinação (FURLAN; CARAMELLI, 2021). O governo federal combateu a adoção de medidas impostas por estados e municípios, como a obrigatoriedade do uso de máscaras faciais e regras de distanciamento social. O presidente também levantou dúvidas sobre a segurança e eficácia das vacinas contra o COVID-19, e buscou promover o uso de medicamentos sem comprovação (hidroxicloroquina e ivermectina) em conjunto com os seus apoiadores, onde adotaram postura semelhante e na maioria das vezes com intensidade exacerbada.

Isso pode ser visto nos gráficos da Regressão Polinomial localizadas na Figura 25, onde os dados do DATASUS obtiveram os melhores resultados em comparação com os outros dois conjuntos de dados, restritos em 2020, usando o mesmo algoritmo de regressão. Notem que as bolas/linhas amarelas representam os dados de treino, enquanto os pontos/retas vermelhas ilustram os dados de teste, e por fim, as esferas/traços verdes refletem nas predições dos modelos.

Logo, os gráficos ilustram um modelo dinâmico não linear que retratam a evolução da pandemia, apresentando um comportamento aleatório e natureza imprevisível do surto do COVID-19 entre detentos, nos quais foram infectados por meio de contato direto e indireto.

Nas regiões mais ricas como os países desenvolvidos, onde os EUA são um exemplo, tendem a ter melhores redes de atenção à saúde e maior capilaridade do sistema de saúde em

Figura 25 – Gráficos gerados dos modelos de Regressão Polinomial



regiões distantes dos centros urbanos. Além disso, em estados com maiores frações de residentes urbanos, tende a decorrer uma maior transmissão do vírus e disseminação da doença.

Porém, há cerca de 7 meses após o início da pandemia, os EUA não tinham um plano estratégico de preparação e resposta, e violaram persistentemente os princípios de epidemiologia de campo, levando três revistas proeminentes de saúde pública, como *Scientific American*, *The Lancet* e *New England Journal of Medicine*, a fazer críticas abrangentes sobre a má gestão do COVID-19 pelo governo na época (MISIR, 2022).

A diferença nas situações em cada Estado do EUA foi um aspecto relevante em meio à pandemia. Muitos analistas consideraram que houve falta de liderança por parte da Casa Branca, e o enfrentamento ao novo coronavírus ficou a cargo dos governadores. Enquanto alguns impuseram restrições precocemente e declararam estado de emergência antes da explosão dos casos, outros optaram por não fazê-lo, o que, na opinião dos especialistas, pode ter contribuído para a ascensão de infecções.

Em 2020, as declarações controversas e contraditórias do presidente americano, comparando a doença com uma simples "gripe", passou a ideia de que o país estava tudo sob controle. E o sistema de saúde nos EUA baseado principalmente em convênios privados, também complicou

a situação no momento em que é necessário identificar casos do COVID-19, de acordo com epidemiologistas. Isso faz com que algumas dessas pessoas que não tem plano evitem de ir ao médico em caso de infecção, sendo que uma consulta pode custar centenas de dólares.

Em vista disso, com as minorias raciais e étnicas levando desvantagem, os dados de *Connecticut* mostraram que as pessoas pretas morreram da doença com o dobro da taxa de pessoas brancas. A população preta foi desproporcionalmente afetada devido à pobreza, encarceramento em massa, mortalidade infantil, acesso limitado a cuidados de saúde e condições relacionadas à saúde, incluindo doenças cardíacas, renais, respiratórias, além do racismo contínuo e a discriminação. A falta de dados descritos e acessíveis sobre a composição racial e étnica dos infectados, pode resultar em grupos historicamente marginalizados com uma carga ainda maior de doenças e um impacto social desproporcionalmente maior (LAURENCIN et al., 2021).

Os dados atuais sobre o coronavírus e seu efeito em diferentes populações, são agregados de toda a população e não refletem as realidades das populações preta e hispânica. Os pretos mantêm maiores cargas de doenças devido a taxas de pobreza mais altas, acesso limitado a cuidados de saúde e aumento de empregos nas indústrias de serviços.

Para testes futuros com ML, neste tipo de experimento podem ser adicionadas mais informações dependendo dos dados obtidos pelo DATASUS, como mortes por idade (já que isso correlaciona com a demência, pois aumenta a suscetibilidade à doenças), fumantes do sexo feminino e masculino, leitos disponíveis em hospitais, número total de testes realizados, áreas de contenção em um município, pontos de acesso e outros. Uma outra sugestão seria também buscar e aferir com dados abertos de outros países, mesmo que ainda há uma ausência de registros dos sistemas prisionais com acesso ao público. E por fim, optou-se em utilizar esses resultados alcançados pelo uso de técnicas de ML e transcrevemos para um artigo, com a finalidade de submeter em uma revista de pesquisa (Qualis Restrito) voltada especificamente para a área de saúde.

O intuito é prever a propagação dessas doenças e ajudar os responsáveis a tomarem ações emergentes. Isso iria identificar o estágio inicial da doença e ilustrar a probabilidade de prever o vírus, onde todo esse esforço ajudaria uma determinada região a diminuir e superar o COVID-19 ou qualquer outra doença. Em vista disso, a ideia é eficiente e pode ser usada durante qualquer tipo de emergência ou pandemia no futuro.

6

Conclusões

O cenário que envolve o encarceramento de pessoas e a pandemia do COVID-19 merece atenção da comunidade científica. Pelo que foi apresentado e pela motivação deste trabalho, pode-se dizer que atualmente há esforços crescentes no estudo dos dados da doença.

Foi fundamental a realização de um Mapeamento Sistemático de pesquisas com dados abertos sobre o COVID-19 no sistema prisional. Portanto, o processo seguido para a realização da revisão consistiu na criação de questões de pesquisa que nortearam o desenvolvimento da busca e seleção dos artigos e sua respectiva análise. Assim, foram encontradas 125 publicações em 7 bases de dados diferentes (*Scopus, Web of Science, El Compendex, ScienceDirect, Pubmed Central, SciELO e Oxford University Press*) e outras 8 publicações após o uso da técnica *Forward Snowballing* (FS). Após a seleção dessas publicações, os critérios foram aplicados, reduzindo-se a 37 artigos considerados relevantes para o objetivo proposto. Consequentemente, a partir da análise de dos estudos selecionados, encontramos estatísticas, tendências e apontamentos a serem desenvolvidos na área de uso de dados do COVID-19. No entanto, também existem algumas dificuldades em lidar com dados que não estão preparados ou estruturados para serem utilizados e cujo respectivo acesso depende de alguma permissão.

Constatou-se que os dados de Saúde Pública e Criminologia são atualmente as áreas mais utilizadas por profissionais e pesquisadores quando se trata de dados de COVID-19. No entanto, há pouca discussão pública sobre tecnologia e metodologias para tratar dessas informações.

Os dados são provenientes de registros médicos, criminais e são parcialmente restritos, o que significa fazer vários esforços para produzir dados apenas para uso privado e secundário. Quanto ao tipo de uso dos dados, nos estudos selecionados, visualizamos a análise descritiva das informações coletadas, muito mais do que o armazenamento e atualização dos dados. Por fim, uma análise detalhada dos países nos estudos selecionados, os Estados Unidos (EUA) é o que tem contribuído para o sistema prisional com dados do COVID-19 (19 artigos). Depois dos EUA,

o Reino Unido tem 3 artigos, Espanha (1 artigo), Itália (1 artigo) e Suíça (1 artigo). Destacamos países como Brasil e Etiópia com três artigos, e Honduras, Guatemala, El Salvador, África do Sul, Namíbia, Zimbábue, Nigéria e Chile com apenas um artigo.

Os resultados discutidos neste mapeamento é um primeiro esforço de fechar a lacuna da ausência de estratégias eficazes de análise de dados entre o estado atual dos registros prisionais. Em vista disso, um novo cenário no qual os dados do COVID-19 estariam disponíveis para as partes interessadas, sem o risco de violação do segredo legítimo, tanto por indivíduos como por organizações.

O avanço na capacidade de gerar e processar dados do COVID-19 de forma digital, sugere que a visualização de dados será cada vez mais relevante. Isso motivou a ideia de desenvolver um painel visual com as informações de casos confirmados da doença, nos municípios do estado da Bahia, como uma representação da pandemia por meio de atributos visuais de fácil compreensão (como gráficos, histogramas e mapas). Esse painel visual (plataforma *web*) possibilitou que as pessoas visualizassem dados do COVID-19 no sistema prisional de uma maneira diferente, reunindo não apenas as informações, mas também detalhes precisos e com novos padrões.

Além disso, foram aplicados critérios heurísticos adequados para a avaliação de sistemas de visualização, contribuindo com uma diminuição de erros de interpretação ou significados gerados por visualizações mal desenvolvidas. Diante disso, a importância de uma lista de verificação de heurísticas de usabilidade, com o intuito de avaliar sistemas de visualização de informações, garante uma alta qualidade em sistemas de dados digitais desenvolvidos para a área da saúde.

Outra contribuição desta dissertação foi o uso de algoritmos de *Machine Learning* nos dados da área de saúde, especificamente o COVID-19, do sistema prisional do estado da Bahia, nos quais permitiram identificar o estágio inicial e prever a propagação dessas doenças com uma melhor análise dos dados coletados. Esses algoritmos oferecem informações que podem auxiliar na tomada de decisões para combater o contágio, bem como os resultados da previsão podem ser usados para decidir ações instantâneas.

Essa pesquisa serve também como fonte de consulta para métodos de redução da mortalidade por doenças cardíacas e respiratórias. Pelos resultados apresentados, pôde-se notar que essa análise e visualização dos dados precisa de pouco processamento e, portanto, não é necessário tanto esforço e investimento em infraestrutura computacional para a implantação de tais técnicas em ambientes reais e sociais da área médica.

6.1 Trabalhos futuros

Após os resultados dos testes de usabilidade e da avaliação heurística, serão feitas correções durante o desenvolvimento do *dashboard* por meio das contribuições obtidas pelos

participantes, com o intuito de aperfeiçoar a aplicação. Também serão adicionadas mais informações conforme as bases exportadas pelo DATASUS, além de buscar e aferir com os dados abertos de outras regiões, tanto em conjunto de registros prisionais de outros estados do Brasil, assim como informações carcerárias obtidas de outros países que tenham acesso público.

Referências

- AGARWAL, E. et al. Covid-19 dashboard: Supervised learning ml model based covid-19 prediction analysis and identifying early-stage symptoms. In: . [S.l.: s.n.], 2021. p. 1–9. Citado 2 vezes nas páginas [62](#) e [63](#).
- AGGARWAL, S. *Flask Framework Cookbook - Second Edition*. [S.l.: s.n.], 2019. ISBN 1789951291. Citado na página [42](#).
- ARRAIS, P. et al. Códigos da classificação internacional de doenças como rastreadores de eventos adversos a medicamentos. *Medicina (Ribeirão Preto)*, v. 54, 12 2021. Citado na página [42](#).
- BA, K.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. v. 2, 01 2007. Citado 3 vezes nas páginas [15](#), [17](#) e [31](#).
- BASILI, V.; WEISS, D. A methodology for collecting valid software engineering data. *IEEE Trans. Software Eng.*, v. 10, p. 728–738, 11 1984. Citado 2 vezes nas páginas [6](#) e [17](#).
- BHUTIA, S.; PATRA, B. Regression analysis and prediction of the covid-19. In: _____. [S.l.: s.n.], 2022. p. 349–361. ISBN 978-981-16-5684-2. Citado na página [66](#).
- BYRNE, J. et al. An imperfect storm: Identifying the root causes of covid-19 outbreaks in the world’s largest corrections systems. In: _____. [S.l.: s.n.], 2021. p. 30–78. Citado 6 vezes nas páginas [23](#), [25](#), [26](#), [28](#), [29](#) e [32](#).
- CALSTER, B. V. et al. Predictive analytics in health care: how can we know it works? *Journal of the American Medical Informatics Association : JAMIA*, v. 26, 08 2019. Citado na página [27](#).
- CANINO, A. Deconstructing google dataset search. *Public Services Quarterly*, v. 15, p. 248–255, 07 2019. Citado na página [63](#).
- CARNEIRO, G. et al. *Perspectives and Challenges in the Analysis of Prison Systems Data: A Systematic Mapping*. 2021. Citado na página [13](#).
- CARVALHO, S.; SANTOS, I.; SANTOS, A. B. dos. *The pandemic in prison: interventions and overisolation*. 2020. Citado 6 vezes nas páginas [23](#), [25](#), [26](#), [28](#), [29](#) e [32](#).
- CASTILLO, O.; MELIN, P. Forecasting of covid-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic. *Chaos, Solitons Fractals*, v. 140, p. 110242, 08 2020. Citado na página [63](#).
- CHI, E.; RIEDL, J. An operator interaction framework for visualization systems. *Information Visualization, IEEE Symposium on*, v. 0, p. 63, 08 1998. Citado 3 vezes nas páginas [5](#), [39](#) e [40](#).
- CHICCO, D.; WARRENS, M.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, v. 7, p. e623, 07 2021. Citado na página [73](#).
- CHITYALA, R.; PUDIPEDDI, S. Convolutional neural network. In: _____. [S.l.: s.n.], 2020. p. 265–273. ISBN 9780429243370. Citado na página [15](#).

CHOPRA, D.; KHURANA, R. Support vector machine. In: _____. [S.l.: s.n.], 2023. p. 58–73. ISBN 9789815124422. Citado na página 68.

COVID-19 in Correctional Facilities. Connecticut State Department of Correction, 2023. Disponível em: <<https://data.ct.gov/Public-Safety/COVID-19-in-Correctional-Facilities/6t8i-du3u>>. Citado na página 64.

COWAN, I.; MACDONALD, S.; FLOYD, R. Measuring and managing radiologist workload: Measuring radiologist reporting times using data from a radiology information system. *Journal of medical imaging and radiation oncology*, v. 57, p. 558–566, 10 2013. Citado na página 15.

CRISPIM, J. et al. Impact and trend of covid-19 in the brazilian prison system: An ecological study [impacto e tendência da covid-19 no sistema penitenciário do brasil: um estudo ecológico]. *Ciencia e Saude Coletiva*, Associacao Brasileira de Pos - Graduacao em Saude Coletiva, v. 26, n. 1, p. 169–178, 2021. ISSN 14138123. Cited By 0. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

DBEAVER Community. In: . [s.n.], 2022. Disponível em: <<https://dbeaver.io/>>. Citado na página 40.

DOWDING, D.; MERRILL, J. The development of heuristics for evaluation of dashboard visualizations. *Applied Clinical Informatics*, v. 09, p. 511–518, 07 2018. Citado 2 vezes nas páginas 14 e 39.

DYBÅ, T.; DINGSØYR, T. Empirical studies of agile software development: A systematic review. *Information and Software Technology*, v. 50, p. 833–859, 08 2008. Citado na página 20.

ERENDOR, M. E.; SUNGURAY, C. The united nations and promoting sustainable development. p. 1727–1743, 04 2019. Citado na página 30.

EZEIBE, C. et al. From protection to repression: State containment of covid-19 pandemic and human rights violations in nigeria. *Victims Offenders*, p. 1–28, 06 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.

FERDIK, F.; FROGGE, G.; DOGGETT, S. “it’s like the zombie apocalypse here”: Correctional officer perspectives on the deleterious effects of the covid-19 pandemic. *Crime Delinquency*, p. 001112872211310, 11 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.

FRANCO-PAREDES, C. et al. Personal view decarceration and community re-entry in the covid-19 era. *The Lancet Infectious Diseases*, v. 21, 09 2020. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

FRANCO-PAREDES, C. et al. Covid-19 in jails and prisons: A neglected infection in a marginalized population. *PLOS Neglected Tropical Diseases*, Public Library of Science, v. 14, n. 6, p. 1–4, 06 2020. Disponível em: <<https://doi.org/10.1371/journal.pntd.0008409>>. Citado na página 12.

FURLAN, L.; CARAMELLI, B. The regrettable story of the “covid kit” and the “early treatment of covid-19” in brazil. *The Lancet Regional Health - Americas*, v. 4, p. 100089, 2021. ISSN 2667-193X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2667193X21000855>>. Citado na página 74.

- G, R. M. T. et al. Data visualization techniques for real-time information — a custom and dynamic dashboard for analyzing surveys' results. In: . [S.l.: s.n.], 2018. p. 1–7. Citado na página 38.
- GALLEGUILLOS, S. et al. The covid-19 pandemic and probation in chile: Remote supervision and regional differences. *International Criminology*, v. 2, p. 3, 03 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- GONCALVES, L. et al. Covid-19 inmate risk appraisal (cira): development and validation of a screening tool to assess covid-19 vulnerability in prisons. *Swiss medical weekly: official journal of the Swiss Society of Infectious Diseases, the Swiss Society of Internal Medicine, the Swiss Society of Pneumology*, v. 151, p. w20471, 02 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- GRUBESIC, T. et al. Geodemographic insights on the covid-19 pandemic in the state of wisconsin and the role of risky facilities. *GeoJournal*, 09 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- GUO, T. et al. Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (covid-19). *JAMA Cardiology*, 03 2020. Citado na página 13.
- GUO, T. et al. Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (covid-19). *JAMA Cardiology*, 03 2020. Citado na página 14.
- HACK, J.; SANTOS, J. Influência do design emocional na interação homem/computador | the influence of emotional design in interaction between humans and the computer. *Liinc em Revista*, v. 6, 09 2010. Citado na página 53.
- HAGAN, L. et al. Mass testing for sars-cov-2 in 16 prisons and jails — six jurisdictions, united states, april–may 2020. *MMWR. Morbidity and Mortality Weekly Report*, v. 69, 08 2020. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- Haiduwa, T.; Hashiyana, V. An android-bases application for convenient visitation at windhoek correctional facility. In: . [S.l.: s.n.], 2020. Citado 6 vezes nas páginas 25, 26, 27, 28, 29 e 32.
- HAMZAH, F. A. B. et al. Coronatracker: World-wide covid-19 outbreak data analysis and prediction. 03 2020. Citado na página 63.
- HASIJA, Y.; CHAKRABORTY, R. Python for data visualization. In: _____. [S.l.: s.n.], 2021. p. 91–122. ISBN 9781003090113. Citado na página 42.
- HERNANDEZ, B.; ROJAS, G. Automatic code generation of data visualization for structural health monitoring. *IEEE Latin America Transactions*, v. 20, p. 1041–1050, 07 2021. Citado na página 37.
- HOUT, M.-C.; BIGLAND, C.; MARINIELLO, T. A legal-realist assessment of the zimbabwean correctional system response to covid-19 during state disaster measures. *International journal of prisoner health*, ahead-of-print, 04 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- IRVINE, M. et al. Modeling covid-19 and its impacts on u.s. immigration and customs enforcement (ice) detention facilities, 2020. *Journal of Urban Health*, v. 97, 05 2020. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

- JAHNS, V. Information visualization: perception for design by colin ware. *ACM SIGSOFT Software Engineering Notes*, v. 39, p. 43–44, 03 2014. Citado 2 vezes nas páginas 39 e 54.
- JUMBE, V. et al. A multi-stakeholder situation assessment of covid-19 disease preparedness and mitigation measures in a large prison complex in malawi. *International journal of prisoner health*, ahead-of-print, 02 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- KUMAR, P. et al. Analytical statistics techniques of classification and regression in machine learning. In: _____. [S.l.: s.n.], 2020. ISBN 978-1-83968-318-3. Citado na página 65.
- KUPFERSCHMIDT, K.; WADMAN, M. End of covid-19 emergencies sparks debate. *Science*, v. 380, n. 6645, p. 566–7, 2023. Citado na página 63.
- LAURENCIN, C. et al. Changes in covid-19-associated deaths during a year among blacks and hispanics compared to whites in the state of connecticut. *Journal of Racial and Ethnic Health Disparities*, v. 9, 09 2021. Citado na página 76.
- LEMASTERS, K. et al. Covid-19 cases and testing in 53 prison systems. *Health and Justice*, v. 8, 12 2020. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- LI, Z.; LI, L. A novel method for lung masses detection and location based on deep learning. In: . [S.l.: s.n.], 2017. p. 963–969. Citado na página 15.
- MALEKI, M. et al. Time series modelling to forecast the confirmed and recovered cases of covid-19. *Travel Medicine and Infectious Disease*, v. 37, p. 101742, 2020. ISSN 1477-8939. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1477893920302210>>. Citado na página 62.
- MANDERSON, L. et al. Containing covid-19 and the social costs on human rights in african countries. *Humanities and Social Sciences Communications*, v. 9, 10 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- MAO, L. et al. Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurology*, v. 77, n. 6, p. 683–690, 06 2020. ISSN 2168-6149. Disponível em: <<https://doi.org/10.1001/jamaneurol.2020.1127>>. Citado na página 12.
- MARTIN-HOWARD, S. Covid-19's impact on black, female correctional officers and justice-involved individuals at rikers island jail. *Crime Delinquency*, v. 68, p. 001112872110736, 02 2022. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- MAYCOCK, M. Covid-19 has caused a dramatic change to prison life. *British Journal of Criminology*, 04 2021. Citado 6 vezes nas páginas 19, 25, 26, 28, 29 e 32.
- MEDEIROS, M. A.; CYBIS, W. d. A. Método de avaliação de usabilidade de software a partir da satisfação de usuários e da aplicação de quesitos da norma iso 9241. *IHC'2000–III Workshop sobre Fatores Humanos em Sistemas de Computação*, p. 93–101, 2000. Citado na página 53.
- MEKONNEN, B. et al. Preparedness and readiness against covid-19 pandemic in prison institutions and detention centers in southwest ethiopia. *International Journal of General Medicine*, 08 2021. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- MISIR, P. Epidemiology of covid-19. In: _____. [S.l.: s.n.], 2022. p. 19–43. ISBN 978-3-030-88765-0. Citado na página 75.

- MOHER, D. et al. Moher d, liberati a, tetzlaff j, altman dg, group ppreferred reporting items for systematic reviews and meta-analyses: the prisma statement. *plos med* 6: e1000097. *Open medicine : a peer-reviewed, independent, open-access journal*, v. 3, p. e123–30, 07 2009. Citado na página 21.
- MUNTINGH, L. Africa, prisons and covid-19. *Journal of Human Rights Practice*, v. 12, 08 2020. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- NIELSEN, J. Usability inspection methods. In: . [S.l.: s.n.], 1995. v. 25, p. 377–378. Citado 2 vezes nas páginas 53 e 54.
- NOSRATI, E. et al. Incarceration and mortality in the united states. *SSM - Population Health*, v. 15, p. 100827, 06 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- NOVISKY, M.; NARVEY, C.; SEMENZA, D. Institutional responses to the covid-19 pandemic in american prisons. In: _____. [S.l.: s.n.], 2021. p. 425–442. ISBN 9781003189572. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- NOVISKY, M. et al. Incarceration as a fundamental social cause of health inequalities: Jails, prisons and vulnerability to covid-19. *The British Journal of Criminology*, 04 2021. Citado 6 vezes nas páginas 19, 25, 26, 28, 29 e 32.
- NOWOTNY, K.; SEIDE, K.; BRINKLEY-RUBINSTEIN, L. Risk of covid-19 infection among prison staff in the united states. *BMC Public Health*, v. 21, 06 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- NUNES, L. d. S.; MORENO, E. D.; CARNEIRO, G. d. F. A review of covid-19 data usage in correctional institutions from developing countries. *IEEE Transactions on Computational Social Systems*, p. 1–15, 2023. Citado na página 36.
- PARK, K.; MEAGHER, T. *A State-By-State Look at 15 Months of Coronavirus in Prisons*. 2021. Disponível em: <<https://www.themarshallproject.org/2020/05/01/a-state-by-state-look-at-coronavirus-in-prisons>>. Citado na página 64.
- PETROPOULOS, F.; MAKRIDAKIS, S. Forecasting the novel coronavirus covid-19. *PLOS ONE*, v. 15, p. e0231236, 03 2020. Citado na página 63.
- PITTS, W.; INKPEN, C. Assessing the effects of covid-19 in prisons in the northern triangle of central america. In: _____. [S.l.: s.n.], 2021. p. 220–237. ISBN 9781003189572. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- RATAN, S.; ANAND, T.; RATAN, J. Formulation of research question – stepwise approach. *Journal of Indian Association of Pediatric Surgeons*, v. 24, p. 15, 01 2019. Citado na página 18.
- REINHART, E.; CHEN, D. Association of jail decarceration and anticontagion policies with covid-19 case growth rates in us counties. *JAMA network open*, v. 4, p. e2123405, 09 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- RIBEIRO, L.; DINIZ, A. A. The brazilian penitentiary system under the threat of covid-19. *Victims Offenders*, v. 15, p. 1–25, 11 2020. Citado 6 vezes nas páginas 23, 25, 26, 28, 29 e 32.
- RUBIN et al. *Handbook of Usability Testing*. [S.l.: s.n.], 2008. Citado na página 47.

- RYCKMAN, T. et al. Outbreaks of covid-19 variants in us prisons: a mathematical modelling analysis of vaccination and reopening policies. *The Lancet Public Health*, v. 6, 08 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- SHARMA, D.; IQBAL, S. Applying decision tree algorithm classification and regression tree (cart) algorithm to gini techniques binary splits. *International Journal of Engineering and Advanced Technology*, v. 12, p. 77–81, 06 2023. Citado na página 68.
- SILVA, L. et al. Desenvolvimento de dashboards interativos utilizando ferramentas de business intelligence no ms excel para auxílio na tomada de decisão empresarial. *Revista Expressão Católica*, v. 7, p. 27, 09 2018. Citado 2 vezes nas páginas 16 e 37.
- SLÁDEK, D. Application of the random forest method on the observation dataset for visibility nowcasting. *Geofizika*, v. 40, p. 75–102, 06 2023. Citado na página 68.
- SONG, Q. et al. Using deep learning for classification of lung nodules on computed tomography images. *Journal of Healthcare Engineering*, v. 2017, p. 1–7, 08 2017. Citado na página 15.
- STONE, P. Popping the (pico) question in research and evidence-based practice. *Applied nursing research : ANR*, v. 15, p. 197–8, 09 2002. Citado na página 18.
- STRODEL, R. et al. Covid-19 vaccine prioritization of incarcerated people relative to other vulnerable groups: An analysis of state plans. *PLOS ONE*, v. 16, p. e0253208, 06 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- STUFANO, A. et al. Efficacy of the measures adopted to prevent covid-19 outbreaks in an italian correctional facility for inmates affected by chronic diseases. *Frontiers in Public Health*, v. 9, p. 694795, 07 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- SUHOMLINOVA, O. et al. Locked up while locked down: Prisoners' experiences of the covid-19 pandemic. *The British Journal of Criminology*, 06 2021. Citado 6 vezes nas páginas 19, 25, 26, 28, 29 e 32.
- TABNET/DATASUS. In: . [s.n.], 2022. Disponível em: <<https://datasus.saude.gov.br/>>. Citado na página 40.
- TABWIN/DATASUS. In: . [s.n.], 2022. Disponível em: <<http://siab.datasus.gov.br/DATASUS/index.php?area=060805&item=3>>. Citado na página 40.
- TUDELA, E. Telematic control and semi-freedom as a response to the pandemic: The spanish penitentiary system experience. In: _____. [S.l.: s.n.], 2021. p. 364–380. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.
- VAHEDI, A. et al. Applications, features and key indicators for the development of covid-19 dashboards: A systematic review study. *Informatics in medicine unlocked*, v. 30, p. 100910, 01 2022. Citado na página 38.
- VASUDEVAN, V. et al. *Disparity in the quality of COVID-19 data reporting across India*. 2020. Citado na página 34.
- VEST, N. et al. Prison population reductions and covid-19: A latent profile analysis synthesizing recent evidence from the texas state prison system. *Journal of Urban Health*, v. 98, 12 2020. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

WALLACE, D. et al. Is there a temporal relationship between covid-19 infections among prison staff, incarcerated persons and the larger community in the united states? *International Journal of Environmental Research and Public Health*, v. 18, p. 6873, 06 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

WALLACE, D. et al. The 2020 coronavirus pandemic and its corresponding data boon: Issues with pandemic-related data from criminal justice organizations. *Journal of Contemporary Criminal Justice*, p. 104398622110279, 07 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

WARD, J. et al. Covid-19 cases among employees of u.s. federal and state prisons. *American Journal of Preventive Medicine*, 02 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

WILBURN, J. et al. Covid-19 within a large prison with high number of vulnerable adults in the uk, march to june 2020: an outbreak investigation and screening event. *International Journal of Infectious Diseases*, v. 104, 01 2021. Citado 5 vezes nas páginas 25, 26, 28, 29 e 32.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. *ACM International Conference Proceeding Series*, 05 2014. Citado na página 22.

ZHANG, P. et al. Computer-aided lung cancer diagnosis approaches based on deep learning. *Journal of Computer-Aided Design Computer Graphics*, v. 30, p. 90, 01 2018. Citado na página 15.

Anexos

ANEXO A – Plano de Testes - Lista de perguntas e tarefas

Questões	Enunciado (itens)
Pergunta 1	Na sua visão, qual é o objetivo de um <i>dashboard</i> ? a) Um painel visual de informações que contém métricas e indicadores; b) Um painel que serve para acompanhar o desempenho e embasar decisões; c) Representar números relevantes para criar estratégias e alcançar os objetivos; d) Monitorar dados em tempo real e representá-los com fácil compreensão visual.
Pergunta 2	Pra você, quais são as vantagens e desvantagens de um <i>dashboard</i> ?
Tarefa 1	Navegue pelo <i>dashboard</i> e identifique o número total de mortes por COVID-19 no município de “Jequié”.
Tarefa 2	Realize uma busca de municípios e selecione por “Feira de Santana”. Encontre a CID com mais mortes na região.
Tarefa 3	Faça uma pesquisa de municípios e opte por “Ilhéus”. Encontre o mês com mais mortes na região, além do nome da CID, sua porcentagem e o número total de mortes (incluindo todas as CIDs). No final busque pela descrição da CID.
Tarefa 4	Agora, na lista de municípios, escolha a opção “Salvador”. Encontre a CID com mais mortes na região e a sua descrição. Após isso, descubra o total de mortes no município.
Tarefa 5	Sabendo que a CID com mais mortes na Bahia é o B342 (COVID-19), descubra a segunda CID com mais mortes no estado.
Pergunta 3	Qual é a sua opinião sobre os mapas? Ficou... a) Ruim b) Bom c) Muito bom d) Ótimo e) Excelente
Pergunta 4	Comente sobre quaisquer pontos que não foram comentados ou atividades que poderiam ser inseridas no <i>dashboard</i> .
Pergunta 5	Quais são as suas sugestões para melhorar o <i>dashboard</i> ?

ANEXO B – Plano de Testes – Termo de Consentimento Livre e Esclarecido (TCLE)

O objetivo desta pesquisa será propor um *dashboard* para auxiliar e apoiar o processo de dados do COVID-19 no estado da Bahia com foco na usabilidade. Será necessário aplicar métodos de avaliação para coletar dados, a fim de identificar a aplicação da ferramenta pelos usuários. Será realizado um teste de usabilidade e os participantes deverão responder um questionário e participar de uma entrevista. Para a aplicação desses métodos haverá um esclarecimento de cada participante, sobre os objetivos e procedimentos da pesquisa, manifestando sua concordância por meio da assinatura deste TCLE.

Os benefícios relacionados aos usuários, participantes nesta pesquisa, estão relacionados ao levantamento de dados sobre a eficiência, eficácia e satisfação da ferramenta e as demandas relacionadas aos esforços cognitivos. Assim serão realizados ajustes na ferramenta a partir das opiniões, das observações e das dificuldades expressadas pelos mesmos.

Não há risco financeiro ao participante, o sistema é disponibilizado gratuitamente pelos pesquisadores para o teste. Suas contribuições serão muito importantes para o desenvolvimento do *dashboard*, para assim compreendermos melhor o seu funcionamento, além da possibilidade de aperfeiçoá-lo ainda mais.

Logo, declaro que a minha participação é voluntária e sem custos financeiros, podendo me recusar a participar e isso não ocasionará em qualquer penalidade, prejuízo ou constrangimento a minha pessoa. Os dados coletados ficam de posse dos pesquisadores responsáveis, restringindo qualquer acesso ou identificação. Fui devidamente informado e esclarecido sobre os objetivos desta pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios envolvidos na minha participação. Tendo em vista os itens acima apresentados de forma esclarecida, manifesto o meu consentimento e concordo em participar da pesquisa, ficando este termo acordado entre as partes.

Desejo em participar voluntariamente da pesquisa, coordenado pelos professores Glauco de Figueiredo Carneiro e Edward David Moreno Ordonez, e pelo pesquisador Lucas dos Santos Nunes”.

Eu concordo _____

ANEXO C – Avaliação Heurística – Termo de Consentimento Livre e Esclarecido – TCLE

O objetivo desta pesquisa será propor um *dashboard* para auxiliar e apoiar o processo de dados do COVID-19 e de outras doenças, no estado da Bahia com foco na usabilidade. A doença é classificada como CID (Classificação Internacional de Doenças), criada pela Organização Mundial da Saúde (OMS) em 1893. A sua vantagem é a diminuição de problemas de tradução e falhas no diagnóstico, facilitando a comunicação entre os diferentes profissionais da saúde do mundo. Além de incluir mapas geográficos, todas essas informações foram exportadas do Departamento de Informática do Sistema Único de Saúde (DATASUS).

Será necessário aplicar métodos de avaliação para coletar dados, a fim de identificar a aplicação da ferramenta pelos usuários. Será realizado um teste de usabilidade e os participantes deverão responder um questionário e participar de uma entrevista. Para a aplicação desses métodos haverá um esclarecimento de cada participante, sobre os objetivos e procedimentos da pesquisa, manifestando sua concordância por meio da assinatura deste TCLE.

Os benefícios relacionados aos usuários, participantes nesta pesquisa, estão relacionados ao levantamento de dados sobre a eficiência, eficácia e satisfação da ferramenta e as demandas relacionadas aos esforços cognitivos. Assim serão realizados ajustes na ferramenta a partir das opiniões, das observações e das dificuldades expressadas pelos mesmos.

Não há risco financeiro ao participante, o sistema é disponibilizado gratuitamente pelos pesquisadores para o teste. Suas contribuições serão muito importantes para o desenvolvimento do *dashboard*, para assim compreendermos melhor o seu funcionamento, além da possibilidade de aperfeiçoá-lo ainda mais.

Logo, declaro que a minha participação é voluntária e sem custos financeiros, podendo me recusar a participar e isso não ocasionará em qualquer penalidade, prejuízo ou constrangimento a minha pessoa. Os dados coletados ficam de posse dos pesquisadores responsáveis, restringindo qualquer acesso ou identificação. Fui devidamente informado e esclarecido sobre os objetivos desta pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios envolvidos na minha participação. Tendo em vista os itens acima apresentados de forma esclarecida, manifesto o meu consentimento e concordo em participar da pesquisa, ficando este termo acordado entre as partes.

Desejo em participar voluntariamente da pesquisa, coordenado pelos professores Glauco de Figueiredo Carneiro e Edward David Moreno Ordonez, e pelo pesquisador Lucas dos Santos

Nunes”.

Eu concordo _____

ANEXO D – Avaliação Heurística - Lista de perguntas e tarefas

Questões	Enunciado (itens)
Pergunta 1	<p>Na sua visão, qual é o objetivo de um <i>dashboard</i>?</p> <p>a) Um painel visual de informações que contém métricas e indicadores;</p> <p>b) Um painel que serve para acompanhar o desempenho e embasar decisões;</p> <p>c) Representar números relevantes para criar estratégias e alcançar os objetivos;</p> <p>d) Monitorar dados em tempo real e representá-los com fácil compreensão visual.</p>
Pergunta 2	<p>Pra você, quais são as vantagens e desvantagens de um <i>dashboard</i>?</p>
Tarefa 1: Visibilidade do status do sistema	<p>a) Cada tela tem um título ou cabeçalho que descreve seu conteúdo?</p> <p>b) Existe um esquema de design de ícone consistente e estilístico tratamento em todo o sistema?</p> <p>c) Há um <i>feedback</i> visual nos menus ou caixas de diálogo sobre quais opções são selecionáveis?</p> <p>d) A terminologia de nomeação do menu é consistente com o domínio de tarefa dos usuários?</p> <p>e) O sistema oferece visibilidade: ou seja, olhando, o usuário pode informar o estado do sistema e as alternativas para ação?</p>
Tarefa 2: Correspondência entre o sistema e o mundo real	<p>a) Os ícones são concretos e familiares?</p> <p>b) Os cabeçalhos e subseções de cada tela estão ordenados da maneira mais lógica?</p> <p>c) Existe uma sequência natural para as opções de menu para um item de dados?</p> <p>d) As cores selecionadas correspondem às expectativas comuns sobre códigos de cores?</p> <p>e) As palavras/conceitos e frases usadas em cada tela são familiares aos usuários?</p>

Tarefa 3: Controle e liberdade do usuário	<ul style="list-style-type: none"> a) Existe uma saída clara em cada tela? b) Todas as telas estão acessíveis no sistema? c) Os usuários têm a opção de clicar nos itens de menu com o mouse ou usar um atalho de teclado? d) Os usuários podem avançar e retroceder facilmente entre as telas?
Tarefa 4: Consistência e padrões	<ul style="list-style-type: none"> a) Os padrões de formatação foram seguidos de forma consistente em todas as telas do sistema? b) Cada janela tem um título? c) Os nomes são consistentes, tanto dentro de cada guia quanto em todo o sistema, em estilo gramatical e terminologia?
Tarefa 5: Reconhecimento em vez de lembrança	<ul style="list-style-type: none"> a) Os avisos, as pistas e as mensagens visíveis são encontradas com facilidade na tela? b) Os itens foram agrupados em zonas lógicas e os cabeçalhos foram usados para distinguir entre as zonas? c) O realce de cor é usado para chamar a atenção do usuário? d) O código de cores é consistente em todo o sistema?
Tarefa 6: Flexibilidade e eficiência de uso	<ul style="list-style-type: none"> a) A navegação entre as telas é simples e visível? b) Se as listas de menu forem curtas, os usuários podem selecionar um item movendo o cursor? c) Nos menus, os usuários têm a opção de clicar diretamente em um item do menu ou usar um atalho do teclado?
Tarefa 7: Design estético e minimalista	<ul style="list-style-type: none"> a) Os objetos grandes, as fontes em negrito e as áreas simples foram usadas para distinguir seções? b) Os rótulos dos campos são breves, familiares e descritivos? c) O layout visual está bem desenhado?
Pergunta 3	Comente sobre quaisquer pontos que não foram comentados ou atividades que poderiam ser inseridas no <i>dashboard</i> .
Pergunta 4	Quais são as suas sugestões para melhorar o <i>dashboard</i> ?

ANEXO E – *Dashboard* - Classificação das CIDs

Capítulo	Grupo(s)	Descrição
1	De A até B	Engloba doenças infecciosas causadas por bactérias e fungos.
2	De C até D	Classifica neoplasias e tumores.
3	De D50 até D89	Rotuladas a doenças do sangue.
4	Iniciam com a letra E	Referentes aos distúrbios hormonais e de metabolismo, além de doenças como diabetes.
5	Iniciam com a letra F	São doenças mentais e comportamentais, como depressão e esquizofrenia.
6	Iniciam com a letra G	São todos os distúrbios do sistema nervoso, como atrofia muscular, paralisia cerebral e doenças inflamatórias.
7	Iniciam com a letra H	São referentes a doenças ligadas à visão humana. Uma delas são a conjuntivite, o glaucoma e a catarata.
8	De H60 até H95	São todas as doenças que atingem os ouvidos, conhecidas como otites.
9	Iniciam com a letra I	Referentes a problemas cardíacos e no sistema circulatório.
10	Iniciam com a letra J	Estão todas as doenças que atingem o sistema respiratório, como gripe, rinite e infecções pulmonares.
11	Iniciam com a letra K	São os problemas bucais, como cáries e gengivites, além de determinados transtornos do aparelho digestivo.
12	Iniciam com a letra L	São todas as doenças de pele, como por exemplo, a dermatite.
13	Iniciam com a letra M	São enfermidades que acometem o tecido conjuntivo, ósseo e muscular, como artroses, artrites e meniscos.
14	Iniciam com a letra N	Que reúnem problemas nos órgãos genitais do sistema urinário, além das doenças que atingem as mamas.

15	Iniciam com a letra O	Referentes a problemas na gravidez e no parto, assim como todos os tipos de complicações que podem acontecer durante a gestação.
16	Iniciam com a letra P	São todos os tipos de doenças e complicações que acontecem próximo ao nascimento de um bebê, como hemorragias, infecções e traumas.
17	Iniciam com a letra Q	São alterações cromossômicas ou anomalias que ocorrem em qualquer parte do corpo humano de uma criança.
18	Iniciam com a letra R	São todos os exames clínicos e laboratoriais que não são classificados em nenhum outro capítulo, ou seja, onde apresentam sintomas que não foram diagnosticados, apresentando resultados anormais.
19	De S até T	São todas as consequências e fatores que podem gerar lesões como queimaduras ou machucados pelo corpo humano.
20	De V até X	São as causas de morte superficiais motivadas por acidentes de trânsito, quedas, afogamento e outras causas.
21	Iniciam com a letra Z	Engloba variações que influenciam o estado de saúde, como as doenças transmissíveis.
22	Iniciam com a letra U	Referentes as novas doenças que foram descobertas recentemente, como aquelas que são causadas por bactérias resistentes a antibióticos.