



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO: MESTRADO PROFISSIONAL EM
INFORMÁTICA EM SAÚDE

Lucas Postal

***Dashboard* de doenças cerebrovasculares e atualização do algoritmo de
análise preditiva: uma proposta para o e-SUS AB**

FLORIANÓPOLIS

2019

Lucas Postal

***Dashboard* de doenças cerebrovasculares e atualização do algoritmo de
análise preditiva: uma proposta para o e-SUS AB**

Dissertação submetida ao Programa de Mestrado Profissional em Informática em Saúde da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Informática em Saúde.

Orientador: Prof. Dr^a. Grace Teresinha Marcon Dal Sasso

Florianópolis

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Postal, Lucas

Dashboard de doenças cerebrovasculares e atualização do algoritmo de análise preditiva: uma proposta para o e-SUS AB / Lucas Postal ; orientadora, Grace Teresinha Marcon Dal Sasso, 2019.
79 p.

Dissertação (mestrado profissional) - Universidade Federal de Santa Catarina, Centro de Ciências da Saúde, Programa de Pós-Graduação em Informática em Saúde, Florianópolis, 2019.

Inclui referências.

1. Informática em Saúde. 2. API REST. 3. Análise Preditiva. 4. Regressão Logística. 5. Doenças Cerebrovasculares. I. Dal Sasso, Grace Teresinha Marcon. II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Informática em Saúde. III. Título.

Lucas Postal

***Dashboard* de doenças cerebrovasculares e atualização do algoritmo de análise preditiva: uma proposta para o e-SUS AB**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof.(a) Grace Teresinha Marcon Dal Sasso, Dr(a).
Universidade Federal de Santa Catarina

Prof. Raul Sidnei Wazlawick, Dr.
Universidade Federal de Santa Catarina

Prof. Paulino Artur Ferreira de Sousa, Dr. (via Webconferência)
Escola Superior de Enfermagem do Porto

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Informática em Saúde.

Prof. Dr.(a) Grace Teresinha Marcon Dal Sasso
Coordenador(a) do Programa

Prof. Dr.(a) Grace Teresinha Marcon Dal Sasso
Orientador(a)

Florianópolis, 22 de agosto de 2019.

Este trabalho é dedicado aos meus colegas de classe e aos meus queridos pais.

AGRADECIMENTOS

Primeiramente, um agradecimento especial para a minha família, meu pai, minha mãe e minha irmã, que sempre acreditaram e estiveram ao meu lado durante esta caminhada me apoiando e me dando forças sempre que precisei.

Um agradecimento especial para minha namorada, que acompanhou de perto esta trajetória e me apoiou, principalmente nas horas mais conturbadas.

Quero agradecer a minha orientadora por guiar-me para que essa conquista se concretizasse.

A todos os meus amigos, sem o apoio de vocês essa caminhada seria muito mais árdua.

Aos meus colegas de mestrado, pela convivência e amizade, das quais jamais esquecerei.

Quero agradecer ao Laboratório Bridge por permitir a utilização dos equipamentos, essenciais para o desenvolvimento da dissertação.

Por fim, quero agradecer a todos que de uma forma ou de outra colaboraram positivamente para o desenvolvimento desta dissertação.

RESUMO

A e-Saúde é a aplicação das Tecnologias de Informação e Comunicação à Saúde. Tem por objetivo qualificar e expandir o acesso à atenção à saúde através do uso das Tecnologias de Informação e Comunicação. O e-SUS AB PEC é o principal sistema da estratégia ministerial e necessita se comunicar com outros sistemas. O *Dashboard* de Doenças Cerebrovasculares (DDCV) é uma ferramenta cuja principal finalidade é a análise preditiva, que possibilita verificar a probabilidade de um cidadão desenvolver doenças cerebrovasculares a partir das suas condições de saúde. As doenças cerebrovasculares são a segunda maior causa de mortes no mundo, inclusive no Brasil, ficando atrás apenas das doenças cardíacas isquêmicas. Atualmente, para realizar a predição, a base de dados do *dashboard* é construída com dados coletados no sistema e-SUS AB PEC. Os processos de importação dos dados no DDCV e de análise preditiva são realizados manualmente, necessitando conhecimentos técnicos do operador do sistema. Portanto, este estudo tem o objetivo de implementar uma API para integração entre o e-SUS AB PEC e o DDCV. Também almeja incluir novos fatores de risco, por exemplo, hábitos alimentares e grau de escolaridade, ao conjunto de variáveis utilizado na análise preditiva que determina probabilisticamente a chance de um cidadão desenvolver doenças cerebrovasculares a partir de um método de regressão logística otimizada por um algoritmo, gradiente descendente estocástico. A predição utilizou dados fictícios oriundos do e-SUS AB PEC por intermédio do Layout e-SUS Atenção Básica de Dados e Interface e os resultados da foram analisados mediante estatística descritiva de frequência absoluta e relativa, média, mediana, moda, desvio padrão e valores mínimo e máximo. Para a construção da API foi utilizado o padrão arquitetural REST juntamente com os métodos de comunicação conforme protocolo HTTP. Para o desenvolvimento deste estudo, foram utilizados os seis princípios e fundamentos de *Design Science Research Methodology*. A predição utilizou a técnica de regressão logística e foi otimizada por um algoritmo de análise preditiva. Assim, permitiu estimar a probabilidade associada à ocorrência de determinado evento de doenças cerebrovasculares por meio de um conjunto de variáveis independentes. A implementação da API REST permitiu a comunicação automatizada entre os sistemas, provendo agilidade e melhora da qualidade dos processos por meio de uma integração leve, ágil e padronizada. A inclusão das novas variáveis aprimorou o algoritmo de análise preditiva, resultando no aumento da probabilidade média de um cidadão desenvolver uma doença cerebrovascular em 45,32%, passando de 47,59% para 69,16%. As variáveis relacionadas ao hábito alimentar apresentaram as maiores probabilidades médias, tanto para o índice mínimo, quanto para o máximo. A variável escolaridade demonstrou que quanto maior o grau de escolaridade menor a probabilidade de desenvolver doença cerebrovascular.

Palavras-chave: API REST. Análise Preditiva. Machine Learning. Regressão Logística. Doenças Cerebrovasculares. Atenção Primária.

ABSTRACT

e-Health is the application of Information and Communication Technologies to Health. Its aims are to qualify and expand the access for health care through the use of Information and Communication Technologies. The e-SUS AB PEC is the main system of Ministerial Strategy that needs to communicate with other systems. The Dashboard of Cerebrovascular Diseases (DDCV) is a tool whose main objective is to perform predictive analysis, which allows to verify the possibility of a citizen to develop cerebrovascular diseases from their health conditions. Cerebrovascular disorders are the second leading cause of death worldwide, including Brazil, behind only ischemic heart disease. Currently, to perform the prediction, the dashboard database is built with data collected from the e-SUS AB PEC. The DDCV data import process and predictive analysis are performed manually, requiring technical knowledge from the system operator. Therefore, this study aims to implement an API for integration between e-SUS AB PEC and DDCV systems. It also aims to include new risk factors, for instance, eating habits and educational level, to the set of variables used in the predictive analysis that probably determines the likelihood of a citizen developing cerebrovascular disorders from a logistic regression method optimized by an algorithm, the stochastic descending gradient. The prediction used empirical data from e-SUS AB PEC through the e-SUS Primary Care Data and Interface Layout and the results were analyzed using descriptive statistics of absolute and relative frequency, mean, median, mode, standard deviation and minimum and maximum values. For the construction of the REST API, architectural standard was used together with the communication methods according to HTTP protocol. The prediction used the logistic regression technique and was optimized by a predictive analysis algorithm. Thus, it was possible to estimate the probability associated with the occurrence of a certain event of cerebrovascular diseases through a set of independent variables. The implementation of the REST API enabled automated communication between systems, providing agility and process quality improvement through a lightweight, agile and standardized integration. The inclusion of the new variables improved the predictive analysis algorithm resulting in an increase the average probability of a citizen developing cerebrovascular disorder by 45.32%, from 47.59% to 69.16%. The variables related to eating habits had the highest average probabilities, both for the minimum and maximum indexes. The education variable showed that the higher the educational level the lower the likelihood of developing cerebrovascular disorder.

Keywords: REST API. Predictive Analysis. Machine Learning. Logistic Regression. Cerebrovascular Disorders. Primary Care.

LISTA DE FIGURAS

Figura 1 - Porcentagem de óbitos causados por AVC.	20
Figura 2 – Mecânica da ferramenta desenvolvida por Hammes (2018).....	33
Figura 3 – Variáveis para análise preditiva de Hammes (2018).....	34
Figura 4 – <i>Design Science Research Methodology</i> (DSRM).	37
Figura 5 – Modelo cliente e servidor.	42
Figura 6 – Dicionário de dados da tabela de histórico.	43
Figura 7 – <i>Body</i> de uma requisição ao <i>endpoint</i> FCI.	44
Figura 8 – <i>Body</i> de uma requisição ao <i>endpoint</i> FCD.....	45
Figura 9 – <i>Body</i> de uma requisição ao <i>endpoint</i> FAI.....	46
Figura 10 – <i>Body</i> de uma requisição ao <i>endpoint</i> FVD.....	47
Figura 11 – <i>Body</i> de uma requisição ao <i>endpoint</i> MCA.....	48
Figura 12 – Requisição ao <i>endpoint</i> MCA através do <i>Postman</i>	49
Figura 13 – Variáveis para análise preditiva do presente estudo.....	52
Figura 14 – Normalização da variável escolaridade.....	54
Figura 15 – Mecânica do DDCV do presente estudo.	57
Figura 16 – Histograma de barras com distribuição normal para o conjunto de variáveis do presente estudo.....	60
Figura 17 – Histograma de barras com distribuição normal para o conjunto de variáveis de Hammes (2018).....	60
Figura 18 – Mapa de calor da correlação no aprendizado do presente estudo.	64
Figura 19 – Mapa de calor da correlação no teste do presente estudo.	64

LISTA DE TABELAS

Tabela 1 - Valores e classificações da variável escolaridade	53
Tabela 2 - Resultados estatísticos de Hammes (2018) e do presente estudo.....	58
Tabela 3 - Distribuição da população por faixa percentual segundo Hammes (2018) e do presente estudo	59
Tabela 4 - Influência das variáveis na análise preditiva do presente estudo	61

LISTA DE ABREVIATURAS E SIGLAS

AP	Atenção Primária
API	<i>Application Programming Interface</i>
APS	Atenção Primária à Saúde
AVC	Acidente Vascular Cerebral
CDS	Coleta de Dados Simplificada
CID-10	Classificação Internacional de Doenças, versão 10
CN	Centralizador Nacional
CNS	Cartão Nacional de Saúde
CSV	<i>Comma-separated Value</i>
DCNT	Doenças Crônicas Não Transmissíveis
DCV	Doenças Cardiovasculares
DDCV	<i>Dashboard</i> de Doenças Cerebrovasculares
DSRM	<i>Design Science Research Methodology</i>
e-Gov	Governo Eletrônico
e-SUS AB	e-SUS Atenção Básica
FHIR	<i>Fast Healthcare Interoperability Resources</i>
HIC	Hemorragia Intracerebral
HSA	Hemorragia Subaracnóidea
HTTP	<i>Hypertext Transfer Protocol</i>
ICSAP	Interações por Condições Sensíveis à Atenção Primária
IDS	Índice de Desenvolvimento Socioeconômico
IHME	<i>Institute for Health Metrics and Evaluation</i>
JSON	<i>JavaScript Object Notation</i>
LEDI AB	Layout e-SUS AB de Dado e Interface
MS	Ministério da Saúde
OMS	Organização Mundial da Saúde
ONC	<i>Office of the National Coordinator for Health Information Technology</i>
PEC	Prontuário Eletrônico do Cidadão
RES	Registro Eletrônico de Saúde
REST	<i>Representational State Transfer</i>
S-RES	Sistemas de Registro Eletrônico em Saúde

SAPS Secretaria de Atenção Primária à Saúde
SES-SC Secretaria de Estado da Saúde de Santa Catarina
SGD Gradiente Descendente Estocástico
SNC Sistema Nervoso Central
TIC Tecnologias de Informações e Comunicações
SUS Sistema Único de Saúde
URL *Uniform Resource Locator*

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS.....	22
1.1.1	Objetivo Geral	22
1.1.2	Objetivos Específicos	22
2	REVISÃO DA LITERATURA	24
2.1	<i>APPLICATION PROGRAMMING INTERFACE</i>	24
2.2	DOENÇAS CEREBROVASCULARES	26
2.3	ANÁLISE PREDITIVA.....	27
2.4	<i>DASHBOARD</i> DE DOENÇAS CEREBROVASCULARES	31
2.4.1	Mecânica do DDCV	32
2.4.2	Análise Preditiva do DDCV	33
3	MÉTODO	35
3.1	NATUREZA DO ESTUDO.....	35
3.2	TIPO DO ESTUDO.....	35
3.3	VARIÁVEIS DO ESTUDO	35
3.4	PROCESSO DE DESENVOLVIMENTO	36
3.5	INSTRUMENTO DE COLETA DE DADOS.....	38
3.6	CONSIDERAÇÕES ÉTICAS	39
3.7	ORGANIZAÇÃO E ANÁLISE DOS DADOS.....	39
4	DDCV: INTEGRAÇÃO, ANÁLISE PREDITIVA E MECÂNICA DA FERRAMENTA	40
4.1	IMPLEMENTAÇÃO DA API REST	40
4.1.1	Endpoint de Cadastro Individual	44
4.1.2	Endpoint de Cadastro Domiciliar	45
4.1.3	Endpoint de Atendimento Individual	46
4.1.4	Endpoint de Visita Domiciliar	46

4.1.5	Endpoint de Marcadores de Consumo Alimentar	47
4.1.6	Simulação de Uso da API	48
4.2	APRIMORAMENTO DA ANÁLISE PREDITIVA	50
4.2.1	Variáveis Referentes à Dieta	52
4.2.2	Variável Escolaridade	53
4.2.3	Base de Dados	55
4.2.4	Treinamento do Algoritmo	55
4.3	GEORREFERENCIAMENTO	56
4.4	MECÂNICA DO DDCV	57
5	RESULTADOS	58
6	CONCLUSÃO	66
	REFERÊNCIAS	70
	ANEXO A – Lista das Condições Sensíveis à Atenção Primária para Doenças Cerebrovasculares	78

1 INTRODUÇÃO

e-Saúde é uma definição da Organização Mundial da Saúde (OMS) para a aplicação das Tecnologias de Informação e Comunicação à Saúde (WHO, 2007). Tem por objetivo qualificar e expandir o acesso à atenção à saúde por meio do uso das Tecnologias de Informação e Comunicação (TIC). É também um termo genérico que abrange ampla gama de serviços de saúde e cuidados prestados por meio das TIC, tais como: Registro Eletrônico de Saúde (RES), sistema de informação de saúde, monitoramento remoto e serviços de consulta (telessaúde, telemedicina, telecuidados), bem como ferramentas para gestão e análise de dados em saúde, entre outros (BARBELLA et al., 2017).

A e-Saúde tem potencial de melhorar o atendimento e oferecer novos serviços para pessoas com multimorbidade conforme ressalta Barbella et al. (2017). Neste sentido, o Brasil está em processo de adesão a e-Saúde e definiu como meta até 2020 incorporar a e-Saúde ao Sistema Único de Saúde (SUS), sendo reconhecida como estratégia de melhoria consistente dos serviços de saúde por meio da disponibilização e uso da informação abrangente, precisa e segura, que agilize e melhore a qualidade da atenção e dos processos de saúde, nas três esferas de governo e no setor privado, beneficiando pacientes, cidadãos, profissionais, gestores e organizações de saúde (BRASIL, 2017).

Portanto, existe a necessidade de uma base de RES minimamente comunicável entre as Redes de Atenção à Saúde. Assim, entende-se que avançar na construção de uma infraestrutura de padrões e de tecnologia da informação em saúde, proporcionará maior interoperabilidade de sistemas, considerando a informação como importante ferramenta na promoção da equidade.

Desde sua instituição em 1988, o SUS tem o objetivo de atender ao direito de acesso à saúde a todos os cidadãos brasileiros. Considerado um dos maiores sistemas públicos do mundo, oferece de maneira gratuita, serviços de saúde na Atenção Primária (AP), Média e Alta complexidade. Por ser um sistema com princípios, culturas, práticas e objetivos bem definidos, o SUS apresenta facilidade para definir objetivos a serem atingidos pela Estratégia de e-Saúde.

Por sua vez, a Atenção Primária, porta de entrada de uma rede de serviços de acesso universal, deve coordenar os cuidados na rede de atenção à saúde e efetivar a integralidade nas diversas dimensões (SUMAR; FAUSTO, 2014). A AP é o primeiro nível de atenção à saúde e tem como principal objetivo atuar de modo preventivo e fazer a ligação com os demais pontos de atenção.

A Atenção Primária se caracteriza por um conjunto de ações de saúde, no âmbito individual e coletivo, que abrange a promoção e a proteção da saúde, a prevenção de agravos, o diagnóstico, o tratamento, a reabilitação, redução de danos e a manutenção da saúde com o objetivo de desenvolver uma atenção integral que impacte na situação de saúde e autonomia das pessoas e nos determinantes e condicionantes de saúde das coletividades (BRASIL, 2011). Em 2014, a AP atingiu o patamar de 56,41% de cobertura da população brasileira, o que representa aproximadamente 110 milhões de cidadãos (BRASIL, 2014).

As ações da AP vão ao encontro dos benefícios trazidos pela adoção da estratégia de e-Saúde, na qual prevê o acesso à informação pelo cidadão sobre a sua saúde e aos dados agregados relativos à comunidade em que vive, bem como sobre as doenças que os afetam; apoio ao profissional de saúde no planejamento do cuidado ao paciente, diagnóstico, tratamento ou prognóstico; permitir ao gestor de saúde acesso às informações necessárias para a gestão clínica, financeira e administrativa; monitoramento e avaliação de análises de tendências dos processos e resultados de saúde; e detecção oportuna de eventos de saúde pública por meio de sistemas de informação que permitem a análise sistemática de dados clínicos, epidemiológicos e de fluxo de paciente (BRASIL, 2017).

A visão de e-Saúde para o Brasil incentiva todas as partes interessadas no ramo da informática em saúde, sejam elas públicas, privadas ou sem fins lucrativos, a alinharem seus esforços, de maneira a aumentar o impacto das iniciativas de concepção, desenvolvimento, aquisição e implantação de sistemas de informação em saúde, incluindo dispositivos, modelos e processos no SUS.

A Secretaria de Atenção Primária à Saúde (SAPS) do Ministério da Saúde (MS) elaborou uma estratégia denominada estratégia e-SUS AB formada pelo Sistema de Coleta de Dados Simplificado (e-SUS AB CDS) e pelo Sistema com Prontuário Eletrônico do Cidadão (e-SUS AB PEC) (BRASIL, 2013a) com a proposta de reestruturação do registro de informações das ações realizadas na AP em nível

nacional, da gestão da informação, a melhoria das condições de infraestrutura e a melhoria dos processos de trabalho por intermédio da informatização e qualificação da informação em busca de um SUS eletrônico (e-SUS) que apoie os municípios e os serviços de saúde na gestão efetiva da AP e na qualificação do cuidado dos usuários (BRASIL, 2015; BRASIL, 2019a; SOARES, 2016). Estes sistemas são disponibilizados gratuitamente para todos os municípios brasileiros e são os principais coletores de dados de saúde da população brasileira que utilizam o Sistema Único de Saúde.

Até dezembro de 2017, praticamente 100% dos municípios do Brasil haviam instalado pelo menos um dos sistemas da estratégia ou ainda realizavam a integração com outros sistemas, sejam eles próprios ou terceirizados. Todas essas instalações enviam dados de saúde para um repositório de dados nacional chamado de Centralizador Nacional (CN). Estes sistemas permitem o registro de informações individualizadas por meio do Cartão Nacional de Saúde (CNS), promovendo a gestão do cuidado do cidadão e também a integração com outros sistemas. Atualmente, o prontuário eletrônico está instalado em 2.770 municípios. Só em 2017 foram registrados 105 milhões de procedimentos, 72 milhões de consultas médicas, 48 milhões de consultas odontológicas e 232 milhões de visitas domiciliares. (BRASIL, 2016).

Porém, a estratégia e-SUS AB não possui uma ferramenta para preparação dos dados de forma a ter uma fonte de informação que possa dar suporte permanente e dinâmico ao processo gerencial de apoio a tomada de decisão pelos gestores federais, estaduais e municipais e pelos coordenadores de equipes e profissionais de saúde (HAMMES, 2018).

Um estudo quantitativo de análise de dados desenvolvido por Cavalcanti, Oliveira Neto e Sousa (2015) analisou os principais desafios para a qualificação da AP na perspectiva dos gestores municipais de saúde. O resultado evidenciou a importância de aperfeiçoar os processos de monitoramento e avaliação da Atenção Primária à Saúde (APS) com base em indicadores e auxiliar as equipes na análise e manejo de informações em saúde de modo que tenham valor de uso e sirvam para a tomada de decisão no âmbito local.

Neste sentido, importante ressaltar que os sistemas de informação são primordiais para a gestão das políticas públicas, pois, com o advento das novas tecnologias da informação e comunicação e da estratégia e-Saúde, tem-se o aumento da captação de dados e da necessidade de transformá-los em informação com agilidade e segurança, de modo que consigam gerar insumos para as tomadas de decisão (SOARES, 2016). Isto é essencial para entender o porquê de algumas doenças apresentarem índices elevados de morbimortalidade não só no Brasil, mas também no mundo, como é o caso das Doenças Crônicas Não Transmissíveis (DCNT) (DUNCAN et al., 2012).

Os Sistemas de Saúde, dentre as suas diversas funções, têm a responsabilidade de identificar os cidadãos com fatores de risco elevados, possibilitando aos profissionais realizarem a busca ativa a estes cidadãos. Quanto mais cedo iniciar o tratamento, seja ele preventivo (orientação sobre hábitos saudáveis) para casos menos graves ou medicamentoso (mediante diagnósticos comprovados por exames clínicos) para os casos mais graves, maiores as chances de evitar o desenvolvimento de DCNT, como é o caso das doenças cardiovasculares (BRASIL, 2012).

As doenças cardiovasculares (DCV) são a principal causa de morte no mundo (ROTH et al., 2017; LOTUFO et al., 2017). Em 1990 foram registradas mais de 12,5 milhões de mortes causadas por DCV. Este número subiu para aproximadamente 18 milhões em 2015, representando $\frac{1}{3}$ de todas as mortes registradas. Cerca de 423 milhões de pessoas apresentaram doenças cardiovasculares prevalentes, isso representa aproximadamente uma a cada 17 pessoas do mundo (ROTH et al., 2017).

Especificamente, as doenças cerebrovasculares são a segunda maior causa de mortes no mundo, ficando atrás apenas das doenças cardíacas isquêmicas. Estima-se que em 2015 foram registrados 42,43 milhões de casos de doenças cerebrovasculares, sendo 6,33 milhões de óbitos (ROTH et al., 2017). Raramente uma pessoa sobrevive a um acidente vascular cerebral sem precisar readequar seu estilo de vida devido às complicações deixadas pelo episódio, muitas vezes dependendo de cuidados de outras pessoas. (ASADI et al., 2014; TRUELSEN; BEGG; MATHERS, 2005). Estes resultados mostram a importância de intensificar os investimentos na prevenção e tratamento de DCV em todo o mundo.

Um dos objetivos do Desenvolvimento Sustentável das Nações Unidas é reduzir a taxa de mortalidade prematura por DCNT em um terço até 2030 (UNITED NATIONS, 2015). Segundo Mozaffarian (2017), para alcançar este objetivo é necessário uma reestruturação do sistema de saúde mundial como um todo, desde as taxas de cobertura, acolhimento, acesso a tratamentos cardiovasculares e terapias preventivas até inovações de sistemas e políticas efetivas direcionadas para o comportamento e estilo de vida das pessoas, especialmente relacionados ao tabagismo, dieta inadequada (alto consumo de sódio) e inatividade física. Os benefícios trazidos por esta reestruturação interferem diretamente em outras doenças, por exemplo, diabetes do tipo 2, câncer e obesidade.

Em 2008, o MS publicou a Portaria GM/MS nº 221, de 17 de abril de 2008, a qual define a Lista Brasileira de Internações por Condições Sensíveis à Atenção Primária (ICSAP) que deverá ser utilizada como instrumento de avaliação da Atenção Primária (AP) e poderá ser aplicada para avaliar o desempenho do sistema de saúde nos âmbitos nacional, estadual e municipal. Esta lista apresenta 19 grupos de causas de internações e diagnósticos de acordo com a Décima Revisão da Classificação Internacional de Doenças (CID-10). Em específico, o grupo 12 aborda as doenças cerebrovasculares. Estão presentes neste grupo as CID-10 I63 a I67, I69, G45 e G46, conforme listado no ANEXO A – Lista das Condições Sensíveis à Atenção Primária para Doenças Cerebrovasculares (BRASIL, 2008).

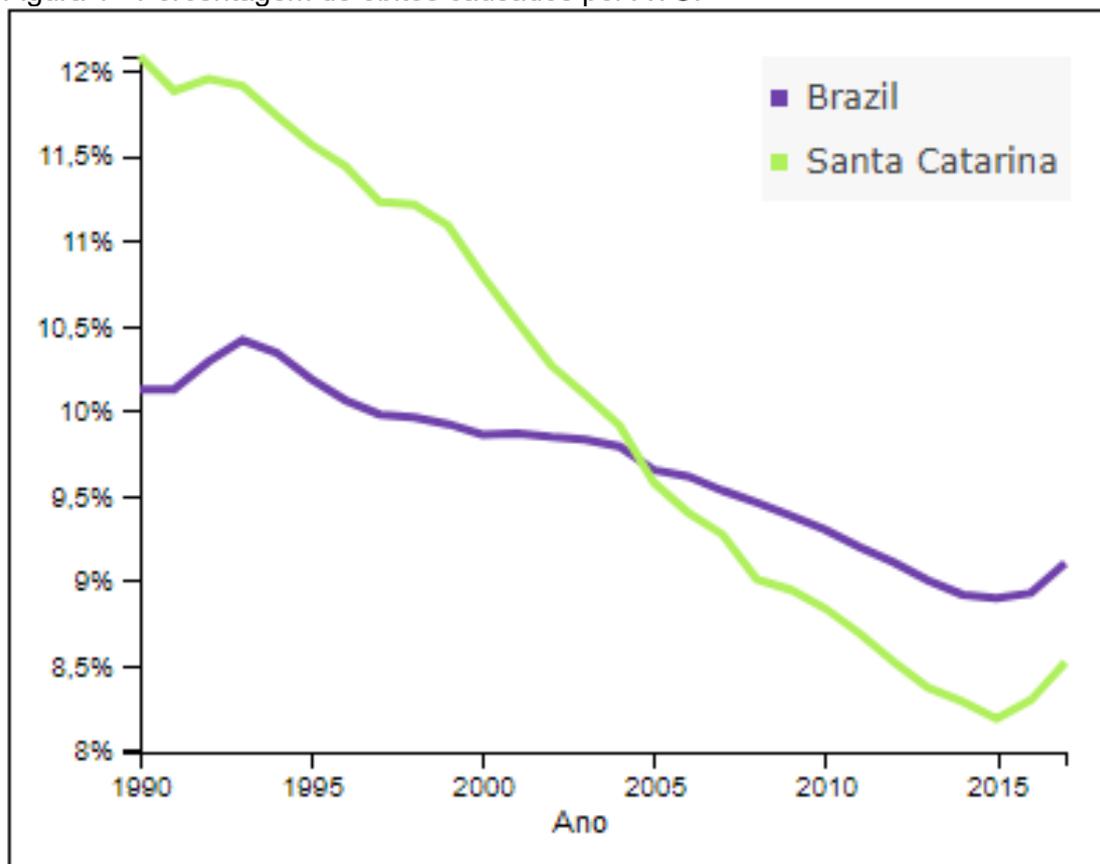
O Brasil apresenta um dos índices mais elevados da América do Sul em se tratando de mortalidade por doenças do aparelho circulatório (WANG et al., 2016; DUNCAN et al., 2012; RIBEIRO et al., 2016). Dentre as doenças cardiovasculares, a cerebrovascular é uma das mais negligenciadas no país por apresentar características peculiares perante a realidade brasileira e devido às características socioeconômicas do país (LOTUFO et al., 2017).

Segundo Lotufo et al. (2017), em 2015 foram registrados pouco mais de 144 mil óbitos por doença cerebrovascular em todo território nacional. Destes, 51% eram pessoas do sexo masculino. O número de casos prevalentes em 2015 foi de aproximadamente 593 mil, sendo que não foi observada diferença considerável entre o sexo masculino e feminino, 50,22% e 49,78%, respectivamente. Quando a

análise foi realizada por estado, percebeu-se a correlação entre a doença cerebrovascular e o índice de desenvolvimento socioeconômico (IDS). Estados com alto IDS apresentaram taxas menores de doença cerebrovascular quando comparados a estados com IDS inferiores.

O estado de Santa Catarina (SC), classificado com alto IDS, apresentou 8.956 internações hospitalares relacionadas às doenças cerebrovasculares em 2014. Isto representa 11,2% do total de internações (SES-SC, 2014). No entanto, há poucas informações sobre o comportamento desta doença à medida que sua granularidade diminui (WANG et al., 2016). Contudo, conforme apresentado na Figura 1, não apenas em nível nacional, mas também no Estado de Santa Catarina, as taxas de mortalidade por acidente vascular cerebral (AVC) estão em ascensão desde 2015, quando foi registrado o menor índice: 8,52%.

Figura 1 - Porcentagem de óbitos causados por AVC.



Fonte: GHDx¹ (2019).

¹ Global Health Data Exchange (GHDx): GBD Results Tool (<http://ghdx.healthdata.org/gbd-results-tool>)

Segundo a Secretaria de Estado da Saúde de Santa Catarina (SES-SC), as Internações por Condições Sensíveis à Atenção Primária (ICSAP) podem ser reduzidas por meio de uma AP oportuna e eficaz. Isto inclui as doenças cerebrovasculares. Portanto, a partir da necessidade de se ter estruturas informatizadas para suprir a carência de informações sobre doenças cerebrovasculares a fim de aperfeiçoar o cuidado e reduzir os índices de prevalência e óbito, Hammes (2018), desenvolveu um *dashboard* georreferenciado juntamente com um modelo de análise preditiva que determinou probabilisticamente a chance de um cidadão desenvolver doenças cerebrovasculares na Atenção Primária.

Por outro lado, verificou-se que o modelo de análise preditiva de Hammes (2018) utilizou um conjunto de dados limitado, disponível no sistema e-SUS AB. Isto pode resultar em uma probabilidade imprecisa. Existem outros fatores de risco importantes para o processo de análise probabilística de um cidadão desenvolver doenças cerebrovasculares que podem ser incluídos no algoritmo de predição, por exemplo, características fisiológicas e características sociais (TRUELSEN; BEGG; MATHERS, 2005; FEIGIN; NORRVING; MENSAH, 2017; PARMAR et al., 2015). Estes fatores podem aperfeiçoar o aprendizado do algoritmo, aumentando a confiabilidade da análise preditiva e identificando padrões entre os dados.

Além disto, o processo de funcionamento do *Dashboard* é realizado mediante interações do usuário com a ferramenta. Estas interações estão relacionadas com importar os dados do e-SUS AB na base de dados do *dashboard*, gerar o arquivo a ser utilizado pelo algoritmo de análise preditiva por meio de um sistema gerenciador de banco de dados, carregar o arquivo no algoritmo (ferramenta não acoplada ao *dashboard*), importar o resultado da predição no *dashboard* para então visualizar o resultado na ferramenta desenvolvida por Hammes (2018). Isto é, para cada etapa, o usuário precisa exportar e importar arquivos nos respectivos sistemas.

Entender o comportamento de uma doença a longo prazo é fundamental para aprimorar o processo de prevenção e tratamento, bem como reduzir o número de óbitos por ela causados, afirma Hammes (2018). O desenvolvimento da

compreensão e da análise do que é coletado e como as informações podem ser utilizadas para gerir o coletivo é essencial para o sucesso de um sistema de informação voltado para a área da saúde. Assim, a informação é discutida, debatida e consensuada, a fim de qualificar as melhores práticas de cuidado (CAVALCANTE, 2018).

Diante deste cenário, busca-se aprimorar o estudo de Hammes (2018) com as seguintes questões norteadoras:

- Quais ferramentas tecnológicas são necessárias para automatizar o processo de importação de dados do *Dashboard* de Doenças Cerebrovasculares proposto por Hammes (2018) em todas as etapas de desenvolvimento visando sua integração com o e-SUS AB PEC?
- Que variáveis precisam ser implementadas ao modelo preditivo proposto por Hammes (2018) a fim de identificar padrões e melhorar a confiança da previsão das doenças cerebrovasculares?

1.1 OBJETIVOS

Nas seções abaixo estão descritos o objetivo geral e os objetivos específicos.

1.1.1 Objetivo Geral

O objetivo geral deste estudo é apresentado em duas temáticas. A primeira é automatizar o processo de importação de dados do *Dashboard* de Doenças Cerebrovasculares (DDCV) em todas as etapas de seu desenvolvimento visando sua integração com o e-SUS AB PEC. A segunda é aprimorar o modelo preditivo proposto por Hammes (2018) a fim de identificar padrões e melhorar a confiabilidade da previsão das doenças cerebrovasculares.

1.1.2 Objetivos Específicos

Para almejar os objetivos deste trabalho, algumas etapas foram definidas conforme apresentadas nos objetivos específicos a seguir:

- a) Criar uma API (*Application Programming Interface*) para a integração entre o DDCV e o sistema de informação e-SUS AB PEC;
- b) Incluir novos fatores de risco ao algoritmo de predição para aumentar o grau de confiança da probabilidade de um cidadão desenvolver doenças cerebrovasculares;
- c) Apresentar no *dashboard* o mapa de calor das probabilidades de cidadãos desenvolverem as doenças cerebrovasculares;
- d) Realizar análise comparativa entre os resultados obtidos por Hammes (2018) e a nova proposta para o estudo.

2 REVISÃO DA LITERATURA

Este capítulo destina-se ao estudo acerca dos temas que serão abordados neste trabalho com ênfase em API, doenças cerebrovasculares, análise preditiva e *Dashboard* de Doenças Cerebrovasculares desenvolvido por Hammes (2018).

2.1 APPLICATION PROGRAMMING INTERFACE

Uma API ou *Application Programming Interface*, é uma maneira de programas de computadores se comunicarem através de um conjunto de padrões e rotinas estabelecidas para a utilização de suas funcionalidades por um aplicativo que não pretende envolver-se em detalhes de implementação do respectivo programa de computador, mas apenas usar os serviços (JONES, 2016). A troca de informações entre diferentes sistemas exige apenas o conhecimento sobre as instruções que precisam ser passadas para a API, descartando todas as regras e como os dados estão armazenados neste outro sistema.

Nos últimos 20 anos, as APIs têm ajudado a conectar dados, aplicações e processos de negócio. No entanto, as regras e funções destas estruturas vêm mudando rapidamente, tornando-se um importante elemento na área da saúde, melhorando a interoperabilidade, a troca de informações e a aceleração digital (DELOITTE DEVELOPMENT LLC, 2018).

Com a promessa de integração rápida, leve e padronizada, não há limite para as possibilidades que uma API pode permitir (JONES, 2016). Os benefícios podem ser um banco de dados de saúde da população com dados e análises em tempo real; concessão de acesso dos pacientes a seus registros médicos; e oportunidades de pesquisa médica que surgem do novo acesso a dados-chave. É fundamental para ramos de negócio, pois faz parte de elementos essenciais da transformação digital, capaz conectar profissionais, parceiros, aplicativos e dispositivos de dados (Ca TECHNOLOGIES, 2015).

No mundo digital atual, a relevância das APIs aumentou. Talvez seja a tecnologia mais essencial no projeto de uma empresa digital (GILL; HEFFNER, 2015). Estas implementações também permitem que as organizações atendam às demandas constantes de funcionalidades novas ou melhoradas, além de facilitar o

processo de melhoria de desempenho, segurança e capacidade de gerenciamento (Ca TECHNOLOGIES, 2015).

À medida que APIs em saúde vão tornando-se conhecidas, novos fluxos de dados podem permitir que sistemas acessem os dados assim que eles estiverem disponíveis. Esta nova estratégia potencializará iniciativas de atendimentos baseadas no valor da informação, como medicina de precisão, saúde da população e análise preditiva (JONES, 2016).

Nos Estados Unidos foi criado o *Office of the National Coordinator for Health Information Technology* (ONC) com a missão de desenvolver políticas e promover habilidades aos pacientes e fornecedores para troca de informações em saúde eletronicamente em colaboração com a indústria da tecnologia da informação em saúde (EUA, 2016). Foram definidas especificações para implementações de APIs voltados para a área da saúde. A adoção e implementação do padrão *Fast Healthcare Interoperability Resources* (FHIR) em tecnologia da informação em saúde também está voltado para que as organizações sigam todas as exigências no intuito de alcançarem os critérios de certificação destas APIs.

Em 2015 a ONC reconheceu o potencial das APIs para revolucionar o compartilhamento de dados de saúde. Jones (2016) afirma que APIs modernas têm potencial de desbloquear dados e abrir uma infinidade de possibilidades para o cuidado de pacientes. O uso de APIs integradas aos Sistemas de Registro Eletrônico em Saúde (S-RES) pode facilitar aos pacientes a consulta e compartilhamento de informações importantes sobre sua saúde, desencadeando o poder transformador dos dados (EUA, 2016; DELOITTE DEVELOPMENT LLC, 2018).

Em uma perspectiva futura, um dos maiores benefícios seria a melhoria na qualidade do cuidado ao paciente e conseqüentemente a redução dos custos com o tratamento. Isto seria possível através do uso de dispositivos vestíveis (*Wearable Technologies*) que, por meio de uma API, coletam e enviam dados dos sinais vitais do paciente para o sistema. Este sistema é responsável por avaliar e, com base nos dados recebidos, corrigir as doses de medicamentos deste paciente em tempo real.

Em teoria, está comprovado o quão importante é para a área da saúde a interação de APIs e S-RES, além da infinidade de benefícios que esta tecnologia

pode trazer para a área da saúde. No entanto, na vida real, apesar das APIs estarem consolidadas há muito tempo, isto evolui morosamente devido a critérios rigorosos de segurança, privacidade, propriedade das informações de saúde e principalmente pelo fato de os S-RES ainda não estarem consolidados nacionalmente.

2.2 DOENÇAS CEREBROVASCULARES

Entre as doenças crônicas não transmissíveis, as que afetam o aparelho circulatório constituem a principal causa de morbidade em todo mundo, incluindo o Brasil. Dentre as doenças cardiovasculares (DCV), a cerebrovascular apresenta características específicas dentro da realidade brasileira, por ser uma das doenças mais negligenciadas no país, com incidência e mortalidade elevadas (LOTUFO et al., 2017).

De acordo com Lotufo et al. (2017), um dos aspectos principais da mortalidade cerebrovascular é o fato da redução do risco de morte pela doença ser muito desigual em macrorregiões, estados ou distritos de cidades quando comparada a de outras doenças crônicas não transmissíveis.

Para funcionar adequadamente, o cérebro necessita de oxigênio e nutrientes que são fornecidos pelo sangue. A doença cerebrovascular ou acidente vascular cerebral (AVC) também é conhecido popularmente como derrame decorrente da insuficiência de sangue em uma determinada área de cérebro (BRASIL, 2012). As principais causas são malformação arterial cerebral (aneurisma), hipertensão arterial, cardiopatia, tromboembolia (bloqueio da artéria pulmonar).

O acidente vascular cerebral é classicamente caracterizado como um déficit neurológico atribuído a uma lesão focal aguda do sistema nervoso central (SNC) por uma causa vascular. O AVC é dividido em dois subtipos: isquêmico e hemorrágico. O AVC isquêmico é um coágulo que bloqueia o fluxo sanguíneo em uma artéria cerebral, causando falta de circulação no seu território vascular. Por sua vez, o AVC hemorrágico é a ruptura de um vaso sanguíneo causando hemorragia intracerebral, intraventricular ou subaracnoidea (BOEHME; ESENWA; ELKIND, 2017).

Segundo o Ministério da Saúde (2012), o fumo é responsável por 25% das doenças vasculares, dentre elas o AVC. Geralmente um cidadão que sofre um AVC

vem a óbito ou apresenta dificuldade na fala e na locomoção devido à paralisação de parte do corpo. O diagnóstico é realizado por exames de imagem que permitem identificar o tipo e a região do cérebro afetada pelo episódio.

Esta doença geralmente não apresenta avisos prévios e é comum pessoas idosas apresentarem o AVC mais de uma vez, contribuindo fortemente para os casos de demência e elevação nas taxas de mortalidade. O tratamento preventivo engloba o controle de vários fatores de risco vasculares modificáveis, como a pressão arterial, diabetes, colesterol, triglicérides, doenças cardíacas, além da necessidade de não fumar, ter uma alimentação saudável e praticar exercícios físicos (BRASIL, 2012). Para Boehme, Esenwa e Elkind (2017), além dos fatores modificáveis, também existem os fatores de riscos não modificáveis, como sexo, idade, genética e raça / cor.

Geralmente os fatores de risco, modificáveis ou não modificáveis, estão ligados tanto para o AVC isquêmico quanto para o hemorrágico. Entretanto, alguns fatores tendem a ser particularmente ligados a um dos tipos, como é o caso da hipertensão arterial, considerada um fator de risco do ACV hemorrágico, embora possa contribuir para o AVC isquêmico (BOEHME; ESENWA; ELKIND, 2017).

Com o objetivo de mapear os casos de doenças cerebrovasculares e avaliar o desempenho do sistema de saúde nos âmbitos Nacional, Estadual e Municipal, o Ministério da Saúde, através da Portaria nº 221, de 17 de abril de 2008 (BRASIL, 2008) estabeleceu a Lista Brasileira de Internações por Condições Sensíveis à Atenção Primária. Dentre os grupos apresentados, encontra-se a lista para doenças cerebrovasculares, conforme apresentada ANEXO A – Lista das Condições Sensíveis à Atenção Primária para Doenças Cerebrovasculares. A lista foi desenvolvida de acordo com a Décima Revisão da Classificação Internacional de Doenças (CID-10).

2.3 ANÁLISE PREDITIVA

A partir dos anos 2000, o termo análise preditiva se tornou sinônimo do termo mineração de dados ou *data mining*. (FINLAY, 2014). É um termo amplo,

empregado para descrever técnicas de análise e estatística utilizadas para desenvolver modelos capazes de prever eventos futuros ou comportamentos (NYCE, 2007; FINLAY, 2014; BARTLEY, 2017).

As técnicas estatísticas de análise preditiva incluem modelagem de dados, aprendizado de máquina (*machine learning*), inteligência artificial, algoritmos de aprendizado profundo (*deep learning*) e a mineração de dados. Esses componentes da ciência de dados analisam detalhes novos e históricos para realizar determinadas tarefas (HAN; KAMBER; PEI, 2012; GRAHAM, et al. 2018).

As principais funcionalidades da mineração de dados são usadas para especificar os tipos de padrões encontrados nas tarefas realizadas. Geralmente são classificadas na categoria descritiva ou preditiva. A categoria descritiva caracteriza propriedades em um conjunto de dados, enquanto que a categoria preditiva executa a indução nos dados atuais para realizar previsões futuras (HAN; KAMBER; PEI, 2012).

A essência da análise preditiva é usar dados existentes para criar um modelo consistente o qual será usado para prever dados que não existem. Pode ser aplicada para gerar qualquer tipo de dados desconhecidos, incluindo passado e presente. Além disso, a previsão pode ser aplicada a casos de uso não temporais, como modelagem de progressão de doenças. A análise preditiva, devido à natureza dos *insights* criados, é probabilística e não determinística. Isto não será capaz de garantir que a previsão realmente aconteça (NATARAJAN; FRENZEL; SMALTZ, 2017).

A modelagem preditiva é competitiva e difícil. Axelrod e Vogel (2003) identificaram sete etapas para a modelagem preditiva:

- a) Primeira etapa: refere-se à coleta dos dados, podendo ser arquivos ou banco de dados;
- b) Segunda etapa: responsável pela agregação dos dados, ela pode combiná-los para gerar um processo relacional e compreensivo, além de realizar a limpeza dos dados;
- c) Terceira etapa: é avaliado como os dados são coletados, surgindo a necessidade de criar novas variáveis;

- d) Quarta etapa: refere-se à transformação das variáveis originais em um conjunto de variáveis derivadas para que possam ser utilizadas pelas várias técnicas de modelagem preditiva;
- e) Quinta etapa: responsável pela mineração de informações, esta etapa revisa as informações disponíveis de maneira intuitiva;
- f) Sexta etapa: responsável pela modelagem de informações, ela revisa o relacionamento entre as variáveis e procurar por melhores variáveis dentro do processo;
- g) Sétima etapa: na última etapa ocorre a definição da relação matemática entre o conjunto final de preditores (variáveis independentes) e os resultados previstos (variáveis dependentes).

Várias tecnologias foram adotadas para prever a ocorrência e os resultados do AVC. Estas tecnologias podem ser divididas em duas categorias principais: métodos estatísticos convencionais e métodos de modelagem computacional inteligente.

Atualmente, os métodos estatísticos convencionais são amplamente utilizados e comumente aplicados para análise de dados de predição de acidente vascular cerebral. Por exemplo, estatísticas descritivas são usadas para calcular a frequência de acidentes vasculares cerebrais na população geral, entre gêneros, grupos étnicos, entre outros. O método de correlação, por exemplo, correlação de *Spearman*, foi utilizado para calcular a correlação entre duas escalas diferentes, tais como o índice de *Barthel* e o SF-36. Foi aplicada regressão logística para analisar os fatores associados às subescalas do SF-36, a fim de descobrir quais destas variáveis melhor discriminam entre pacientes com escores baixo e alto nas subescalas do SF-36. O teste unidirecional de variância e χ^2 (qui quadrado) foram adotados para avaliar as diferenças entre os diferentes desfechos do curso do AVC (LIANG et al., 2014).

No entanto, segundo Liang et al. (2014), os métodos estatísticos convencionais têm limitações na eficiência e na precisão da previsão, em comparação com os métodos de aprendizado de máquina. O aprendizado de máquina é o processo de descobrir e interpretar informações significativas, como

novas correlações, padrões e tendências, analisando grandes quantidades de dados armazenados, usando tecnologias de reconhecimento de padrões, bem como técnicas estatísticas e matemáticas. Em outros contextos, o aprendizado de máquina é um processo de usar diferentes técnicas de análise para observar informações potencialmente desconhecidas e potencialmente significativas, descobrindo padrões e relações fortes de um grande conjunto de dados que pode ser aplicado com mais precisão a uma pessoa específica.

Para Asadi et al. (2014), a abordagem usual para analisar os dados dos resultados do AVC é desenvolver modelos de regressão logística. No entanto, os algoritmos de aprendizado de máquina têm sido propostos como uma alternativa, em particular para dados multi-institucionais em grande escala, com a vantagem de incorporar novos dados disponíveis para melhorar o desempenho da previsão.

A regressão é uma importante técnica para *data mining*, frequentemente usado para descobrir relacionamentos entre variáveis (KHALID; ABDELWAHAD, 2015). Modelos de regressão logística permitem a identificação e validação de variáveis preditivas (ASADI et al., 2014). A regressão logística, na estatística, é um modelo de regressão onde a variável dependente é categorizada ou binária (KALECHOFISKY, 2016). Na área da medicina, a regressão logística tem sido usada para realizar cálculos probabilísticos com a finalidade de prever a ocorrência de determinadas doenças ou a capacidade de um paciente sobreviver (KHALID; ABDELWAHAD, 2015).

Os algoritmos de *Machine Learning* podem ser aplicados e treinados, em dois grandes cenários diferentes: supervisionado e não supervisionado. No cenário supervisionado, os resultados previstos são conhecidos e usados para treinar os modelos. Em máquinas não supervisionadas, a saída desejada é desconhecida. O objetivo é descobrir a estrutura nos dados, não generalizar um mapeamento de entradas para as saídas. Dois dos métodos de aprendizado de máquina mais usados incluem rede neural artificial e máquina de vetores de suporte. Estes modelos são treinados e supervisionados com algoritmos de redes neurais capazes de treinamento não supervisionado (ASADI, 2014).

Na área da saúde, a análise preditiva ainda é considerada pouco explorada. O maior uso está direcionado para a compreensão do passado para prever atividades futuras, servindo de suporte para a medicina baseada em evidências e

sistemas de apoio à decisão clínica (PALEM, 2013). A análise permite a compilação de informações sobre tendências, padrões, desvios, anomalias e relações nos dados da história médica dos pacientes em diferentes culturas e regiões para detectar qualquer possível epidemia.

Prever os casos que necessitarão de cuidado qualificado com precisão aceitável é atuar na prevenção de agravos clínicos, é reduzir os custos com tratamentos caros, é direcionar os investimentos da saúde pública construindo estratégias para intervir com antecedência (YANG et al., 2017). O cuidado preventivo é uma das formas mais eficazes de reduzir os gastos com saúde oferecendo melhor qualidade no atendimento (MDH, 2015).

2.4 DASHBOARD DE DOENÇAS CEREBROVASCULARES

Dashboards são interfaces acessíveis que apresentam as informações em painéis gráficos com o objetivo de auxiliar analistas e executivos na tomada de decisões (GHAZISAEIDI et al., 2015). Um *dashboard* somente será útil se os dados apresentados são confiáveis, de fácil entendimento e mapeados para objetivos estratégicos específicos com intuito de fornecer aos painéis a capacidade de medir, monitorar e analisar a sua capacidade de entretenimento (SEITZ, 2014; WEGGELAAR-JANSEN; BROEKHARST; BRUIJNE, 2018).

O *Dashboard* de Doenças Cerebrovasculares (DDCV) desenvolvido por Hammes (2018) está organizado em dois módulos principais. O módulo do *dashboard* propriamente dito, responsável por relatórios específicos de doenças cerebrovasculares na Atenção Primária à Saúde (APS), apresentando de forma individualizada os cidadãos que já tiveram ou têm alguma doença cerebrovascular. Também apresenta outro relatório com os indicadores de cada CID-10 do grupo de doenças cerebrovasculares, com quantidade de casos, quantidade por sexo, média de idade e distribuição por período. Ainda no *dashboard*, os casos de cidadãos que já tenham sido acometidos por esta patologia, são apresentados de forma georreferenciada em um mapa. O módulo de análise preditiva, onde um algoritmo aprende com dados históricos de características e fatores de risco para doenças

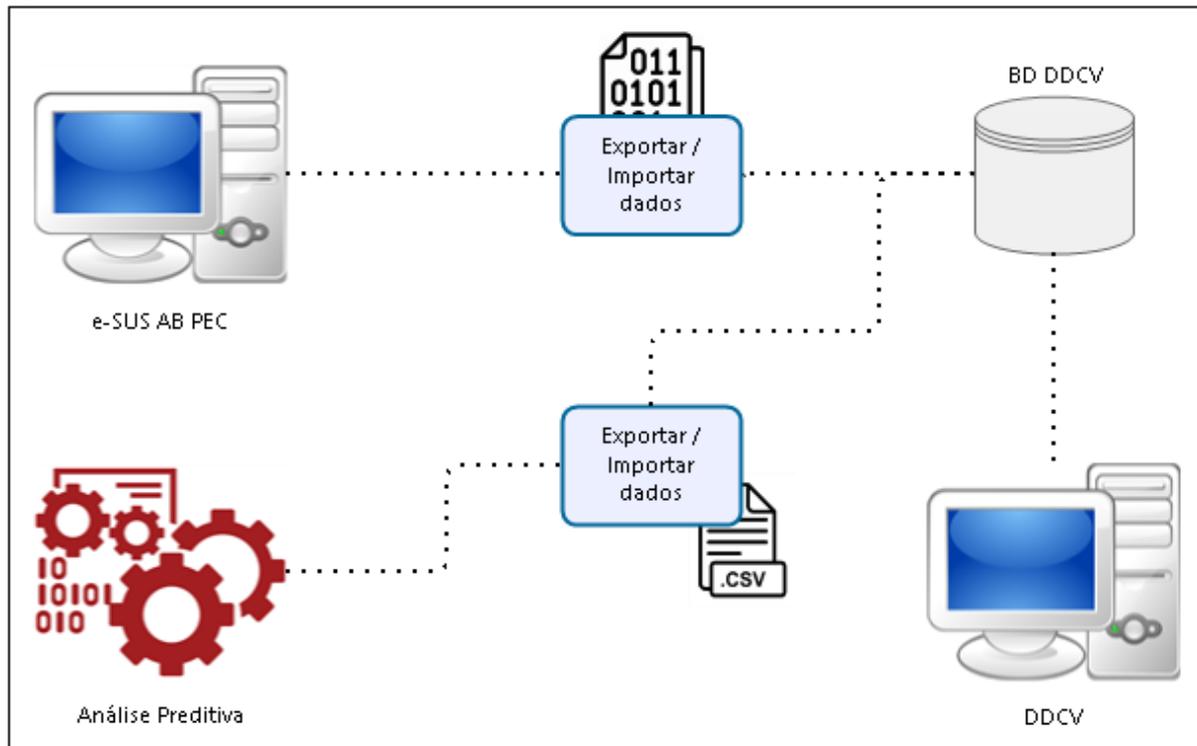
cerebrovasculares com pessoas que já tem uma doença do grupo e prediz a chance de um cidadão ter, por exemplo, um AVC. O resultado desta análise preditiva é apresentado no *dashboard* na forma de relatório.

2.4.1 Mecânica do DDCV

A ferramenta foi desenvolvida utilizando a linguagem de programação Java. Possui um modelo de dados próprio desenvolvido em *PostgreSQL*®. A fonte de informação utilizada para popular o banco dados é oriunda das fichas de Coleta de Dados Simplificada (CDS) do e-SUS AB PEC, ambos sistemas da estratégia e-SUS AB para informatização da Atenção Primária.

A Figura 2 ilustra toda a estrutura de desenvolvimento do DDCV. Os dados são exportados do e-SUS AB e importados na ferramenta. Neste momento alguns relatórios já podem ser visualizados no *dashboard*, por exemplo, os casos de cidadãos que já tiveram doenças cerebrovasculares. Para realizar a análise preditiva é necessário gerar um arquivo no formato *Comma-separated Values* (CSV) e importar na ferramenta de predição. Concluído esta etapa, a ferramenta de predição gera um novo arquivo no formato CSV com a probabilidade do cidadão desenvolver doenças cerebrovasculares, o qual precisa ser importado no DDCV. A partir deste momento, é possível visualizar no relatório qual a probabilidade do cidadão desenvolver algum agravo deste grupo de doenças.

Figura 2 – Mecânica da ferramenta desenvolvida por Hammes (2018).



Fonte: Elaborado pelo autor (2019).

O processo apresentado na Figura 2 não representa ser oneroso. No entanto, requer interação do usuário em cada uma das etapas, realizando a exportação e importação dos arquivos manualmente em cada uma delas. Este diálogo entre os sistemas poderia ser automatizado.

Há muitas formas de automatizar este processo. Dentre elas, destaca-se a *Application Programming Interface* (API). Embora pouco explorada na área da saúde, apresenta grande potencial, permitindo a inclusão dos dados praticamente em tempo real, à medida que vão sendo coletados por diferentes sistemas.

2.4.2 Análise Preditiva do DDCV

Para a análise preditiva foi utilizada a linguagem de programação *Python* em conjunto com uma regressão logística otimizada por um algoritmo de otimização, Gradiente Descendente Estocástico (SGD). Este algoritmo utiliza-se de um conjunto

de variáveis predefinidas para realizar o cálculo probabilístico de um cidadão apresentar a doença cerebrovascular.

O algoritmo usufrui apenas de um subconjunto dos dados armazenados. A seleção é realizada através de uma reamostragem dos dados originais, na qual é considerada o componente estocástico do algoritmo. A população total é de 170.567 cidadãos, sendo que 561 possuem doenças cerebrovasculares. Este grupo (fictício) foi utilizado para treinar o algoritmo de predição. Os fatores de risco considerados por Hammes (2018) são apresentados na Figura 3.

Figura 3 – Variáveis para análise preditiva de Hammes (2018).

```
columns = ["id", "cns", "nome", "idade", "sexo", "raca_cor", "obeso",  
"ex_obeso", "fumante", "ex_fumante", "alcool", "ex_alcool", "drogas",  
"ex_drogas", "hipertensao_arterial", "diabetes", "dcv", "gestante"]
```

Fonte: Hammes (2018).

Existem muitos fatores de risco para doenças cerebrovasculares. Características biológicas: como idade e sexo; características fisiológicas: como colesterol sérico e fibrinogênio; comportamentos como tabagismo, dieta com alta concentração de sódio, açúcares e gorduras saturadas, consumo de álcool e sedentarismo; e características sociais como escolaridade, classe social e raça/cor (TRUELSEN; BEGG; MATHERS, 2005; FEIGIN; NORRVING; MENSAH, 2017; PARMAR et al., 2015). Dentre eles, a hipertensão arterial é o fator de risco mais relevante, com um risco atribuível de opulência de 50% (LOTUFO et al., 2017; TRUELSEN; BEGG; MATHERS, 2005).

Portanto, novas variáveis consideradas importantes para o desenvolvimento de doenças cerebrovasculares podem ser incluídas no intuito de melhorar a eficiência da predição. Além do mais, a apresentação do resultado em forma de relatório não facilita a visualização das regiões com maiores índices da doença. Disponibilizar o resultado na forma de um mapa de calor, análogo a forma em que são apresentados os cidadãos que já foram diagnosticados com o agravo, pode ser um diferencial no planejamento das ações de controle das doenças cerebrovasculares.

3 MÉTODO

Nesta seção será apresentado o método utilizado para o desenvolvimento deste estudo.

3.1 NATUREZA DO ESTUDO

Este trabalho apresenta um estudo de natureza quantitativa, pois visa apresentar em um mapa georreferenciado a probabilidade dos cidadãos desenvolverem doenças cerebrovasculares na Atenção Primária à Saúde (APS) como apoio a tomada de decisão às condições sensíveis a partir da integração com dados do e-SUS AB PEC que requer modelos matemáticos e análise estatística.

3.2 TIPO DO ESTUDO

Estudo inovador de produção tecnológica e analítico uma vez que se aprofunda no desenvolvimento de um *dashboard* interativo e dinâmico para as doenças cerebrovasculares e na análise preditiva a partir de um modelo para as doenças cerebrovasculares.

3.3 VARIÁVEIS DO ESTUDO

A existência de comportamentos para os quais interessa realizar medições objetivas é identificada como variáveis (WAZLAWICK, 2009). As variáveis podem ser caracterizadas como dependentes ou independentes. Variáveis dependentes são influenciadas por outras variáveis, enquanto que as variáveis independentes são as influenciadoras. Neste estudo, as variáveis independentes são aquelas utilizadas para alimentar o algoritmo de análise preditiva, por exemplo, “hipertensao_arterial”. A variável dependente, neste caso, é o desfecho, o qual apresenta a probabilidade de um cidadão desenvolver doenças cerebrovasculares. Portanto, as variáveis

independentes são manipuladas e as variáveis dependentes são as medidas, pois não se trata de um estudo de intervenção.

Complementando o contexto de variáveis, elas sempre são tipadas. Isto é, toda variável necessita de uma definição acerca do tipo do dado que irá armazenar, podendo ser um número, por exemplo, a variável “idade”, um texto, no caso da variável “raça / cor”, ou ainda um valor booleano (verdadeiro ou falso), por exemplo, a variável “dieta_alta_concentracao_sodio”.

3.4 PROCESSO DE DESENVOLVIMENTO

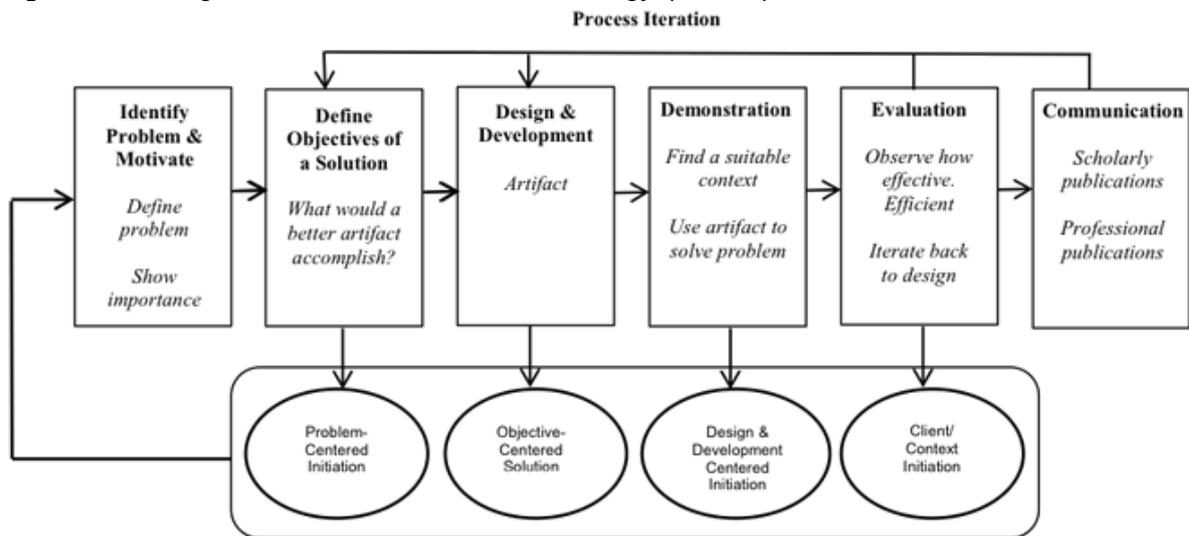
Neste contexto, o artefato a ser desenvolvido é uma API para automatização do processo de importação dos dados da base do sistema e-SUS AB PEC para a ferramenta DDCV implementada por Hammes (2018). Além disso, será adicionado ao banco de dados novas estruturas de dados permitindo a inclusão de informações referentes às dietas com alta concentração de sódio e açúcar, além das informações sobre o grau de escolaridade dos cidadãos. O objetivo de adicionar estes dados é permitir a inclusão de novas variáveis ao algoritmo de análise preditiva, no qual é responsável por apresentar a probabilidade de um cidadão desenvolver doenças cerebrovasculares.

Para orientar o desenvolvimento deste estudo, serão utilizados os princípios e fundamentos de *Design Science Research Methodology* (DSRM). *Design Science* ou Ciência de Projeto abordam os aspectos “o que” e “como” as coisas devem ser e especialmente a concepção de artefatos que tenham por propósito a realização de objetivos (SIMON, 1996; LACERDA et al., 2013; ÇAĞDAŞ; STUBKJÆR, 2011; PEFFERS, 2007).

Segundo Lacerda et al. (2013), o *Design Science* é a base epistemológica e o *Design Science Research* é o método que operacionaliza a construção do conhecimento neste contexto. O *Design Science Research* se constitui em um processo rigoroso de projetar artefatos para resolver problemas, avaliar o que foi projetado ou o que está funcionando, e comunicar os resultados obtidos (ÇAĞDAŞ; STUBKJÆR, 2011).

A DSRM é organizada em seis etapas (conforme apresentada na Figura 4), ordenadas, mas que não precisam necessariamente ser realizadas nesta ordem.

Figura 4 – *Design Science Research Methodology (DSRM)*.



Fonte: Peffers (2007).

Na primeira etapa é definido o problema de pesquisa e as justificativas para a solução proposta. É nesta etapa que se busca entender quais são os problemas relevantes, as soluções existentes e suas fraquezas.

A segunda etapa é responsável por apresentar os objetivos e quais os critérios e passos necessários para solucionar o problema. Iniciativa centrada no problema.

Na terceira etapa ocorre o desenvolvimento do artefato e busca-se a solução centrada no objetivo.

A demonstração e uso do artefato construído são apresentados na quarta etapa. Nesta etapa também se busca provar que o problema previamente relatado foi solucionado. É a iniciativa centrada no desenvolvimento e no design.

Na quinta etapa, busca-se discutir a solução desenvolvida observando sua eficiência comparando os objetivos traçados e os resultados obtidos, iniciativa contexto-cliente.

Na sexta e última etapa, faz-se a divulgação para pesquisadores e demais interessados do problema encontrado, a solução desenvolvida, sua utilidade, inovação e eficácia.

3.5 INSTRUMENTO DE COLETA DE DADOS

A ferramenta desenvolvida por Hammes (2018) utiliza a estrutura de dados dos sistemas da estratégia e-SUS AB, ou seja, os sistemas e-SUS AB PEC e e-SUS CDS, para gerar uma base de dados com dados fictícios. A estratégia e-SUS AB possui um modelo de dados denominado Layout e-SUS AB de Dados e Interface (LEDI AB²). É através deste layout que todos os municípios brasileiros enviam os dados de saúde para o Centralizador Nacional, sejam eles registrados nos sistemas da estratégia e-SUS AB ou através de sistemas terceiros.

O LEDI AB permite a integração entre sistemas próprios e o e-SUS AB PEC através das tecnologias *Apache Thrift* e *eXtensible Markup Language (XML)*. É independente de linguagem de programação e está organizado em 12 estruturas, representando os cuidados em saúde da Atenção Primária. São elas:

- a) Ficha de Cadastro Individual;
- b) Ficha de Cadastro Domiciliar e Territorial;
- c) Ficha de Atendimento Individual;
- d) Ficha de Atendimento Odontológico;
- e) Ficha de Atividade Coletiva;
- f) Ficha de Procedimentos;
- g) Ficha de Visita Domiciliar e Territorial;
- h) Marcadores de Consumo Alimentar;
- i) Ficha de Avaliação de Elegibilidade;
- j) Ficha de Atendimento Domiciliar;
- k) Ficha Complementar - Síndrome Neurológica por Zika / Microcefalia;
- l) Ficha de Vacinação.

Para atender a expectativa deste estudo, além das estruturas de Ficha de Cadastro Individual, Ficha de Cadastro Domiciliar e Territorial, Ficha de Atendimento Individual e Ficha de Visita Domiciliar e Territorial utilizadas por Hammes (2018), foi incluída também a estrutura de Marcadores de Consumo Alimentar. Ela apresenta os dados referentes à dieta nutricional dos cidadãos e servirá como base para a composição das variáveis “dieta_alta_concentracao_sodio” e

² LEDI AB: <https://integracao.esusab.ufsc.br/ledi/index.html>

“dieta_alta_concentracao_acucar”. A variável “escolaridade” será extraída da estrutura Ficha de Cadastro Individual. No próximo capítulo estas variáveis são apresentadas detalhadamente.

3.6 CONSIDERAÇÕES ÉTICAS

Na apresentação de relatórios deste estudo, não foi necessário avaliação pelo Comitê de Ética em Pesquisa pelo fato de utilizar dados gerados aleatoriamente de forma fictícia, não representando cidadãos reais.

Para a prova do modelo em questão, a base de dados contém somente dados fictícios, gerados de forma automatizada. Os dados criados foram baseados nas estruturas de dados do LEDI AB: Ficha de Cadastro Individual, Ficha de Cadastro Domiciliar e Territorial, Ficha de Atendimento Individual, Ficha de Visita Domiciliar e Territorial e Marcadores de Consumo Alimentar. Todas as variáveis necessárias para realizar o cálculo probabilístico de risco para doenças cerebrovasculares estão presentes nos atributos destas estruturas.

3.7 ORGANIZAÇÃO E ANÁLISE DOS DADOS

A predição utilizou a técnica de regressão logística e foi otimizada por um algoritmo de análise preditiva, uma vez que a variável dependente é categórica. Assim, permitiu estimar a probabilidade associada à ocorrência de determinado evento de doenças cerebrovasculares por meio de um conjunto de variáveis independentes. Os resultados da análise ficam contidos no intervalo de zero a um, juntamente com os fatores nos quais o algoritmo julgou importantes para a predição. Estes dados foram organizados em planilhas e analisados mediante estatística descritiva de frequência absoluta e relativa, média, mediana, moda, desvio padrão, valores máximo e mínimo.

4 DDCV: INTEGRAÇÃO, ANÁLISE PREDITIVA E MECÂNICA DA FERRAMENTA

Neste capítulo serão apresentados os resultados do desenvolvimento da tecnologia. Inicialmente foi realizado um estudo técnico sobre a ferramenta previamente desenvolvida por Hammes (2018). O intuito foi conhecer a organização do código-fonte e as tecnologias utilizadas para então dar início ao desenvolvimento das propostas trazidas neste contexto. Não foram encontradas evidências para utilizar outras tecnologias senão as mesmas utilizadas por Hammes (2018). Portanto, neste estudo, também foram mantidas as tecnologias utilizadas anteriormente pelo autor.

O restante deste capítulo irá apresentar as etapas do processo de aprimoramento e evolução da ferramenta, iniciando pelo desenvolvimento da API, conforme será apresentado na seção 4.1. Na sequência, a seção 4.2 irá descrever as evoluções no algoritmo de análise preditiva. Na seção 4.3 será relatado o mecanismo para busca das coordenadas para plotagem no componente de mapas. Por fim, a nova mecânica do *Dashboard* de Doenças Cerebrovasculares é apresentada na seção 4.4.

4.1 IMPLEMENTAÇÃO DA API REST

O *Dashboard* de Doenças Cerebrovasculares (DDCV) necessita se comunicar com o sistema de Prontuário Eletrônico do Cidadão da Atenção Básica (e-SUS AB PEC) para receber os dados dos cidadãos, dos atendimentos e das condições de saúde da população que utiliza o Sistema Único de Saúde (SUS). Sem o recebimento destes dados, o DDCV não tem insumos necessários para operar. Portanto, após estudos realizados, optou-se pelo desenvolvimento de uma API. A função desta API é possibilitar, em um futuro próximo, a interoperabilidade entre o e-SUS AB PEC e o DDCV com dados reais dos cidadãos.

A API desenvolvida utiliza o padrão arquitetural desenvolvido por Fielding e Taylor (2000), denominado *Representational State Transfer* (REST). O REST ignora os detalhes da implementação do componente e da sintaxe do protocolo para se concentrar nas funções dos componentes, nas restrições sobre sua interação com outros componentes e na interpretação de elementos de dados significativos

(FIELDING; TAYLOR, 2000). Com isto, os serviços REST vêm ganhando espaço no mercado e têm sido utilizados pelo Google®, Facebook®, Yahoo!®, Amazon®, eBay®, Microsoft®, dentre outros (DEVMEDIA, 2013).

A forma de conectar os dispositivos por meio da API é utilizando conectores cliente e servidor. Este modelo caracteriza-se por ser uma aplicação distribuída que distribui tarefas entre os fornecedores, denominados servidores. Quem solicita as tarefas ou serviços são designados clientes. Portanto, o cliente inicia a comunicação através do envio de uma requisição enquanto o servidor escuta as conexões e responde para fornecer acesso aos seus serviços. Neste contexto, o e-SUS AB PEC é o cliente e o DDCV, o servidor.

Uma API REST utiliza os métodos básicos *get*, *post*, *put* e *delete* presentes no protocolo de comunicação *Hypertext Transfer Protocol* (HTTP) para ler, inserir, atualizar e excluir dados, respectivamente (RODRIGUEZ, 2008). Isto faz com que os dados sejam transmitidos pela rede de forma mais leve e simples, descartando a criação de camadas intermediárias para encapsular os dados.

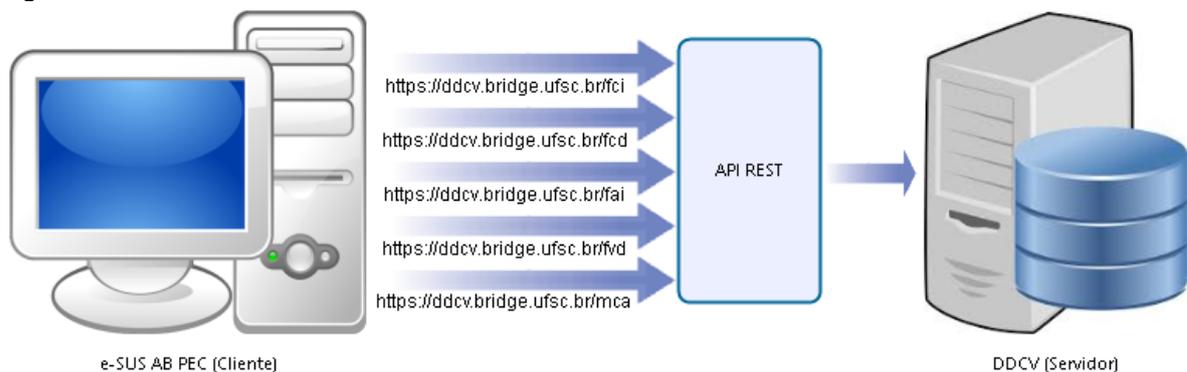
Analisando o escopo da API neste estudo, constatou a necessidade de utilizar apenas o método *post* devido ao fato de o cliente apenas enviar dados para serem processados e armazenados no servidor. O sistema e-SUS AB PEC, no papel de cliente, não poderá solicitar ao servidor o resultado da análise probabilística, tão pouco enviar uma requisição para alterar ou excluir informações enviadas anteriormente. Portanto, não há evidências para implementar os métodos de consultar, alterar ou excluir dados do DDCV.

Todas as interações REST são *stateless*. Ou seja, cada solicitação contém todas as informações necessárias para a aplicação entender a solicitação, independentemente de quaisquer solicitações que possam ter precedido. Esta restrição elimina a necessidade da aplicação manter o estado do aplicativo entre as solicitações, reduzindo assim o consumo de recursos físicos e melhorando a escalabilidade. Além disso, permite que as interações sejam processadas em paralelo sem exigir que o mecanismo de processamento entenda a semântica da interação. Outra vantagem é forçar todas as informações que podem levar à

reutilização de uma resposta armazenada em *cache* em cada solicitação (FIELDING; FLOYD, 2000).

As mensagens enviadas do cliente para o servidor são tratadas como requisições HTTP. As requisições são recebidas pela API (interface entre o cliente e o servidor) e encaminhadas para o servidor processar os dados, conforme apresentado na Figura 5.

Figura 5 – Modelo cliente e servidor.



Fonte: Elaborado pelo autor (2019).

Para que a requisição seja válida, ela precisa estar bem formada. Portanto, seguindo o padrão HTTP, a requisição deve ter quatro partes (DEVMEDIA, 2013), a constar:

- a) URL (*Uniform Resource Locator*): é o endereço a ser acessado, por exemplo, “<https://ddcv.bridge.ufsc.br>”;
- b) Método: é a ação que será realizada, neste caso, o método *post*;
- c) *Header*: contém detalhes de como as requisições serão interpretadas;
- d) *Body*: é o local onde estão presentes os dados que são enviados.

O endereço da requisição, conforme mostrado na Figura 5, deve conter o recurso final a ser acessado pelo cliente, denominado de *endpoint*. A API REST desenvolvida apresenta cinco *endpoints*, um para cada contexto de negócio. Para representar o *body* de cada uma das requisições, as mensagens enviadas precisam ser especificadas no formato *JavaScript Object Notation* (JSON).

De maneira sucinta, JSON³ pode ser definido como um formato de troca de dados simples e rápida, que utiliza texto legível para os humanos no formato de

³ *JavaScript Object Notation* (JSON): <https://www.json.org/json-pt.html>

atributo-valor. Os atributos devem respeitar os tipos de dados que representam, conforme definido no dicionário de dados presente na Figura 6.

Figura 6 – Dicionário de dados da tabela de histórico.

historico
id int8(19) NOT NULL (PK)
id_origem int8(19) NOT NULL
data date(13) NOT NULL
cnes varchar(15) NOT NULL
unidade_saude varchar(200) NULL
ano int4(10) NOT NULL
cns varchar(15) NOT NULL
cns_responsavel varchar(15) NULL
nome varchar(255) NULL
nome_social varchar(255) NULL
data_nascimento date(13) NOT NULL
faixa_etaria varchar(10) NOT NULL
idade int8(19) NOT NULL
sexo varchar(30) NOT NULL
raca_cor varchar(30) NULL
cep varchar(8) NULL
municipio_id int8(19) NULL (FK)
ibge varchar(20) NULL
bairro varchar(100) NULL
uf varchar(2) NULL
tipo_logradouro varchar(20) NULL
logradouro varchar(255) NULL
numero varchar(10) NULL
lat varchar(30) NULL
lng varchar(30) NULL
gestante bool(1) NOT NULL
obeso bool(1) NOT NULL
fumante bool(1) NOT NULL
alcool bool(1) NOT NULL
drogas bool(1) NOT NULL
hipertensao_arterial bool(1) NOT NULL
diabetes bool(1) NOT NULL
avc bool(1) NOT NULL
dcv bool(1) NULL
cid10_1 varchar(4) NULL
cid10_2 varchar(4) NULL
cid10_3 varchar(4) NULL
cid10_4 varchar(4) NULL
cid10_5 varchar(4) NULL
nu_anos_estudo int8(19) NULL
st_dieta_sal bool(1) NOT NULL
st_dieta_acucar bool(1) NOT NULL

Fonte: Elaborado pelo autor (2019).

Nas subseções seguintes, será detalhado cada um dos *endpoints* juntamente com as respectivas mensagens, definidas de acordo com o dicionário de dados. Por fim, será abordado as simulações realizadas utilizando a ferramenta *POSTMAN Learning Center*⁴. É importante ressaltar que os dados apresentados são meramente ilustrativos, não representando dados verídicos.

4.1.1 *Endpoint* de Cadastro Individual

Este *endpoint* é responsável por receber os dados relacionados ao contexto de negócio cadastro do cidadão (FCI). O endereço de acesso é “<https://ddcv.bridge.ufsc.br/fci>”. A especificação do corpo (*body*) da requisição é apresentada na Figura 7.

Figura 7 – *Body* de uma requisição ao *endpoint* FCI.

```
{
  "id": "1",
  "dataCadastroIndividual": "2019-08-22",
  "cnes": "1234567",
  "cns": "123456789012345",
  "cnsResponsavel": "",
  "nomeCidadao": "Nome do cidadão",
  "nomeSocial": "",
  "dataNascimento": "1965-09-22",
  "sexo": "Feminino",
  "racaCor": "BRANCA",
  "avc": "false",
  "diabetes": "false",
  "drogas": "false",
  "fumante": "false",
  "gestante": "false",
  "hipertensaoArterial": "false",
  "condicaoPeso": "Acima do peso",
  "alcool": "false",
  "escolaridade": "63"
}
```

Fonte: Elaborado pelo autor (2019).

⁴ POSTMAN Learning Center: <https://learning.getpostman.com/>

Neste contexto, os dados referentes às condições de saúde do cidadão, por exemplo, diabetes, são autorreferidos. Além disso, o atributo “escolaridade” foi incluído nesta estrutura para manter a informação no mesmo domínio que se encontra no sistema e-SUS AB PEC.

4.1.2 *Endpoint* de Cadastro Domiciliar

O *endpoint* de cadastro domiciliar é responsável por receber os dados relacionados ao contexto de negócio cadastro do domicílio (FCD). O endereço de acesso para este *endpoint* é “https://ddcv.bridge.ufsc.br/fcd”. A especificação do corpo da requisição é apresentada na Figura 8.

Figura 8 – *Body* de uma requisição ao *endpoint* FCD.

```
{  
  "tipoLogradouro": "Rua",  
  "logradouro": "Lauro Linhares",  
  "ibge": "4205407",  
  "bairro": "Trindade",  
  "localidade": "Florianópolis",  
  "numeroDomicilio": "2055",  
  "cep": "88036003",  
  "uf": "SC",  
  "cnsResponsavel": "123456789012345"  
}
```

Fonte: Elaborado pelo autor (2019).

Além dos dados de endereço, também possui o atributo “cnsResponsavel”, o qual indica quem é o responsável pelo respectivo domicílio. A partir deste atributo é possível descobrir quem são os cidadãos que residem em um mesmo domicílio e também buscar as coordenadas para registro no mapa georreferenciado. Mais adiante, este assunto será abordado detalhadamente.

4.1.3 *Endpoint* de Atendimento Individual

É o *endpoint* que contem as principais informações relacionadas ao contexto de negócio das condições de saúde do cidadão. A especificação do corpo da requisição pode ser visualizada na Figura 9 e o endereço de acesso é através da URL “https://ddcv.bridge.ufsc.br/fai”.

Figura 9 – *Body* de uma requisição ao *endpoint* FAI.

```
{
  "id": 1,
  "cids": "CID1|CID2|CID3",
  "sexo": "Feminino",
  "dataAtendimentoIndividual": "2019-03-04",
  "dataNascimento": "1999-02-10",
  "cnes": "1234567",
  "cns": "123456789012345",
  "peso": "60",
  "altura": "1.70",
  "gestante": "false",
  "alcool": "false",
  "diabetes": "false",
  "drogas": "false",
  "fumante": "false",
  "hipertensaoArterial": "false"
}
```

Fonte: Elaborado pelo autor (2019).

Caso tenha sido registrado mais de uma CID-10 durante o atendimento ao cidadão, estas informações devem ser incluídas no valor do atributo “cids”, separando cada uma delas pelo caractere de barra vertical: “|”. Esta organização permite que os valores sejam separados e persistidos nas colunas da tabela histórico referentes as CID-10.

4.1.4 *Endpoint* de Visita Domiciliar

Este *endpoint* é responsável por receber os dados relacionados ao contexto de negócio condições de saúde do cidadão registradas durante a visita domiciliar. Na Figura 10 é apresentada a especificação do *body* de uma requisição acessada a partir do endereço “https://ddcv.bridge.ufsc.br/fvd”.

Figura 10 – *Body* de uma requisição ao *endpoint* FVD.

```
{
  "id": "1",
  "sexo": "Masculino",
  "dataNascimento": "2000-10-01",
  "dataVisitaDomiciliar": "2016-02-11",
  "altura": "1.87",
  "peso": "80",
  "cnes": "1234567",
  "cns": "123456789012345",
  "alcool": "false",
  "diabetes": "false",
  "drogas": "false",
  "fumante": "false",
  "gestante": "false",
  "hipertensaoArterial": "false"
}
```

Fonte: Elaborado pelo autor (2019).

No âmbito da Atenção Primária à Saúde, a visita domiciliar geralmente é realizada pelo agente comunitário de saúde. No entanto, é possível que outros profissionais de saúde também realizem esta atividade.

4.1.5 *Endpoint* de Marcadores de Consumo Alimentar

Diferente dos *endpoints* descritos nas subseções anteriores, os dados referentes ao contexto de negócio marcadores de consumo alimentar não existiam na base de dados do DDCV. Este *endpoint* permite acesso através da url “<https://ddcv.bridge.ufsc.br/mca>”. A estrutura, conforme apresentada na Figura 11, foi desenvolvida para contemplar as variáveis “dieta_sodio” e “dieta_acucar”, abordadas com mais detalhes nas próximas seções.

Figura 11 – *Body* de uma requisição ao *endpoint* MCA.

```
{
  "id": "10",
  "sexo": "Feminino",
  "dataConsumoAlimentar": "2019-05-04",
  "dataNascimento": "1997-02-10",
  "cnes": "1234567",
  "cns": "123456789012345",
  "embutidos": "false",
  "salgadinhos": "false",
  "bebidasDoces": "false",
  "doces": "false"
}
```

Fonte: Elaborado pelo autor (2019).

É relevante destacar que no sistema e-SUS AB PEC os atributos “embutidos”, “salgadinhos”, “bebidasDoces” e “doces” são apresentados em dois grupos de informações: um para crianças entre 6 e 23 meses e outro para crianças com 2 anos ou mais, adolescentes, adultos, gestantes e idosos. Na API REST de integração entre o e-SUS AB PEC e o DDCV, identificar o grupo onde a informação foi registrada não interfere nos propósitos do *dashboard*. No entanto, é possível detectar o grupo a partir dos atributos “dataNascimento” e “dataConsumoAlimentar”, calculando a idade do cidadão.

4.1.6 Simulação de Uso da API

A API REST desenvolvida permite flexibilidade aos clientes, independente de quem seja. As requisições realizadas segundo o protocolo HTTP não possuem interface gráfica, somente atuam no envio e recebimento de dados a partir das próprias requisições.

Neste estudo, o cliente é apenas um, o sistema e-SUS AB PEC, conforme mencionado na seção 4.1. Isto poderia permitir a implementação das requisições diretamente no cliente. Porém, mesmo sendo conhecido, o acesso ao código-fonte do sistema cliente é restrito ao Ministério da Saúde e ao Laboratório Bridge⁵, o qual

⁵ Laboratório Bridge: <https://bridge.ufsc.br/>

é o responsável pelo seu desenvolvimento. Portanto, o teste da API REST, buscando identificar possíveis falhas ou melhorias, precisou ser realizado através da simulação de requisições a partir de um cliente conhecido e que se tenha controle sobre os dados.

Para esta etapa, foi necessário o auxílio da ferramenta *POSTMAN Learning Center*. Esta ferramenta permite realizar requisições HTTP a partir de uma interface simples e intuitiva, simplificando e permitindo simular cenários de teste que representam o uso real (POSTMAN, 2019).

Na Figura 12, é possível perceber a praticidade que a ferramenta proporciona. Na barra superior, é selecionado o método “*POST*” e no campo ao lado informa-se o *endpoint* que será acessado para requisição. O próximo passo é selecionar a aba “*Body*”, definir o formato da mensagem, (neste caso “*JSON (application/json)*”) e então escrever a mensagem no formato JSON no campo de texto. Concluída a escrita da mensagem, o envio é realizado acionando a opção “*Send*”.

Figura 12 – Requisição ao *endpoint* MCA através do *Postman*.



Fonte: Elaborado pelo autor (2019).

Toda a requisição enviada retorna um resultado, o qual é apresentado no campo “*Response*”, localizado abaixo do campo onde a mensagem JSON foi definida. O resultado pode ser de sucesso ou então apresentar um erro que possa ter ocorrido.

Foram realizados testes com todos os *endpoints* simulando casos de sucesso e também vários cenários de erros, dentre eles, o envio de requisições com atributos inexistentes, envio de requisições com valores inválidos e requisições sintaticamente incorretas. Após os testes, não foram constatados problemas graves, apenas pequenos ajustes, já realizados. Portanto, a API REST demonstrou ser capaz de permitir a integração entre o e-SUS AB PEC e o *Dashboard* de Doenças Cerebrovasculares.

4.2 APRIMORAMENTO DA ANÁLISE PREDITIVA

A API REST exerce função fundamental para a integração entre os sistemas. É a partir desta ferramenta que o e-SUS AB PEC envia os dados para o *dashboard*. Então, alguns processamentos são realizados e os resultados exibidos nas telas da aplicação. A análise preditiva é um destes processamentos. Tem a responsabilidade de prever a probabilidade de um cidadão desenvolver doenças cerebrovasculares a partir de um conjunto de dados existentes.

Hammes (2018) utilizou a regressão logística otimizada por um algoritmo de otimização, o Gradiente Descendente Estocástico (SGD) escrito na linguagem de programação *Python* para realizar a predição. As variáveis incluídas ao modelo preditivo são categorizadas assumindo valores zero ou um, sendo o modelo de regressão logística o mais utilizado para este cenário.

O algoritmo de Hammes (2018) conta com as variáveis apresentadas no conjunto H, o qual é definido a seguir.

$$H = \{\text{id, cns, nome, idade, sexo, raça_cor, obeso, ex_obeso, fumante, ex_fumante, alcool, ex_alcool, drogas, ex_drogas, hipertensao_arterial, diabetes, dcv, gestante}\} \quad (1)$$

Análogo à forma em que foi definido o conjunto H, também foi definido o conjunto TFP composto pelos fatores de risco segundo Truelsen, Begg & Mathers (2005); Feigin, Norrving & Mensah (2017); Parmar et al. (2015). A seguir, é apresentado o conjunto TPF.

$$\text{TFP} = \{\text{idade, sexo, colesterol_serico, fibrinogenio, fumante, dieta_alta_concentracao_gordura_saturada, dieta_alta_concentracao_sodio, dieta_alta_concentracao_acucar, alcool, sedentarismo, escolaridade, classe_social, raca_cor}\} \quad (2)$$

Após a definição dos conjuntos, foi realizada a operação matemática diferença entre TFP e H (matematicamente representado por TFP - H). O resultado da operação é apresentado a seguir no conjunto DIF.

$$\text{DIF} = \{\text{colesterol_serico, fibrinogenio, dieta_alta_concentracao_gordura_saturada, dieta_alta_concentracao_sodio, dieta_alta_concentracao_acucar, sedentarismo, escolaridade, classe_social}\} \quad (3)$$

Com o propósito de identificar quais dos fatores de risco pertencentes no conjunto DIF são registrados pelo sistema e-SUS AB PEC, foram analisadas todas as informações de todas as estruturas de dados presentes no Layout e-SUS AB de Dado e Interface (LEDI AB⁶), versão 3.2.1. Ao término desta análise, foi constatado que não há evidências de registros de dados para as variáveis “colesterol_serico”, “fibrinogenio”, “dieta_alta_concentracao_gordura_saturada”, “sedentarismo” e “classe_social” pertencentes ao conjunto DIF. O “colesterol_serico” foi encontrado na estrutura da Ficha de Atendimento Individual, no grupo de exames, porém a única informação presente é se este exame foi solicitado ou avaliado, não contemplando a informação do resultado do exame. Portanto, somente os elementos “dieta_alta_concentracao_sodio”, “dieta_alta_concentracao_acucar” e “escolaridade”

⁶ Layout e-SUS AB de Dados e Interface (LEDI AB): <https://integracao.esusab.ufsc.br/ledi/index.html>

foram adicionados ao conjunto de variáveis utilizadas pelo algoritmo de análise preditiva, conforme mostra a Figura 13. As demais variáveis foram desconsideradas neste estudo.

Figura 13 – Variáveis para análise preditiva do presente estudo.

$$V = \left\{ \begin{array}{l} \text{"ID", "CNS", "NOME", "IDADE", "SEXO", "RACA_COR", "OBESO",} \\ \text{"EX_OBESO", "FUMANTE", "EX_FUMANTE", "ALCOOL", "EX_ALCOOL",} \\ \text{"DROGAS", "EX_DROGAS", "HIPERTENSAO_ARTERIAL", "DIABETES",} \\ \text{"DCV", "GESTANTE", "DIETA_ALTAS_CONCENTRACOES_SODIO",} \\ \text{"DIETA_ALTAS_CONCENTRACOES_ACUCARES", "ESCOLARIDADE"} \end{array} \right\}$$

Fonte: Elaborado pelo autor (2019).

Nas subseções seguintes, serão apresentadas as variáveis “dieta_alta_concentracao_sodio”, “dieta_alta_concentracao_acucar” e “escolaridade”.

4.2.1 Variáveis Referentes à Dieta

Segundo o *Institute for Health Metrics and Evaluation* (IHME⁷, 2019), são consideradas dietas com altas concentrações de sódio e açúcares quando alcançarem os indicadores de consumo maior ou igual a 1g por dia e o consumo de qualquer bebida com pelo menos 50 calorias de açúcar por porção, incluindo bebidas carbonatadas, refrigerantes, bebidas energéticas, bebidas de frutas, mas excluindo sucos 100% de frutas e vegetais, respectivamente.

Conforme apresentado anteriormente, o *endpoint* de marcadores de consumo alimentar é o responsável pelas informações das variáveis relacionadas às dietas. No entanto, pode ser observado que não há um atributo claro suficiente para representar estes valores. Assim, inferiu-se que cidadãos com hábito de consumir embutidos e/ou salgadinhos possuem dieta com alta concentração de sódio. Enquanto que cidadãos com hábito de consumir bebidas doces e/ou doces possuem dieta com alta concentração de açúcar. Portanto, estas variáveis foram

⁷ *Institute for Health Metrics and Evaluation* (IHME, 2019): <http://www.healthdata.org/terms-defined>

categorizadas assumindo valores zero ou um, seguindo o padrão previamente adotado para variáveis com este comportamento.

4.2.2 Variável Escolaridade

A variável escolaridade possui um conjunto de valores predefinidos conforme o LEDI AB⁸ e apresentados na Tabela 1.

Tabela 1 - Valores e classificações da variável escolaridade

Código	Descrição	Classificação
51	Creche	Básica
52	Pré-escola (exceto CA)	Básica
53	Classe de alfabetização – CA	Básica
54	Ensino fundamental 1 ^a a 4 ^a séries	Básica
55	Ensino fundamental 5 ^a a 8 ^a séries	Básica
56	Ensino fundamental completo	Média
57	Ensino médio especial	Média
58	Ensino fundamental EJA – séries iniciais (supletivo 1 ^a a 4 ^a)	Básica
59	Ensino fundamental EJA – séries finais (supletivo 5 ^a a 8 ^a)	Básica
60	Ensino médio, médio 2 ^o ciclo (científico, técnico, etc)	Média
61	Ensino médio especial	Média
62	Ensino médio EJA (supletivo)	Média
63	Superior, aperfeiçoamento, especialização, mestrado e doutorado	Superior
64	Alfabetização para adultos (Mobral, etc)	Média
65	Nenhum	Básica

Fonte: Elaborado pelo autor (2019).

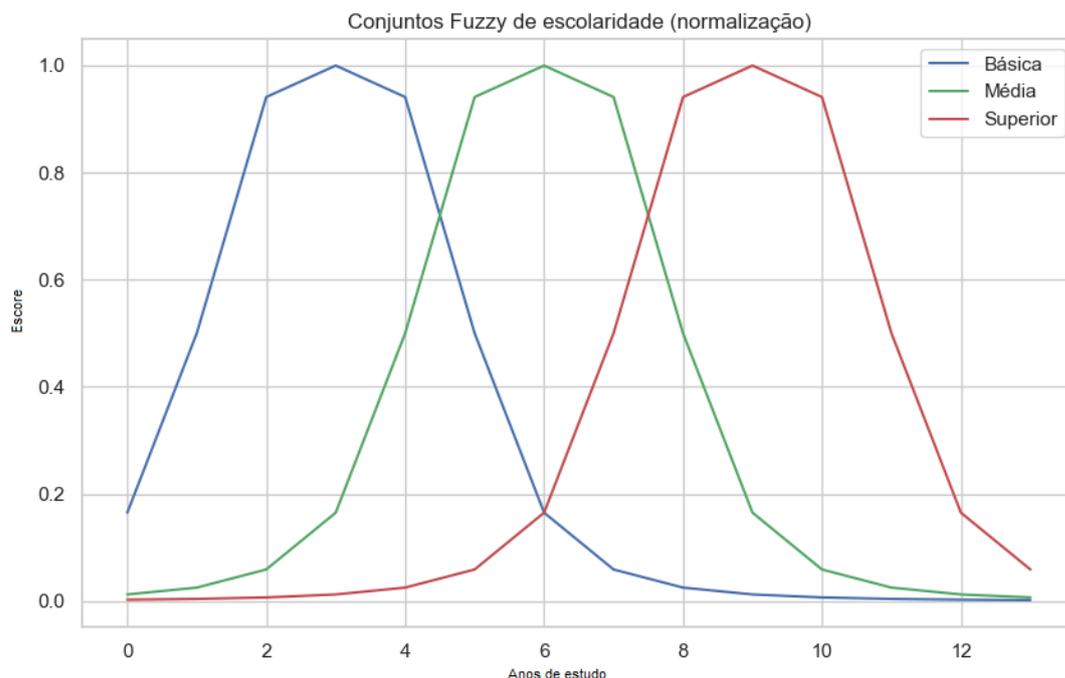
Para realizar a análise preditiva, foi necessário classificar esta variável escolaridade nas categorias básica, média e superior, além de converter esta informação em anos de estudo. A categoria básica engloba os códigos 51, 52, 53, 54, 55, 58, 59 e 65 e compreende o período inicial de alfabetização, totalizando nove anos de estudo. A categoria média compreende o período de três anos de estudo posteriores à categoria básica. Fazem parte da categoria média os códigos 56, 57,

⁸ Fonte de informação para o atributo escolaridade (<https://integracao.esusab.ufsc.br/ledi/referencias/dicionario.html#cursomaiselevado>)

60, 61, 62 e 64. Por fim, a categoria alta engloba o código 63 e compreendo os anos de estudo posteriores à categoria média. Para fins de atribuição de um limite, a categoria alta poderia ser contemplada com um valor máximo de dez anos.

Após a classificação da variável escolaridade em básica, média e superior, foi necessário normalizar os dados atribuindo valores entre zero e um para os anos de estudo apresentados em cada uma das categorias, conforme apresentado na Figura 14. Para realizar este processo, foi aplicada a teoria dos conjuntos Fuzzy. Esta teoria foi introduzida em 1965 pelo matemático Lotfi Asker Zadeh, com a intenção de dar um tratamento matemático a certos termos linguísticos subjetivos como: “aproximadamente”, “em torno de”, dentre outros. Pode-se dizer que a teoria dos conjuntos Fuzzy representa um primeiro passo no sentido de se programar e armazenar conceitos vagos em computadores, tornando possível a produção de cálculos com informações imprecisas, a exemplo do que faz o ser humano. (SOUZA, 2010).

Figura 14 – Normalização da variável escolaridade.



Fonte: Elaborado pelo autor (2019).

4.2.3 Base de Dados

A base de dados contém informações individualizadas de cidadãos. Para evitar expor a privacidade dos dados de saúde dos cidadãos que utilizam a rede de saúde pública brasileira, foi desenvolvido um gerador de dados fictícios para popular as tabelas. Isto proporcionou simular os registros de atendimentos realizados no sistema e-SUS AB PEC. Também permitiu que a análise probabilística de um cidadão desenvolver doenças cerebrovasculares fosse realizada com maior eficiência. Fato este inviável através da API REST, uma vez que cada requisição HTTP envia apenas um único registro e gerar uma grande massa de dados seria extremamente oneroso.

Apesar de que a geração dos dados ocorreu de maneira automática e aleatória, teve-se a preocupação de gerar uma base de dados com indicadores representativamente próximos da realidade. Para isto, cada atributo era moderado por um coeficiente de controle. Ao total, a base de dados gerada conta com 170.567 registros.

Com a implementação da API REST para receber os dados oriundos do e-SUS AB PEC, as estruturas de “views” criadas por Hammes (2018) tornam-se desnecessárias. As informações recebidas a partir desta versão do DDCV são processadas e persistidas diretamente na tabela histórico.

4.2.4 Treinamento do Algoritmo

O algoritmo foi implementado conforme o método Gradiente Descendente Estocástico. Este método faz uso apenas de uma amostra dos dados da população e representam o componente estocástico do algoritmo. Assim, é inviável definir critérios de parada do algoritmo baseados em diferenças de valores consecutivos, pois o traço da função objetiva deixa de ser suave. Portanto, estipula-se o número máximo de iterações que o algoritmo irá executar de forma a garantir suficiente proximidade com o ponto de ótimo da função (ZEVIANI; FERREIRA, 2017).

Para fins de análise dos resultados obtidos entre o algoritmo implementado por Hammes (2018) e o implementado neste estudo, foram realizados os respectivos treinamentos com o mesmo extrato da base de dados. A diferença está no conjunto de variáveis utilizadas em cada um dos algoritmos. Especificamente, neste estudo, foram adicionadas as variáveis “dieta_alta_concentracao_sodio”, “dieta_alta_concentracao_acucar” e “anos_estudo”, as quais não estavam presentes no trabalho de Hammes (2018).

O treinamento do algoritmo foi realizado com os registros cuja variável “dcv” possui valor um. Desta forma, o algoritmo identifica os fatores que contribuem para o desenvolvimento de doenças cerebrovasculares. Também foram incluídos ao conjunto de dados os cidadãos mais velhos, sem a doença, com hábitos alimentares saudáveis e com alta escolaridade. Este segundo conjunto tem o objetivo de informar ao algoritmo quais características levam a uma baixa chance de ter a doença. O conjunto final usado no treinamento totaliza 13.561 cidadãos. Finalmente, é analisado o conjunto de dados restante, estimando a probabilidade de um cidadão desenvolver esta situação clínica.

4.3 GEORREFERENCIAMENTO

A ferramenta desenvolvida por Hammes (2018) fez uso Google® *Geocoding* API, oferecido pela companhia Google® para realizar a conversão de endereços presentes na base de dados em coordenadas geográficas, esse processo é chamado de geocodificação. (GOOGLE MAPS PLATAFORM, 2019). A versão gratuita desta API apresenta limitações relacionadas ao número máximo de requisições, podendo receber até 2.500 requisições gratuitas por dia.

Durante a fase de testes foi identificado que este número máximo de requisições permitidas não era suficiente para atender a demanda do DDCV. Portanto, buscou-se uma ferramenta alternativa que atendesse a necessidade. Após estudos realizados, foi selecionada a HERE® *Geocoder* API⁹.

A HERE® *Geocoder* API oferece as mesmas operações realizadas pela API do google. Porém, tem a vantagem de permitir realizar um número de requisições

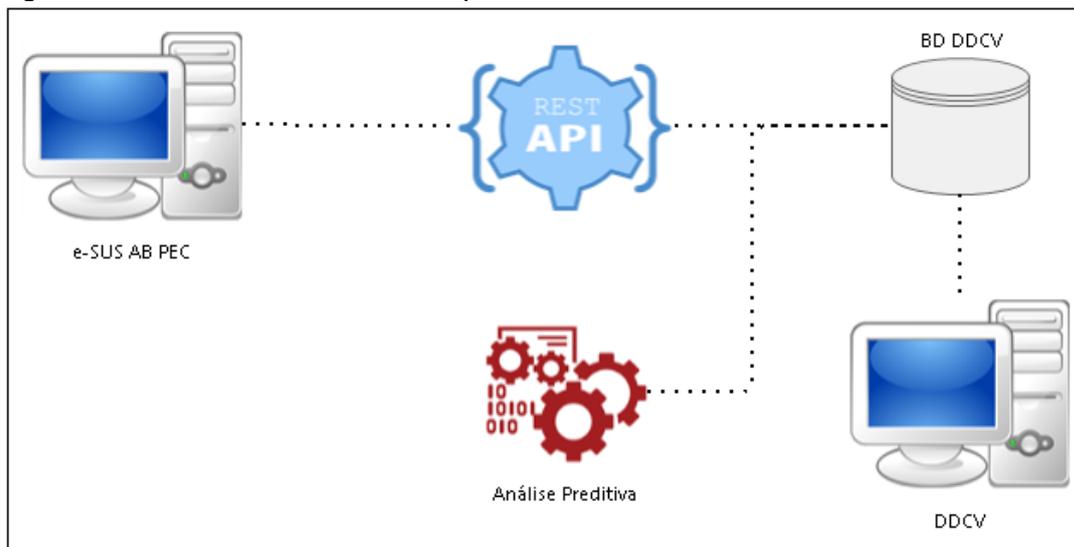
⁹ HERE Geocoder API: <https://developer.here.com/documentation/geocoder/topics/what-is.html>

muito superior, 250.000 requisições gratuitas por mês (HERE GEOCODER API, 2019). Após a realização dos testes, constatou-se que esta API atende a necessidade do *Dashboard* de Doenças Cerebrovasculares.

4.4 MECÂNICA DO DDCV

Concluída as etapas do desenvolvimento, podemos visualizar na Figura 15 duas grandes mudanças. A primeira, referente à etapa de exportar os dados do e-SUS AB PEC e importar no DDCV, a qual foi substituída e automatizada com a utilização da API REST. A segunda diz respeito à inexistência da etapa de exportação e importação dos dados no algoritmo de análise preditiva. Esta etapa foi totalmente automatizada pelo DDCV, passando a ser acessada por meio da interface do *dashboard*.

Figura 15 – Mecânica do DDCV do presente estudo.



Fonte: Elaborado pelo autor (2019).

Na versão desenvolvida por Hammes (2018), estes dois processos, bem como os passos executados em cada um deles, eram realizados manualmente. Além disso, no segundo, era necessário que o usuário tivesse conhecimento em linguagem SQL para desenvolver a consulta SQL e posteriormente salvar o resultado em um arquivo no formato CSV.

5 RESULTADOS

A inclusão das novas variáveis (“dieta_alta_concentracao_sodio”, “dieta_alta_concentracao_acucar” e “anos_estudo”) aprimorou o algoritmo de análise preditiva resultando no aumento da probabilidade média de um cidadão desenvolver doença cerebrovascular em 45,32%, passando de 47,59% para 69,16%, conforme apresenta a Tabela 2. No entanto, em 5.452 cidadãos a probabilidade diminuiu e em 798 o resultado foi o mesmo em ambas as análises. Estes números representam 3,20% e 0,47% da população utilizada neste estudo, respectivamente.

Tabela 2 - Resultados estatísticos de Hammes (2018) e do presente estudo

Indicador	Hammes (2018)	Presente estudo
Probabilidade mínima	24,00%	19,00%
Probabilidade máxima	87,00%	95,00%
Probabilidade média	47,59%	69,16%
Mediana	47,00%	70,00%
Moda	52,00% (15.198 vezes)	75,00% (8.611 vezes)
Desvio padrão	10,32	11,48
Variância	106,42	131,71

Fonte: Elaborado pelo autor (2019).

Neste estudo, os dados sobre a probabilidade estão mais dispersos devido à amplitude dos valores (probabilidade mínima 19,00% e máxima 95,00%). Isto justifica o desvio padrão de 11,48 ser maior com relação ao trabalho de Hammes (2018), 10,32. Em ambos os trabalhos, é possível observar que a mediana se encontra muito próxima da probabilidade média. Já a moda se distancia da média 0,43 desvios padrão no trabalho de Hammes (2018) e 0,51 desvios padrão neste.

A Tabela 3 exibe a distribuição da população entre as faixas percentuais. As faixas percentuais expressam a probabilidade de um cidadão desenvolver doenças cerebrovasculares.

Tabela 3 - Distribuição da população por faixa percentual segundo Hammes (2018) e do presente estudo

Faixa percentual	Número de cidadãos	
	Hammes (2018)	Presente estudo
Até 9%	0	0
De 10% a 19%	0	5
De 20% a 29%	5.878	341
De 30% a 39%	13.171	2.581
De 40% a 49%	79.534	7.450
De 50% a 59%	59.192	22.156
De 60% a 69%	2.757	46.839
De 70% a 79%	6.337	59.632
De 80% a 89%	3.698	29.064
Acima de 89%	0	2.499
Total	170.567	170.567

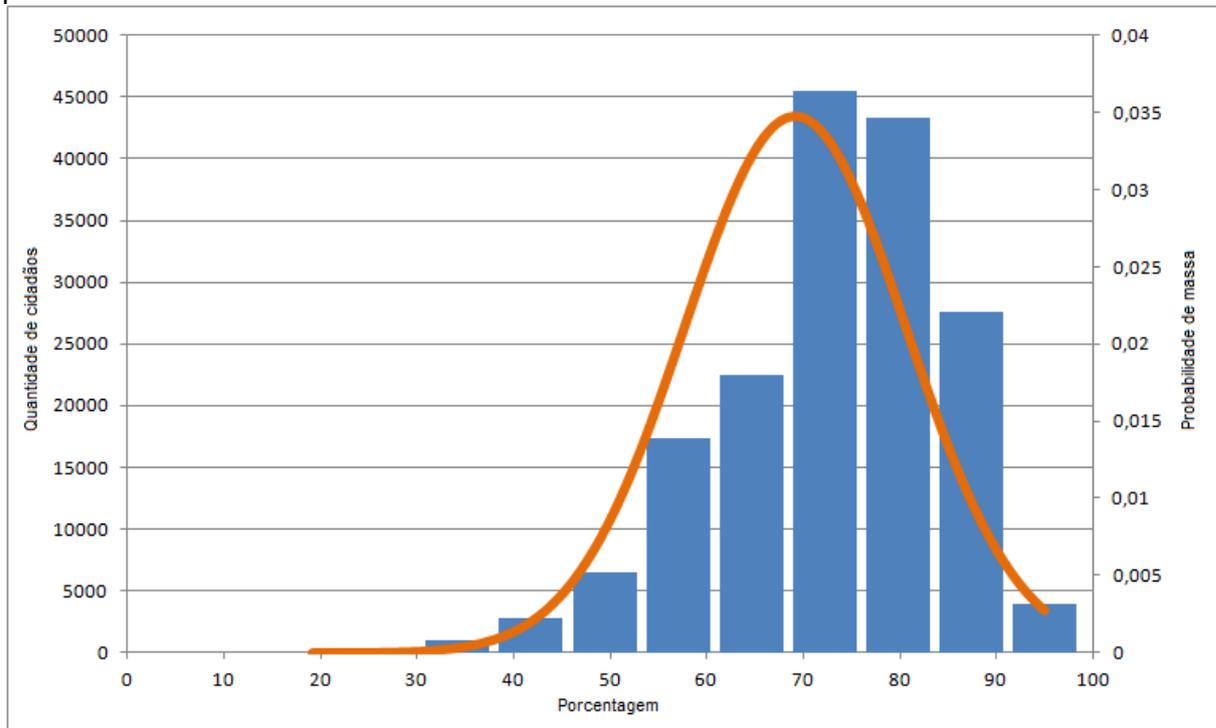
Fonte: Elaborado pelo autor (2019).

Utilizando o conjunto de variáveis de Hammes (2018), não foi identificado nenhum cidadão com risco nas faixas até 19,00% e acima de 89,00%. Para 81,33% dos cidadãos, o risco de desenvolvimento de doenças cerebrovasculares é de 40 a 59,00%.

Ao analisar os dados com a inclusão das variáveis sobre dietas com altas concentrações de açúcar e sódio e a escolaridade, apenas a faixa até 9,00% não apresenta nenhum cidadão. Para 62,42% dos indivíduos, o risco encontra-se entre 60 e 79,00%. A maior dispersão dos cidadãos entre as faixas percentuais sinaliza uma tendência à distribuição normal, conforme mostra a

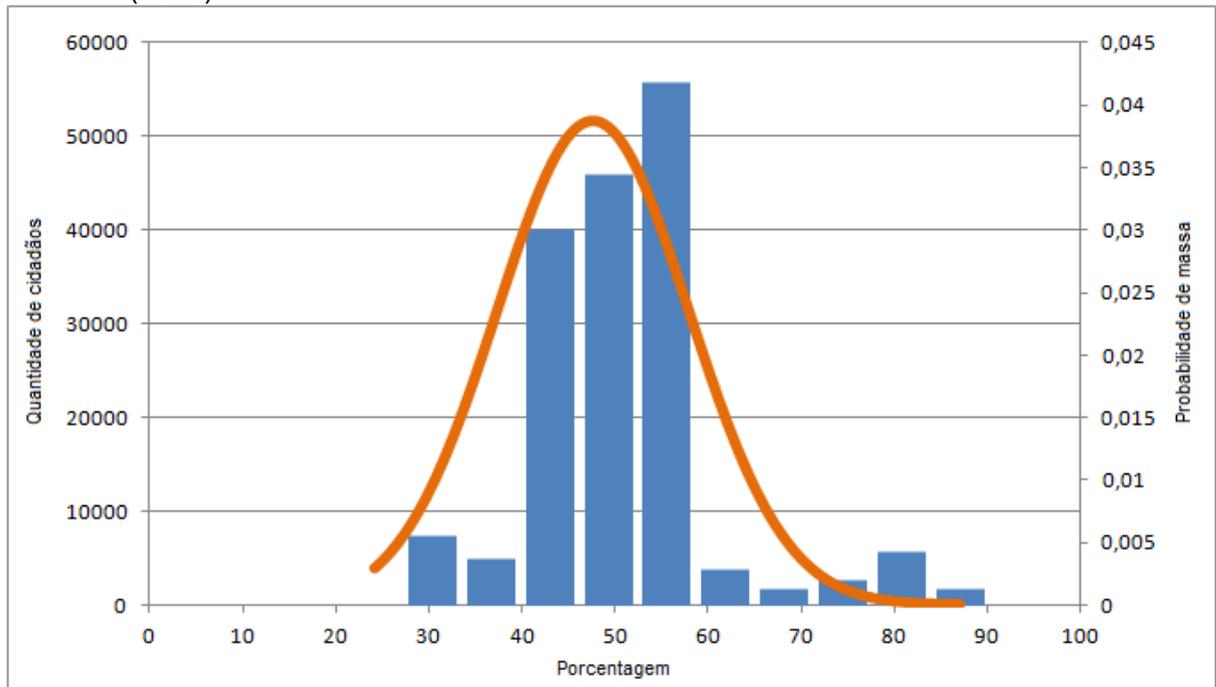
Figura 16. Este comportamento não é observado na proposta de Hammes (2018), conforme apresentado na Figura 17.

Figura 16 – Histograma de barras com distribuição normal para o conjunto de variáveis do presente estudo.



Fonte: Elaborado pelo autor (2019).

Figura 17 – Histograma de barras com distribuição normal para o conjunto de variáveis de Hammes (2018).



Fonte: Elaborado pelo autor (2019).

O resultado da análise preditiva possibilita avaliar cada uma das variáveis individualmente. Além da probabilidade de o cidadão desenvolver a doença, é possível identificar quais fatores o algoritmo utilizou para realizar a predição. A Tabela 4 apresenta o conjunto de variáveis utilizadas, o número de vezes em que a variável foi determinante para o resultado e a probabilidade mínima, máxima e média.

Tabela 4 - Influência das variáveis na análise preditiva do presente estudo

Variáveis	Número de predições	Probabilidade		
		Mínima	Máxima	Média
Faz uso de álcool	1.019	23,00%	93,00%	70,17%
Tem ou teve doença cerebrovascular	13.561	38,00%	95,00%	81,52%
Tem diabetes	4.565	22,00%	95,00%	67,63%
Dieta alta concentração açúcar	104.842	28,00%	95,00%	73,00%
Dieta alta concentração sódio	132.814	31,00%	95,00%	72,15%
Faz uso de drogas	739	21,00%	93,00%	70,41%
Está fumante	3.379	20,00%	94,00%	70,39%
Está gestante	1.449	36,00%	93,00%	70,38%
Tem hipertensão arterial	14.490	22,00%	95,00%	67,92%
Idade	7.127	21,00%	95,00%	68,56%
Escolaridade básica	14.437	36,00%	95,00%	74,86%
Escolaridade média	11.036	31,00%	94,00%	70,69%
Escolaridade superior	9.621	19,00%	88,00%	57,80%
Obesidade	5.623	21,00%	94,00%	68,95%

Fonte: Elaborado pelo autor (2019).

Exceto a variável escolaridade superior, todas as demais interferiram com probabilidades acima de 90,00% em pelo menos um indivíduo. A probabilidade média destaca a importância de todas as variáveis no processo de análise preditiva. As variáveis escolaridade superior e DCV são discrepantes da média geral (69,16%) apresentando probabilidade média de 57,80% e 81,52%, respectivamente. As probabilidades mínimas oscilaram consideravelmente. Novamente, as variáveis escolaridade superior e DCV apresentaram as porcentagens extremas, 19,00% e 38,00%, respectivamente. É válido destacar que as variáveis gestante e escolaridade baixa apresentaram probabilidade mínima de 36,00%.

Conforme esperado, o fator DCV é determinante para a predição. Apresenta os maiores percentuais entre todas as variáveis. A variável escolaridade superior apresentou as menores probabilidades, 19,00% para mínima, 88,00% para máxima e 57,80% em média. Estes indicadores corroboram com a afirmação dos autores Lotufo et al. (2017) de que quanto maior o grau de instrução dos cidadãos, menores são as chances de desenvolverem doenças cerebrovasculares.

A variável escolaridade baixa apresentou o maior índice, 95,00%. Este fator de risco pode estar relacionado ao local de residência do cidadão, normalmente comunidades carentes, com infraestrutura precária, sem condições adequadas de saneamento básico e acesso à informação. Outra característica provável é o ingresso no mercado de trabalho muito cedo para complementar a renda familiar, abdicando dos estudos. Isto prejudica o crescimento intelectual, dificulta o acesso às instruções sobre os malefícios de uma alimentação não saudável, com a ingestão de embutidos, salgadinhos, doces e alimentos processados.

Dietas com altas concentrações de sódio e açúcar foram importantes para a análise probabilística de um cidadão desenvolver doenças cerebrovasculares. Quando presentes entre as variáveis que interferiram diretamente no resultado, os piores índices foram 31,00% para dieta hipersódica e 28,00% para dieta alto teor de açúcar. Com relação aos melhores índices, ambas as variáveis alcançaram a porcentagem máxima da predição, 95,00% de probabilidade. As médias destas variáveis ficaram acima da média geral, 73,00% para dieta com alta concentração de açúcar e 72,15% para dieta com alta concentração de sódio.

Os altos índices encontrados nas variáveis de dietas com alta concentração de sódio e açúcar vão ao encontro das afirmações nas quais apontam que o consumo excessivo de sódio e açúcar aumenta o risco de ter um acidente vascular cerebral (STROKE ASSOCIATION, 2015; HANKEY, 2017; FOROUGHFI et al., 2013). Particularmente, o consumo exagerado de sódio potencializar o desenvolvimento de hipertensão arterial, principal variável causadora de doenças cerebrovasculares (STROKE ASSOCIATION, 2015).

A variável obesidade aparece praticamente com a mesma média (68,95%) que a média geral, não tendo destaque com relação às demais. No entanto, nos próximos anos, há probabilidade deste fator influenciar de maneira mais aguda no processo de análise preditiva. Isto pode ser inferido devido ao crescimento de

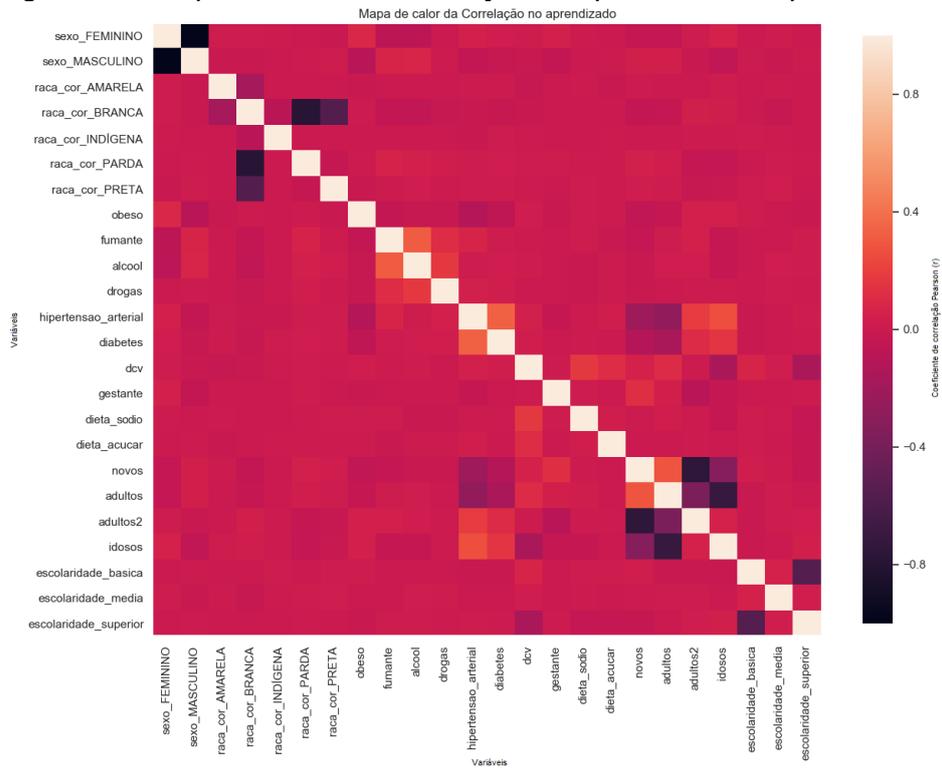
67,80% nos últimos treze anos da prevalência de obesidade entre os brasileiros, saindo de 11,80% em 2006 para 19,80% em 2018 (BRASIL, 2019b).

Em probabilidade e estatística, correlação é qualquer relação estatística (causal ou não causal) entre duas variáveis (BUSSAB; MORETTIN, 2010). Em relacionamentos correlacionais não se tem nenhum controle sobre as variáveis estudadas. As variáveis são observadas sem interferência externa, de maneira natural. A análise de correlação fornece um número, denominado coeficiente de correlação Pearson (r), que resume o grau de relacionamento entre as duas variáveis.

O coeficiente de correlação Pearson (r) pode variar de -1 a 1. Um coeficiente 1 indica a existência de correlação linear positiva perfeita. Inversamente, um coeficiente -1 indica a existência de correlação linear negativa perfeita. Um coeficiente de correlação zero significa que não existe relação linear entre as variáveis. Todavia, como valores extremos dificilmente são encontrados na prática, Cohen (1988), interpreta os escores entre 0,10 e 0,29 como correlação fraca, escores entre 0,30 e 0,49 como correlação média e valores entre 0,50 e 1 podem ser interpretados como correlações fortes.

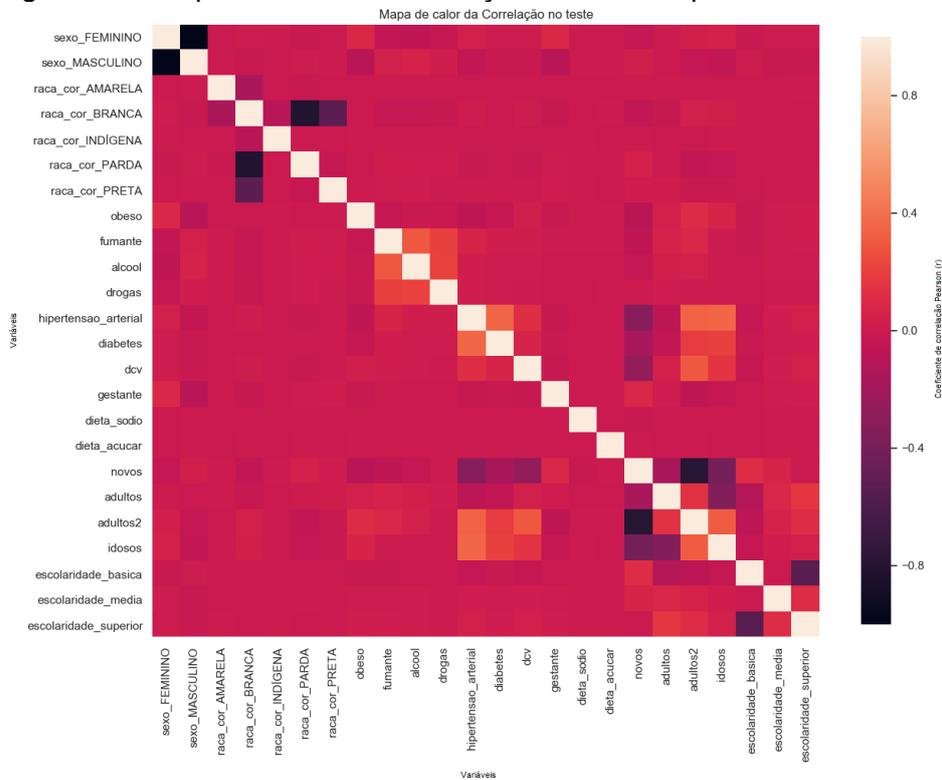
Observar a correlação entre as variáveis de treinamento e de teste é uma etapa importante dentro do processo de análise preditiva utilizando regressão logística otimizada por um algoritmo de otimização. A correlação pode mostrar se há relacionamento entre as variáveis, além de detectar falhas nos dados. Caso exista diferença considerável entre o treinamento e o teste, isto seria um problema para o algoritmo. Conforme pode ser observado nas Figura 18 e Figura 19, o mapa de calor da correlação apresenta grande semelhança.

Figura 18 – Mapa de calor da correlação no aprendizado do presente estudo.



Fonte: Elaborado pelo autor (2019).

Figura 19 – Mapa de calor da correlação no teste do presente estudo.



Fonte: Elaborado pelo autor (2019).

Os mapas de calor da correlação no aprendizado (Figura 18) e no teste (Figura 19) mostram de forma visual a relação negativa entre sexo masculino e feminino. Isto comprova que os dados não apresentam erro em relação a essas colunas. A variável DCV apresentou correlação positiva fraca com a variável hipertensão arterial. Isto indica que essas duas variáveis impactam de forma semelhante no resultado.

No aprendizado, a variável DCV quando comparada com as variáveis escolaridades básica e média apresentou correlação neutra, $r = 0,09$ e $r = 0,03$, respectivamente. Já ao avaliar com a escolaridade superior, apresentou correlação negativa fraca, $r = -0,15$. Porém, a mesma variável, comparada com o teste, apresentou correlação neutra para a escolaridade básica, $r = -0,02$, permaneceu com mesmo escore para a escolaridade média e também apresentou correlação neutra para a escolaridade superior, $r = 0,05$.

As variáveis referentes às dietas com altas concentrações de sódio e açúcar apresentaram comportamento muito similar ao da variável escolaridade. No aprendizado, quando avaliada a variável DCV com dieta com alta concentração de sódio e dieta com alta concentração de açúcar apresentou coeficientes de correlação positiva fraca, $r = 0,16$ e $r = 0,12$, respectivamente. No entanto, no teste, estas duas variáveis se mostraram neutras ($r = 0$) com relação a todo conjunto de variáveis, incluindo a variável DCV.

Embora existam pequenas diferenças no coeficiente de correlação Pearson (r), não faz sentido interpretar uma correlação de 0,15 como sendo 15%. Também não é possível afirmar que uma correlação de 0,6 é o dobro de uma correlação de 0,3. O coeficiente tem um caráter adimensional, isto é, é desprovido de unidade física que o defina (CHEN; POPOVICH, 2002). Portanto, neste cenário de teste, as variáveis de escolaridade e dietas com altas concentrações de açúcar e sódio apresentam coeficientes de correlação de Pearson (r) muito próximos à zero. Isto indica apenas a ausência de relacionamento linear.

6 CONCLUSÃO

O DDCV é uma ferramenta que contribui para o processo gerencial de tomada de decisão, suprindo uma das muitas carências relacionadas ao conhecimento das condições de saúde da população que utilizam o SUS. Contudo, provêm insumos para os gestores atuarem de maneira preventiva no tratamento de cidadãos com altas chances de apresentarem doenças cerebrovasculares. A partir de iniciativas como esta é possível reduzir índices registrados para as doenças crônicas não transmissíveis, atualmente um dos maiores no Brasil e também no mundo.

A estratégia e-SUS AB vem se consolidando ano após ano. Atualmente, 100% dos municípios brasileiros estão usufruindo de pelo menos um dos sistemas da estratégia, e-SUS AB CDS ou e-SUS AB PEC. Aproximadamente 2.770 municípios estão utilizando o principal sistema fornecido pelo Ministério da Saúde, o e-SUS AB PEC. Entretanto, os sistemas da estratégia e-SUS AB são limitados de relatórios e ferramentas para auxiliar os gestores na tomada de decisão e cuidado continuado do cidadão.

Nos dias atuais, existem muitos sistemas centrados em finalidades específicas e que poderiam potencializar suas funcionalidades a partir da comunicação entre eles. A API desenvolvida neste trabalho supriu esta lacuna criando mecanismos de comunicação entre os sistemas e-SUS AB PEC e o *Dashboard* de Doenças Cerebrovasculares.

Esta API vai ao encontro das diretrizes da e-Saúde no que diz respeito à agilidade e melhora da qualidade da atenção e dos processos de saúde. Torna o DDCV minimamente comunicável com a APS através de uma integração rápida, leve e padronizada, fornecendo insumos e análises em tempo real para os gestores, enriquecendo ainda mais o principal sistema da estratégia ministerial. Além disso, a interoperabilidades entre os sistemas traz benefícios de melhora na qualidade do cuidado ao cidadão.

O aprimoramento do algoritmo de análise preditiva vai ao encontro do principal objetivo da atenção primária, que é atuar de modo preventivo auxiliando os profissionais de saúde no planejamento do cuidado ao paciente. Esta ação reflete diretamente na redução dos custos com tratamentos que poderiam ter sido evitados

se medidas preventivas tivessem sido tomadas. Contudo, é interessante realizar estudos sobre o tamanho da amostra para avaliar a real efetividade do resultado. Em conjunto, também pode ser realizado o treinamento do algoritmo com diferentes tamanhos de amostra, buscando identificar qual o tamanho ideal para melhor estimar a probabilidade de cidadãos desenvolverem doenças cerebrovasculares, reduzindo assim a massa de dados utilizada na predição, uma vez que ela implica diretamente no desempenho do algoritmo. Seria importante avaliar outros algoritmos que desempenham a mesma função e analisar os resultados encontrados, comparando com os resultados obtidos neste estudo.

A inclusão das variáveis escolaridade, dieta com alta concentração de açúcar e dieta com alta concentração de sódio fez com que o algoritmo aprimorasse o processo de análise preditiva, resultando no aumento da probabilidade média de um cidadão desenvolver doença cerebrovascular em 45,32%, passando de 47,59% para 69,16%. Isto prova a importância de analisar estes fatores quando se trata de doenças crônicas não transmissíveis, como é o caso das doenças cerebrovasculares. Contudo, estudos ainda precisam ser realizados, principalmente no que diz respeito ao uso de dados fictícios. Embora alguns parâmetros tenham sido considerados no momento da construção da base de dados, é possível que os dados gerados não representem a realidade da situação de saúde dos municípios brasileiros.

Os dados utilizados para a análise preditiva tem origem apenas do sistema e-SUS AB PEC. Isto representa uma pequena parcela de informações de saúde quando comparada ao conjunto de dados registrados pelas unidades de pronto atendimento, clínicas e hospitais. Igualmente, alguns fatores importantes para as condições de risco de doenças cerebrovasculares não puderam ser avaliadas na análise preditiva, pois não existem de maneira estruturada no modelo de informação do LEDI AB, por exemplo, colesterol, sedentarismo, fibrinogênio e dieta com alta concentração de gordura saturada. A interoperabilidade das informações em saúde entre os diferentes níveis dos serviços de saúde (Atenção Primária à Saúde, Média e Alta Complexidade), seja públicos ou privados, pode fortalecer o cuidado continuado

e a confiabilidade na predição não só de doenças cerebrovasculares, mas também de outras condições de saúde.

O desenvolvimento deste estudo agregou valor ao *Dashboard* de Doenças Cerebrovasculares desenvolvido por Hammes (2018). Aprimorou o processo de importação dos dados para o DDCV, aumentou o índice de confiança do algoritmo utilizado na análise preditiva, resultando na probabilidade de cidadãos desenvolverem doenças cerebrovasculares, apresentou o mapa de calor das regiões com maior probabilidade de apresentarem doenças cerebrovasculares, melhorou a funcionalidade de mapeamento georreferenciado, além de permitir a integração entre o e-SUS AB PEC e o DDCV.

Com estes dois sistemas comunicando através de uma API simples e prática, poderão ser disponibilizados insumos aos gestores municipais e profissionais de saúde sobre a condição de saúde da população no que diz respeito às doenças cerebrovasculares. Isto auxilia os profissionais no planejamento, bem como na tomada de decisões relacionadas à gerência e à assistência aos cidadãos, fornecendo informações adequadas. Contudo, busca-se maior eficiência gestora na orientação e direcionamento das ações e tomadas de decisões através da qualificação e uso da informação na gestão e no cuidado em saúde.

Para perspectiva futura, é relevante testar o DDCV em uma base de dados real, orientado pela Lei Geral de Proteção de Dados Pessoais nº 13.709, de 14 de Agosto de 2018 e pela Resolução nº 466/12, do Conselho Nacional de Saúde no qual aprova as diretrizes e normas regulamentadoras de pesquisa envolvendo seres humanos. Desta forma será possível validar as funcionalidades do DDCV, principalmente no que diz respeito à análise preditiva. Seguramente com acesso restrito a profissionais de saúde, destacando-se o quão delicado é utilizar somente do aprendizado de máquina para informar ao cidadão a chance de contrair um possível AVC ou outra CID-10 deste mesmo grupo de doenças crônicas não transmissíveis. Salienta-se que esta ferramenta deve ser usada como um instrumento de apoio ao cuidado do cidadão. Jamais deve ser usada como única fonte de alerta de riscos de doenças cerebrovasculares. Em nenhuma hipótese, o DDCV cobre todo o conjunto de fatores e avaliações clínicas realizadas pelos profissionais de saúde.

Com relação à API, fazendo alguns ajustes, será possível interoperar com outros sistemas. Receber dados de saúde, não só do sistema de prontuário eletrônico da Atenção Primária à Saúde (e-SUS AB PEC), mas também de sistemas da Média e Alta Complexidade, públicos ou privados, enriquecerá ainda mais o algoritmo de análise preditiva. Além de possibilitar a inclusão de novas variáveis, por exemplo, colesterol, sedentarismo, fibrinogênio e dieta com alta concentração de gordura saturada, aumentando a confiabilidade da probabilidade de um cidadão desenvolver doenças cerebrovasculares e o aprendizado do algoritmo. A API também poderá assumir novas funções, por exemplo, retornar a probabilidade de um cidadão desenvolver doenças cerebrovasculares por meio da implementação do método *get*, pertencente ao protocolo de comunicação HTTP.

Por fim, seria de grande valia refatorar a interface do *Dashboard* de Doenças Cerebrovasculares utilizando o padrão *Bold Design System*¹⁰. Este padrão contempla um estilo de componentes totalmente focado em usabilidade, acessível e aberto para a comunidade. Foi construído conforme as especificações da WCAG¹¹ nível AA.

¹⁰ *Bold Design System*: <https://bold.bridge.ufsc.br/>

¹¹ *Web Content Accessibility Guidelines* (WCAG): <https://www.w3.org/WAI/standards-guidelines/wcag/>

REFERÊNCIAS

_____. Portaria MS/GM nº 1.412, de 10 de julho de 2013. **Institui o Sistema de Informação em Saúde para a Atenção Básica (SISAB)**. Diário Oficial da União, Brasília-DF, Seção 1. p 294, 11 jul. 2013a. Disponível em: <https://www.jusbrasil.com.br/diarios/56517645/dou-secao-1-11-07-2013-pg-294?ref=serp>. Acesso em: 23 jul. 2018.

_____. Portaria MS/GM nº 1.653, de 2 de outubro de 2015. **Acrescenta o art. 2º-A à Portaria nº 1.412/GM/MS, de 10 de julho de 2013, que Institui o Sistema de Informação em Saúde para a Atenção Básica (SISAB), com o objetivo de garantir a transição entre o Sistema de Registro das Ações Ambulatoriais de Saúde (RAAS) e o SISAB**. Diário Oficial da União, Brasília-DF, Seção 1. p 673, 5 out. 2015. Disponível em: <https://www.jusbrasil.com.br/diarios/101473260/dou-secao-1-05-10-2015-pg-673?ref=serp>. Acesso em: 23 jul. 2018.

_____. Portaria MS/GM nº 221, de 17 de abril de 2008. **Lista Brasileira de Internações por Condições Sensíveis à Atenção Primária**. Diário Oficial da União, Brasília-DF, Seção 1. p 70, 18 abr. 2008. Disponível em: <https://www.jusbrasil.com.br/diarios/546710/pg-70-secao-1-diario-oficial-da-uniao-dou-de-18-04-2008?ref=serp>. Acesso em: 03 mar. 2019.

_____. Ministério da Saúde. Conselho Nacional de Saúde. Resolução nº 466, de 12 de dezembro de 2012. **Aprova Normas Regulamentadoras de Pesquisas Envolvendo Seres Humanos**. Diário Oficial da União, Brasília-DF, Seção 1. p 59, 13 jun. 2013b. Disponível em: <https://www.jusbrasil.com.br/diarios/55483111/dou-secao-1-13-06-2013-pg-59?ref=serp>. Acesso em: 25 mar. 2019.

_____. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Diário Oficial da União, Brasília-DF, p 59, 15 ago. 2018. Disponível em: <https://presrepublica.jusbrasil.com.br/legislacao/612902269/lei-13709-18>. Acesso em: 26 ago. 2019.

BRASIL. Ministério da Saúde. Secretaria de Atenção Primária à Saúde. **Manual de Uso do Sistema com Prontuário Eletrônico do Cidadão PEC v3.2**. Brasília, 2019a. Disponível em: http://aps.saude.gov.br/ape/esus/manual_3_2/. Acesso em: 16 jul. 2019.

CAVALCANTI, Pauline Cristine da Silva; OLIVEIRA NETO, Aristides Vitorino de; SOUSA, Maria Fátima de. **Quais são os desafios para a qualificação da Atenção Básica na visão dos gestores municipais?**. Saúde debate, Rio de Janeiro, v. 39, n. 105, p. 323-336, Junho de 2015. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-11042015000200323&lng=en&nrm=iso. Acesso em: 03 set. 2018. <http://dx.doi.org/10.1590/0103-110420151050002323>.

CAVALCANTE, Ricardo Bezerra et al. **Computerization of Primary Health Care Information Systems: Advances and Challenges**. Cogitare Enfermagem, [S.l.], v.

23, n. 3, ago. 2018. ISSN 2176-9133. Disponível em:
<https://revistas.ufpr.br/cogitare/article/view/54297>. Acesso em: 10 set. 2018.
 doi:<http://dx.doi.org/10.5380/ce.v23i3.54297>.

SUMAR, Norhan; FAUSTO, Márcia Cristina Rodrigues. **Atenção Primária à Saúde: a construção de um conceito ampliado**. Journal of Management & Primary Health Care, Pernambuco, v. 5, n. 2, p. 202-212, 2014.

SOARES, Eva Vilma Barbosa. **Atenção básica e informação: análise do Sistema de Informação em Saúde para Atenção Básica (SISAB) e estratégia e-SUS AB e suas repercussões para uma gestão da saúde com transparência**. 2016. 42 f., il. Trabalho de conclusão de curso (Especialização em Gestão Pública na saúde)— Universidade de Brasília, Brasília, 2016.

SIMON, Herbert Alexander. **The Sciences of the Artificial**. Cambridge: MIT Press, 1996.

LACERDA, Daniel Pacheco et al. **Design Science Research: método de pesquisa para a engenharia de produção**. Gest. Prod., São Carlos, v. 20, n. 4, p. 741-761, 2013. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X2013000400001&lng=en&nrm=iso. Acesso em: 18 set. 2018. Epub Nov 26, 2013. <http://dx.doi.org/10.1590/S0104-530X2013005000014>.

ÇAĞDAŞ, Volkan; STUBKJÆR, Erik. **Design research for cadastral systems**. Computers, Environment and Urban Systems, 35, p. 77-87, 2011.

PEFFERS, Ken et al. **A Design Science Research Methodology for Information Systems Research**. Journal of Management Information Systems, vol. 24, no. 3, 2007, pp. 45–77. JSTOR, JSTOR, www.jstor.org/stable/40398896.

WORLD HEALTH ORGANIZATION. **Building foundations for eHealth: progress of Member States**. Report of the WHO Global Observatory for eHealth. Geneva, 2007. Disponível em: https://www.who.int/goe/publications/build_foundations/en/. Acesso em: 17 fev. 2019.

BRASIL. Ministério da Saúde. Comitê Gestor da Estratégia e-Saúde. **Estratégia e-Saúde para o Brasil**. Brasília, 2017. Disponível em: <http://portalarquivos.saude.gov.br/images/pdf/2017/julho/12/Estrategia-e-saude-para-o-Brasil.pdf>. Acesso em: 17 fev. 2019.

HAMMES, Jades Fernando. **DASHBOARD E UM MODELO DE ANÁLISE PREDITIVA PARA DOENÇAS CEREBROVASCULARES NA ATENÇÃO PRIMÁRIA EM SAÚDE**. 2018. 97f. Dissertação (Mestrado Profissional em Informática em Saúde) – Programa de Mestrado Profissional em Informática em Saúde, Universidade Federal de Santa Catarina, Florianópolis, 2018.

BRASIL. **Prontuário Eletrônico chega a 57 milhões de brasileiros**. 2016. Disponível em: <http://datasus.saude.gov.br/noticias/atualizacoes/1073-prontuario-eletronico-chega-a-57-milhoes-de-brasileiros>. Acesso em: 18 fev. 2018.

Ca TECHNOLOGIES. **O Manual de Gerenciamento de APIs Compreendendo as soluções para o gerenciamento de APIs**. 2015.

GILL, Martin; HEFFNER, Randy. **Brief: Four Ways APIs Are Changing Your Business**. How to Design an API Strategy that Fuels Digital Business. Forrester Research, Inc. 17 Jun. 2015.

EUA. Department of Health and Human Services. National Coordinator for Health Information Technology. **About APIs**. Washington D.C., 2016. Disponível em: https://www.healthit.gov/api-education-module/story_content/external_files/hhs_transcript_module.pdf. Acesso em: 24 fev. 2019.

JONES, Collin. **APIs in healthcare: what healthcare providers should know about APIs**. Corepoint Health. 20 Setembro. 2016. Disponível em: <https://corepointhealth.com/resource-center/white-papers/api-primer/>. Acesso em: 25 fev. 2019.

DELOITTE DEVELOPMENT LLC. **The API Imperative in Life Sciences and health care. From IT Concern to Business Mandate to Drive Digital Acceleration**. Tech Trends 2018. A Life Sciences and Health Care Perspective. 2018. Disponível em: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-api-imperative-in-life-sciences-and-health-care.pdf>. Acesso em: 26 fev. 2019.

LOTUFO, Paulo Andrade et al. **Cerebrovascular disease in Brazil from 1990 to 2015: Global Burden of Disease 2015**. Rev. bras. epidemiol. São Paulo, v. 20, supl. 1, p. 129-141, Maio 2017. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2017000500129&lng=en&nrm=iso. Acesso em: 03 mar. 2019. <http://dx.doi.org/10.1590/1980-5497201700050011>.

WANG, Haidong et al. **Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015**. GBD 2015. Mortality and Causes of Death Collaborators. The Lancet, v. 388, Issue 10053. p. 1459-1544, Oct 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0140673616310121>. Acesso em: 03 mar. 2019. [https://doi.org/10.1016/S0140-6736\(16\)31012-1](https://doi.org/10.1016/S0140-6736(16)31012-1).

DUNCAN, Bruce Bartholow et al. **Doenças crônicas não transmissíveis no Brasil: prioridade para enfrentamento e investigação**. Rev. Saúde Pública, São Paulo, v. 46, supl. 1, p. 126-134, Dez. 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-

89102012000700017&lng=en&nrm=iso. Acesso em: 03 mar. 2019.
<http://dx.doi.org/10.1590/S0034-89102012000700017>.

RIBEIRO, Antonio Luiz P. et al. **Cardiovascular Health in Brazil: Trends and Perspectives**. *Circulation*, v. 133, No. 4, p. 422-433, Jan. 2016. Disponível em: <https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.114.008727>. Acesso em: 03 mar. 2019. <https://doi.org/10.1161/CIRCULATIONAHA.114.008727>.

ROTH, Gregory A. et al. **Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015**. *Journal of the American College of Cardiology*. v 70, Issue 1, p. 1-25, Julho 2017. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0735109717372443>. Acesso em: 05 mar. 2019. <https://doi.org/10.1016/j.jacc.2017.04.052>.

MOZAFFARIAN, D. **Global Scourge of Cardiovascular Disease: Time for Health Care Systems Reform and Precision Population Health**. *Journal of the American College of Cardiology*, v. 70, n. 1, p. 26–28, 2017.

UNITED NATIONS. **Sustainable Development Goals: 17 Goals to Transform our World**. 2015. Disponível em: <http://www.un.org/sustainabledevelopment/>. Acesso em: 05 mar. 2019.

TRUELSEN, T.; BEGG, S.; MATHERS, C. **The global burden of cerebrovascular disease**. *Global Burden of Disease*, p. 1–67, 2005.

SES-SC. **Secretaria de Estado da Saúde de Santa Catarina. Internações por Condições Sensíveis à Atenção Primária – ICSAP. 2014**. Disponível em: http://portalses.saude.sc.gov.br/index.php?option=com_content&view=article&id=2622&Itemid=489. Acesso em: 05 mar. 2019.

BRASIL. **Acidente Vascular Cerebral (AVC): Uma das principais causas de mortes no mundo, doença pode ser prevenida com hábitos saudáveis de vida**. 2012. Disponível em: <http://www.brasil.gov.br/saude/2012/04/acidente-vascular-cerebral-avc>. Acesso em: 05 mar. 2019.

NYCE, Charles. **Predictive Analytics**. White Paper. American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America. 2007. Disponível em: <https://www.the-digital-insurer.com/wp-content/uploads/2013/12/78-Predictive-Modeling-White-Paper.pdf>. Acesso em: 10 mar. 2019.

FINLAY, Steven. **Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods**. 2014. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=LDntAwAAQBAJ&oi=fnd&pg=PP1&dq=Predictive+Analytics,+Data+Mining+and+Big+Data.+Myths,+Misconceptions+and+Methods&ots=pkrR_d4UAc&sig=hYCLi4scZFt1pHxQWzNoMuv8j6c#v=onepage&q&f=false. Acesso em: 10 mar. 2019.

FEIGIN, Valery L.; NORRVING, Bo; MENSAH, George A. **Global Burden of Stroke. Circulation Research**. vol. 120, no. 3, Feb 2017, pp. 439–448. Disponível em: <https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.116.308413>. Acesso em: 11 mar. 2019. <https://doi.org/10.1161/CIRCRESAHA.116.308413>.

PARMAR, P. et al. **The Stroke Riskometer™ App: validation of a data collection tool and stroke risk predictor**. Stroke Riskometer™ Collaboration. International Journal of Stroke. Feb 2015. v 10, ed 2. p 231– 244. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4335600/>. Acesso em: 11 mar. 2019. doi: 10.1111/ijss.12411

FIELDING, Roy T.; TAYLOR, Richard N. **Architectural styles and the design of network-based software architectures**. Irvine, USA: University of California, Irvine, 2000.

RODRIGUEZ, Alex. **RESTful Web Services: Learn the Basics Principles of REST Web Services. IBM Developer**. Bengaluru. Nov 2008. Disponível em: <https://developer.ibm.com/articles/ws-restful/>. Acesso em: 13 mar. 2019.

BARTLEY, Andrew. **Predictive Analytics in Healthcare: A Data-Driven Approach to Transforming Care Delivery**. White Paper. Intel Corporation. 2017. Disponível em: <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/gmc-analytics-healthcare-whitepaper.pdf>. Acesso em: 26 mar. 2019.

PALEM, Gopalakrishna. **The Practice of Predictive Analytics in Healthcare**. Abril 2013. Disponível em: https://www.researchgate.net/publication/236336250_The_Practice_of_Predictive_Analytics_in_Healthcare. Acesso em: 26 mar. 2019.

YANG, Chengliang et al. **Machine Learning Approaches for Predicting High Utilizers in Health Care**. Rojas I., Ortuño F. Bioinformatics and Biomedical Engineering. IWBBIO 2017. Lecture Notes in Computer Science, vol 10209. p 382-395. Springer, Cham. Abril, 2017. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-56154-7_35. Acesso em: 28 mar. 2019. https://doi.org/10.1007/978-3-319-56154-7_35

Minnesota Department of Health (MDH). **An introductory analysis of potentially preventable health care events in Minnesota**. 2015. Disponível em: <https://www.leg.state.mn.us/docs/2015/other/150694.pdf>. Acesso em: 28 mar. 2019.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data mining: concepts and techniques**. Third edition. Amsterdam: Elsevier, 2012.740p.

NATARAJAN, Prashant; FRENZEL, John C; SMALTZ, Detlev.H. **Demystifying Big Data and Machine Learning for healthcare**. Boca Raton, FL: CRC Press Taylor & Francis Group, 2017. 210p

AXELROD, Randy C.; VOGEL, David. **Predictive Modeling in Health Plans.** Disease Management & Health Outcomes, v. 11, n. 12, p. 779-787, 01 de Dezembro de 2003. Disponível em: <https://link.springer.com/article/10.2165/00115677-200311120-00003>. Acesso em: 01 abr. 2019. <https://doi.org/10.2165/00115677-200311120-00003>.

KALECHOFISKY, Hal. **A Simple Framework for Building Predictive Models. A Little Data Science Business Guide.** M Squared Consulting and Solomon Edwards. Setembro de 2016. Disponível em: <http://www.msquared.com/wp-content/uploads/2017/01/A-Simple-Framework-for-Building-Predictive-Models.pdf>. Acesso em: 02 abr. 2019.

WAZLAWICK, Raul S. **Metodologia de Pesquisa para Ciência da Computação.** 1 ed. Rio de Janeiro: Elsevier Editora Ltda, 2009.

BARBELLA, F. et al. **How can eHealth improve care for people with multimorbidity in Europe?** POLICY BRIEF 25. Berlin, 2017.31p. Disponível em: http://www.icare4eu.org/pdf/PB_25.pdf. Acesso em: 15 abr. 2019.

SOUZA, Osmar do Nascimento. **Introdução à Teoria dos Conjuntos Fuzzy.** Março, 2010. Disponível em: <http://www.ime.unicamp.br/~valle/PDFfiles/osmar10.pdf>. Acesso em: 13 jul. 2019.

ZEVIANI, Walmes M.; FERREIRA, Eduardo V. **Métodos de Gradiente Descendente.** Setembro, 2017 Acesso em: 14 jul. 2019. Disponível em: <http://leg.ufpr.br/~walmes/ensino/ML/tutorials/02-gradient-methods.html#gradiente-descendente-estocastico>.

GOOGLE MAPS PLATAFORM. **Geocoding API, Get Started.** Junho, 2019. Disponível em: <https://developers.google.com/maps/documentation/geocoding/start>. Acesso em: 15 jul. 2019.

BUSSAB, Wilton de Oliveira; MORETTIN, Pedro Alberto. **Estatística Básica.** 6. ed. São Paulo: Saraiva, 2010. xvi, 540 p.

COHEN, Jacob. **Statistical Power Analysis for the Behavioral Sciences.** 2. ed. New York: Routledge, 1988. 567 p. <https://doi.org/10.4324/9780203771587>

CHEN, Peter Y.; POPOVICH, Paula M. **Correlation: Parametric and Nonparametric Measures.** 1. Ed. Thousand Oaks: SAGE Publication, 2002. 104 p.

BRASIL. Ministério da Saúde. **Vigitel Brasil 2018: Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico.** Brasília, 2019b. Disponível em: <http://portalarquivos2.saude.gov.br/images/pdf/2019/julho/25/vigitel-brasil-2018.pdf>. Acesso em: 25 jul. 2019.

- BOEHME, Amelia K.; ESENWA, Charles; ELKIND, Mitchell S. V. **Stroke Risk Factors, Genetics, and Prevention**. *Circulation Research*, v. 120, n. 3, p. 472-495, Fevereiro de 2017. Disponível em: <https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.116.308398>. Acesso em: 07 ago. 2019. <https://doi.org/10.1161/CIRCRESAHA.116.308398>.
- GRAHAM, B.; BOND, R.; QUINN, M.; MULVENNA, M. **Using Data Mining to Predict Hospital Admissions From the Emergency Department**. *IEEE Access*, v. 6, p. 10458-10469, 2018. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8300528>. Acesso em: 07 ago. 2019. <https://doi.org/10.1109/ACCESS.2018.2808843>.
- ASADI, H.; DOWLING, R.; YAN, B.; MITCHELL, P. J. **Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy**. *PLoS ONE* 9(2): e88225, 2014. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0088225>. Acesso em: 08 ago. 2019. <https://doi.org/10.1371/journal.pone.0088225>.
- LIANG, W.; KRISHNAMURTHI, R.; KASABOV, N.; FEIGIN, V. **Information Methods for Predicting Risk and Outcome of Stroke**. *Springer Handbook of Bio-/Neuroinformatics*. Springer Handbooks. Springer, Berlin, Heidelberg, p. 993-1001, 2014. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-30574-0_55. Acesso em: 08 ago. 2019. https://doi.org/10.1007/978-3-642-30574-0_55.
- DEV MEDIA. **REST Tutorial**. Rio de Janeiro, 2013. Disponível em: <https://www.devmedia.com.br/rest-tutorial/28912>. Acesso em: 08 ago. 2019.
- POSTAMAN. **Postman 7.1: Create APIs Directly within the Postman App**. 2019. Disponível em: https://blog.getpostman.com/2019/05/17/postman-7-1-create-apis-directly-within-the-postman-app/?_ga=2.237409792.1098372479.1565229112-669073863.1562885275. Acesso em: 08 ago. 2019.
- STROKE ASSOCIATION. **HEALTHY EATING AND STROKE**. London, 2015. Disponível em: https://www.stroke.org.uk/sites/default/files/healthy_eating_and_stroke_0.pdf. Acesso em: 08 ago. 2019.
- HANKEY, Graeme J. **The Role of Nutrition in the Risk and Burden of Stroke**. *Stroke*, v. 48, n. 11, p. 3168-3174. 2017. Disponível em: <https://www.ahajournals.org/doi/10.1161/STROKEAHA.117.016993>. Acesso em: 08 ago. 2019. <https://doi.org/10.1161/STROKEAHA.117.016993>.
- FOROUGH, Mehdi et al. **Stroke and Nutrition: A Review of Studies**. *International Journal of Preventive Medicine*, v. 4, n. 2, p. 165-179. 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3678213/>. Acesso em: 08 ago. 2019.
- GHAZISAEIDI, Marjan et al. **Development of Performance Dashboards in Healthcare Sector: Key Practical Issues**. *Acta Informatica Medica*. v. 23, p. 317-321. 2015. Disponível em:

https://www.researchgate.net/publication/283671434_Development_of_Performance_Dashboards_in_Healthcare_Sector_Key_Practical_Issues. Acesso em: 09 ago. 2019. <https://doi.org/10.5455/aim.2015.23.317-321>.

SEITZ, Vanessa et al. **A case study identifying key performance indicators in public sectors**. IIE Annual Conference.Proceedings, Norcross, p. 371-377. 2014. Disponível em: <https://search.proquest.com/docview/1622299410?accountid=26642>. Acesso em: 09 ago. 2019.

WEGGELAAR-JANSEN, A. M. J. W. M.; BROEKHARST, D. S. E.; BRUIJNE, M. **Developing a Hospital-Wide Quality and Safety Dashboard: A Qualitative Research Study**. BMJ Quality & Safety. v. 27, n. 12, p. 1000-1007. 2018. Disponível em: <https://qualitysafety.bmj.com/content/27/12/1000>. Acesso em: 09 ago. 2019. <https://doi.org/10.1136/bmjqs-2018-007784>.

HERE GEOCODER API. **HERE Geocoder API: Documentation**. 2019. Disponível em: <https://developer.here.com/documentation/geocoder/topics/what-is.html>. Acesso em: 09 ago. 2019.

KHALID, Balar; ABDELWAHAB, Naji. **A Model For Predicting Ischemic Stroke Using Data Mining Algorithms**. International Journal of Innovative Science Engineering and Technology (IJSET). v. 2, n. 11. 2015. Disponível em: https://www.researchgate.net/publication/298352841_A_Model_For_Predicting_Ischemic_Stroke_Using_Data_Mining_Algorithms. Acesso em: 09 ago. 2019.

ANEXO A – Lista das Condições Sensíveis à Atenção Primária para Doenças Cerebrovasculares

Planilha 1 – Lista das condições sensíveis à AB para doenças cerebrovasculares.

CID-10	Descrição
I63	Infarto cerebral
I63.0	Infarto cerebral devido a trombose de artérias pré-cerebrais
I63.1	Infarto cerebral devido a embolia de artérias pré-cerebrais
I63.2	Infarto cerebral devido a oclusão ou estenose não especificadas de artérias pré-cerebrais
I63.3	Infarto cerebral devido a trombose de artérias cerebrais
I63.4	Infarto cerebral devido a embolia de artérias cerebrais
I63.5	Infarto cerebral devido a oclusão ou estenose não especificadas de artérias cerebrais
I63.6	Infarto cerebral devido a trombose venosa cerebral não-piogênica
I63.8	Outros infartos cerebrais
I63.9	Infarto cerebral não especificado
I64	Acidente vascular cerebral, não especificado como hemorrágico ou isquêmico
I65	Oclusão e estenose de artérias pré-cerebrais que não resultam em infarto cerebral
I65.0	Oclusão e estenose da artéria vertebral
I65.1	Oclusão e estenose da artéria basilar
I65.2	Oclusão e estenose da artéria carótida
I65.3	Oclusão e estenose de artérias pré-cerebrais múltiplas e bilaterais
I65.8	Oclusão e estenose de outra artéria pré-cerebral
I65.9	Oclusão e estenose de artérias pré-cerebrais não especificadas
I66	Oclusão e estenose de artérias cerebrais que não resultam em infarto cerebral
I66.0	Oclusão e estenose da artéria cerebral média
I66.1	Oclusão e estenose da artéria cerebral anterior
I66.2	Oclusão e estenose da artéria cerebral posterior
I66.3	Oclusão e estenose de artérias cerebelares
I66.4	Oclusão e estenose de artérias cerebrais, múltiplas e bilaterais
I66.8	Oclusão e estenose de outra artéria cerebral
I66.9	Oclusão e estenose de artéria cerebral não especificada
I67	Outras doenças cerebrovasculares
I67.0	Dissecção de artérias cerebrais, sem ruptura
I67.1	Aneurisma cerebral não-roto
I67.2	Aterosclerose cerebral
I67.3	Leucoencefalopatia vascular progressiva
I67.4	Encefalopatia hipertensiva
I67.5	Doença de Moyamoya
I67.6	Trombose não-piogênica do sistema venoso intracraniano
I67.7	Arterite cerebral não classificada em outra parte
I67.8	Outras doenças cerebrovasculares especificadas

I67.9	Doença cerebrovascular não especificada
I69	Seqüelas de doenças cerebrovasculares
I69.0	Seqüelas de hemorragia subaracnoídea
I69.1	Seqüelas de hemorragia intracerebral
I69.2	Seqüelas de outras hemorragias intracranianas não traumáticas
I69.3	Seqüelas de infarto cerebral
I69.4	Seqüelas de acidente vascular cerebral não especificado como hemorrágico ou isquêmico
I69.8	Seqüelas de outras doenças cerebrovasculares e das não especificadas
G45	Acidentes vasculares cerebrais isquêmicos transitórios e síndromes correlatas
G45.0	Síndrome da artéria vértebro-basilar
G45.1	Síndrome da artéria carotídea (hemisférica)
G45.2	Síndrome das artérias pré-cerebrais, múltiplas e bilaterais
G45.3	Amaurose fugaz
G45.4	Amnésia global transitória
G45.8	Outros acidentes isquêmicos cerebrais transitórios e síndromes correlatas
G45.9	Isquemia cerebral transitória não especificada
G46	Síndromes vasculares cerebrais que ocorrem em doenças cerebrovasculares
G46.0	Síndrome da artéria cerebral média
G46.1	Síndrome da artéria cerebral anterior
G46.2	Síndrome da artéria cerebral posterior
G46.3	Síndromes vasculares do tronco cerebral
G46.4	Síndrome vascular cerebelar
G46.5	Síndrome lacunar motora pura
G46.6	Síndrome lacunar sensorial pura
G46.7	Outras síndromes lacunares
G46.8	Outras síndromes vasculares cerebrais em doenças cerebrovasculares

Fonte: Brasil (2008).