

Desenvolvimento de método para extração de conhecimentos de base de dados da área médica

Leonardo B. de Lima; Paulo Henrique J. Amorim, Thiago Franco de Moraes; Pedro Y. Noritomi (CTI)

leonardo.lima@cti.gov.br; paulo.amorim@cti.gov.br; thiago.franco@cti.gov.br; pnoritom@cti.gov

Resumo

Este trabalho pretende relacionar uma revisão bibliográfica bem como abordar o conceito de sistemas de informações hospitalares e as suas mais diversas ferramentas. Tais ferramentas podem ser entendidas desde sistemas de gestão, métodos de controle, sistemas computacionais entre outros. A partir da experiência do ProMED, um grupo de pesquisa funcionando no CTI Renato Archer, e do mecanismo de coleta de informações ProMED Web, pretende-se relacionar as maneiras de como tornar o sistema de informação mais eficiente aos usuários, reduzir tempos de processamentos e tratar a informação a luz das novas leis de proteção de dados.

Palavras-chave: Sistema de informação hospitalares, Gestão dos sistemas hospitalares, Segurança de dados pessoais, métodos de predição de dados com linguagem Python.

1. Introdução

O ProMED é um programa que utiliza tecnologias 3D, criado pelo Dr. Jorge Silva no ano 2000, para desenvolver soluções de engenharia para cirurgias complexas, auxiliando o planejamento cirúrgico e desenvolvimento de próteses. Atualmente, sua missão é contribuir para melhorar a qualidade de vida dos pacientes, com concomitante redução de custos para o SUS e para a previdência social. Essa missão vem sendo cumprida ao transformar pesquisa e inovação em resultados concretos e quantificáveis para a sociedade. O MS (Ministério da Saúde) vem apoiando esse programa desde 2009, sendo está a primeira parceria firmada entre o MS e uma unidade de pesquisa do MCTI (Ministério da ciência, tecnologia e Inovações).

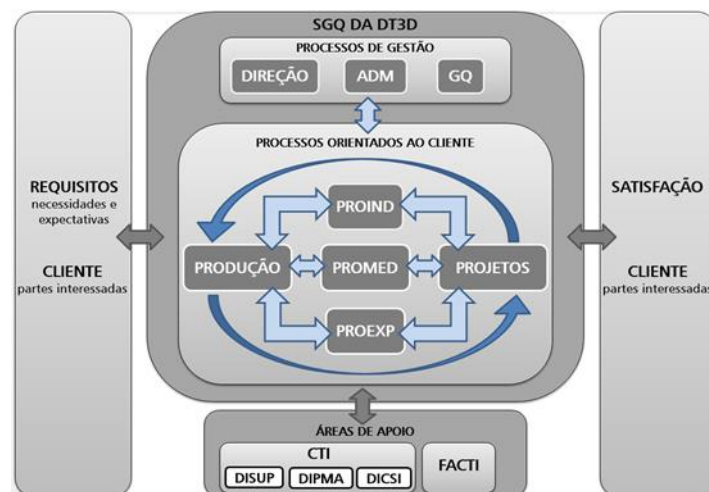


Figura 1 - Fluxograma de interação dos processos, fonte: MQ-NT3D

As cirurgias apoiadas por impressão 3D são uma forma revolucionária de tratar pacientes, pois são mais rápidas e seguras, propiciando melhoria de vida aos pacientes com concomitante redução de custos para o SUS (Sistema Único de Saúde).

Como um dos processos do DITPS, o ProMED explora a mesma tecnologia em moldes para próteses de crânio e face, e há projetos de investir em novas áreas.

Neste sentido, o ProMED atua como um órgão da administração pública sob a hierarquia do DITPS (Divisão de Tecnologia para Produção e Saúde), CTI e MCTI (Ministério da ciência, tecnologia e Inovações) e apoiando tecnologicamente o MS (Ministério da Saúde). Em especial, o sistema único de saúde (SUS), um importante projeto que contribui para a garantia de direitos da população brasileira.

PROMED		
MATERIAL/EQUIPAMENTOS		COMPETÊNCIA/TREINAMENTOS
<ul style="list-style-type: none"> • Infraestrutura: Meios de comunicação, Ambiente de trabalho, DIFIN e DICSÍ • Informática: Computadores, sistemas computacionais, Impressoras. • Software: SolidWorks, Magics, InVesalius, WEB2PROJECT 		<ul style="list-style-type: none"> • Coordenador do NT3D • Assistente Administrativo • Técnicos de modelagem • Bolsista (s) Técnico – PROMED
ENTRADA	ATIVIDADES	SAÍDA
<ul style="list-style-type: none"> • Política e Objetivos da Qualidade • Dados Cadastrais de Clientes (requisitos de clientes) • Imagens (Propriedade de Cliente) • Propostas e Pedidos (Solicitação de serviços) • Relatório Cirúrgico • Ações Propostas (Ação Corretivas e Melhorias) 	<ul style="list-style-type: none"> • Gerar e arquivar imagens • Analisar criticamente a proposta e pedido • Encaminhar para Cliente a Proposta (Solicitação de serviço/termo de compromisso/Modelo Relatório Cirúrgico) • Gerar protocolo e número do caso • Realizar a modelagem dos arquivos • Encaminhar solicitação de protótipo à produção • Gerar informação quanto à execução do Indicador • Análise crítica da Conformidade do Produto. 	<ul style="list-style-type: none"> • Banco de dados do cliente (propriedade do cliente) • Registro de análise crítica de proposta e pedido • Proposta/Pedido • Solicitação de Faturamento • Comunicação com o Cliente • Protótipo • Dados Estatísticos.
DOCUMENTAÇÃO		MONITORAMENTO E MEDIÇÃO
<ul style="list-style-type: none"> • MQ-NT3D – Manual Setorial da Qualidade do NT3D • IT – 001/09 – Instrução de trabalho para prototipagem rápida. 		<ul style="list-style-type: none"> • Prazo de Entrega • Prazo da Proposta • Nº Casos • Nº Hospitais

Figura 2 - Relação dos processos inerentes ao ProMED, fonte: MQ-NT3D

Desta maneira, o ProMED fornece o suporte de auxílio aos hospitais brasileiros de natureza pública sem qualquer custo direto. O processo inicia-se com o manifesto de interesse por parte de um representante do hospital público e em seguida são iniciados uma espécie de entrevista com o cirurgião. Algumas destas informações são necessárias para o bom caminhar do processo. Por exemplo, as características da cirurgia, características da patologia, bem como

suas limitações técnicas, ademais, são solicitados os arquivos gerados por tomografia computadorizada e/ou da ressonância magnética.

Estes arquivos são direcionados, atualmente, por correio eletrônico e, em casos de exceder os limites de banda, eles podem ser enviados fisicamente pelos serviços de correios.

Apesar do programa, desde sua criação em 2000 ser um sucesso e atender plenamente sua missão de auxiliar os mais diversos hospitais com casos complexos, reduzir eventuais despesas adicionais e sobretudo como uma última esperança para muitos casos., o ProMED possui muito potencial para expandir e melhorar continuamente para continuar sua importante tarefa no auxílio dos hospitais.

2. Embasamento teórico

O ProMED como sendo um programa de amplo interesse de universidades, grupos de pesquisa e até mesmo museus atua predominantemente no auxílio aos hospitais, portanto a necessidade de basear-se em boas práticas que sejam capazes de conduzir os esforços nos termos da melhoria continua.

Desta maneira, o presente trabalho pretende evidenciar a definição de saúde como um completo bem-estar, conforme consta nos estatutos da Organização Mundial de Saúde (OMS), aprovados em conferência internacional realizada na cidade de Nova Iorque, em julho de 1946. A definição de saúde incorporada a esses estatutos é bem conhecida: “saúde é o estado de completo bem-estar físico, mental e social e não meramente a ausência de doença ou enfermidade” (WHO, 1946).

Para Roberto P. Nogueira, essa definição decorreu da influência da doutrina utilitarista de Jeremy Bentham sobre alguns médicos europeus propugnadores da medicina social. Contudo, considerando que a ideia de completo bem-estar deve ser interpretada não apenas como convenção e inspiração filosófica, mas também considerando os objetivos estratégicos da OMS, marcada pela conjuntura especial do pós-guerra, naquela ocasião, e pelo desafio da reorganização econômico-social dos países. (NOGUEIRA, ROBERTO P., 2020).

Tal definição nos permite vislumbrar e nos inspirar a luz do que representa o termo saúde e consequentemente orientar os esforços e ações para garantir um serviço de qualidade e que, portanto, respeitem as condições da dignidade humana bem como suas leis regulamentadoras que definem desde os princípios instituídos pela Constituição Federal, em termos de direitos da saúde e bem-estar social até os interesses públicos de contratação de serviços e princípios de eficiência e responsabilidade dos gastos públicos.

Esta linha de estudos médicos acerca do valor econômico da saúde alcançaria repercussão internacional em meados de 1970 através da teoria da transição epidemiológica, concebida pelo médico de origem egípcia Abdel Omram. (NOGUEIRA, ROBERTO P., 2020).

Dois médicos, no Brasil, costumam ser lembrados pela discussão da relação entre saúde e desenvolvimento: Josué de Castro e Mário Magalhães da Silveira. Ambos conduziram seus estudos a partir dos dados e métodos peculiares das desigualdades sociais sobre as condições de saúde e bem-estar da população. Empenharam-se em levar a conhecimento das autoridades nacionais que más condições de saúde decorriam da desnutrição, da mortalidade infantil e da reduzida expectativa geral de vida.

Vários indicadores demográficos e epidemiológicos eram usados por esses e outros médicos brasileiros dessa época com o propósito de evidenciar o valor econômico da saúde. (NOGUEIRA, ROBERTO P., 2020).

É claro, considerando apenas os casos de patologias associadas à má distribuição de saneamento básico dos municípios representados por unidades da federação. Os casos atendidos pelos hospitais em todo o Brasil atendem as mais diversas casos e patologias desde doenças crônicas como diabetes, Alzheimer, Asma, AIDS, Câncer, Depressão, AVC (Acidente Vascular Cerebral), Parkinson.

Em reportagem da jornalista Leticia Alves, de 03 de fevereiro de 2020, publicado no site Poder360, pode-se comparar as principais causas de morte no Brasil, segundo os dados do DataSus.

As 10 Principais causa de morte no Brasil		
#	Causas	Quantidade
1	Doenças do aparelho circulatório	356.178
2	Tumores	227.150
3	Doenças do aparelho respiratorio	155.921
4	Causas externas de morbidade e mortalidade	150.165
5	Doenças endócrinas nutricionais e metabólicas	80.293
6	Sintomas, achados e anormais de exames clinicos	73.793
7	Doenças do aparelho digestivo	67.029
8	Doenças infecciosas e parasitárias	54.814
9	Doenças do aparelho geniturinário	43.290
10	Doenças do sistema nervoso	40.520

fonte: DataSus - departamento de informática do SUS (Sistema Único de Saúde)

Figura 3 – Principais causas de morte no Brasil

No Brasil, de acordo com levantamento feito pelo DataSus, departamento de informática do SUS (Sistema Único de Saúde), as doenças do aparelho circulatório aparecem em 1º lugar, com 356.178 mortes. Destas, as mais comuns são as arteriais, cerebrovasculares e hipertensivas, além do infarto agudo do miocárdio. (ALVES, LETICIA, 2020).

Na 4ª posição, está o que foi denominado pelo SUS como “causas externas de morbidade e mortalidade” (150.165), que incluem mortes motivadas por circunstâncias distintas, como lesões causadas por tragédias ambientais, afogamentos e envenenamentos, por exemplo. O texto sobre a metodologia do estudo não especifica os casos.

Independentemente da origem, todos estes casos acabam indo ocupar as estatísticas das bases de dados do MS. E, portanto, são indicadores que estabelecem o Brasil com um potencial gigantesco para implementação de boas práticas não apenas no âmbito da saúde, mas, sobretudo, no âmbito da justiça social.

Tais números justificam a importância de uma gestão eficiente dos sistemas de informação hospitalares e, também, e pela função social que o ProMED desempenha para todo o Brasil. Pois, uma gestão eficiente está totalmente alinhada com o combate à desigualdade social de

nosso país, sobretudo com a contribuição para melhorar índices que demonstrem um país saudável.

3. Metodologia

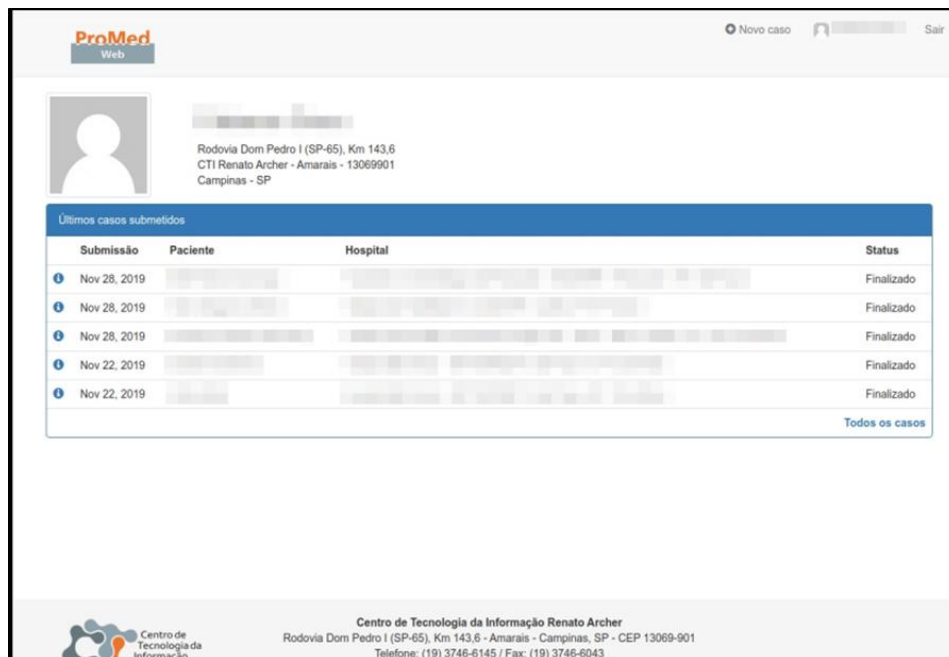
3.1. Plano de trabalho

Este trabalho está sendo conduzido de maneira a respeitar o plano de trabalho do bolsista PCI (Programa de Capacitação Institucional) registrado em CNPq 444303/2018-9. Em sua modalidade PCI-DD com período de 60 meses, na divisão DITPS (Divisão de Tecnologia para Produção e Saúde), antigo NT3D (Núcleo de Tecnologias Tridimensionais). E que tem por título:

“Desenvolvimento e implementação de método para extração de conhecimentos de base de dados da área médica”

Como um dos objetivos deste programa está a capacitação do bolsista frente aos temas de mineração, correlação e visualização de dados médicos textuais. Por outro lado, tornar o sistema ProMED Web mais robusto com a possibilidade de realizar mineração de dados, pesquisas, inserção e visualização de imagens médicas, além da integração com o sistema de armazenamento de imagens médicas (PACS) e com o software InVesalius.

Abaixo é possível visualizar a tela de cadastro de atendimentos do sistema ProMED Web.



The screenshot shows the ProMED Web interface. At the top left, there is a logo for 'ProMed Web'. On the right, there are links for 'Novo caso' and 'Sair'. Below the logo, there is a user profile section with a placeholder for a profile picture and the following address: 'Rodovia Dom Pedro I (SP-65), Km 143,6', 'CTI Renato Archer - Amaraís - 13069901', 'Campinas - SP'. Below this is a table titled 'Últimos casos submetidos' with columns for 'Submissão', 'Paciente', 'Hospital', and 'Status'. The table contains five rows of data, all with a status of 'Finalizado'. At the bottom right of the table, there is a link for 'Todos os casos'. At the bottom of the page, there is a footer with the logo of the 'Centro de Tecnologia da Informação' and contact information for 'Centro de Tecnologia da Informação Renato Archer'.

Submissão	Paciente	Hospital	Status
Nov 28, 2019			Finalizado
Nov 28, 2019			Finalizado
Nov 28, 2019			Finalizado
Nov 22, 2019			Finalizado
Nov 22, 2019			Finalizado

Centro de Tecnologia da Informação Renato Archer
Rodovia Dom Pedro I (SP-65), Km 143,6 - Amaraís - Campinas, SP - CEP 13069-901
Telefone: (19) 3746-6145 / Fax: (19) 3746-6043

Figura 4 – Tela de cadastros de casos ProMED Web

Atualmente o DITPS, armazena uma base de dados de casos cirúrgicos auxiliados por planejamento virtual (Processamento de imagens e Modelagem 3D) e físico (Impressão 3D). É do interesse da divisão realizar melhorias nessa base de dados e em seu respectivo sistema ProMED Web, inserir, consultar novos itens como informações relacionadas a ISO 9001 –

Sistema de Gestão da Qualidade e correlacionar esses dados de forma simples e automatizada. Além dos textuais, estes sistemas serão capazes de se integrar

com os sistemas de armazenamento de imagens médicas PACS e ao software livre de processamento de imagens médicas InVesalius.

O trabalho prevê a revisão bibliográfica dos padrões e protocolos de troca de informação utilizados na área médica, implementação de ferramenta para cadastro de casos cirúrgicos, implementação de ferramenta para consulta de casos médicos cadastrados, configuração de um servidor de imagens médicas PACS, bem como o cadastro de imagens no servidor PACS, a integração do sistema ProMED Web com o servidor PACS, a integração do software InVesalius ao ProMED Web, por fim a revisão bibliográfica de técnicas de mineração de dados e machine learning para extração e correlação de dados médicos.

3.2 Revisão bibliográfica

3.2.1 Sistemas de Informação Hospitalares

Os sistemas de informação em saúde, podem ser considerados como os recursos necessários para guardar e processar a informação a uma organização de saúde, informação que deve ser útil e estar acessível àqueles que dela necessitam. Um sistema de informação é, pois, uma combinação de procedimentos, informação, pessoas, tecnologias e vários outros recursos. Note-se que um sistema de informação pode envolver, ou não, a utilização de tecnologia informática, por isso não se deve confundir sistema de informação com um sistema informático. (ALMEIDA, ANDREIA DA SILVA, 2012).

Os sistemas de informação em saúde devem incluir todos os dados necessários aos profissionais de saúde e utilizadores dos sistemas, como o objetivo de desenvolverem e protegerem a saúde das populações. Segundo a organização Mundial de Saúde, o investimento nestes sistemas de informação possui vários benefícios, como ao nível do auxílio aos tomadores de decisão, no controle e detecção de problemas de saúde endêmicos, na monitorização de progressos e metas preestabelecidos e na promoção da equidade e da qualidade dos serviços.

3.2.2 Software InVesalius

É um software livre, robusto, simples (com funcionalidades restritas e suficientes ao seu propósito), com interface amigável, orientada por objetos, multiplataforma (executável em Windows e Linux) e apropriado para o uso em PC. (AMORIM P., MORAES T., SILVA J., PEDRINI H., 2015).

Como um dos principais projetos do grupo, o software InVesalius ajuda no planejamento do trabalho de cirurgiões. A partir de imagens médicas tomográficas, permite a criação de modelos virtuais 3D de estruturas anatômicas e torna possível a confecção de modelos físicos dessas estruturas com o auxílio da prototipagem rápida.

Os dois sistemas mais conhecidos e usados na aquisição de imagens médicas são a tomografia computadorizada TC e a ressonância magnética RM. O modelo digitalizado é um volume virtual, representando um volume real do corpo do paciente. O volume é mostrado no

monitor por meio da reformatação dos dados, para criar projeções ortográficas ou uma pseudorepresentação 3D usando algoritmos de rendering.

3.2.3 Modelos Estatísticos

Modelo estatístico é um conjunto de hipóteses sobre a geração de dados observados e dados semelhantes de diferentes fatores. Este descreve a relação entre variáveis aleatórias com outras variáveis (não aleatórias). Em termos matemáticos, um modelo estatístico é geralmente considerado como um par (S, P) , em que S é o conjunto de observações possíveis, isto é, o espaço amostral, e P é um conjunto de distribuição de probabilidades em S . (MCCULLAGH, P., 2002).

Algumas variáveis aleatórias adaptam-se muito bem a uma série de problemas práticos. Portanto, um estudo pormenorizado dessas variáveis é de grande importância para a construção de modelos probabilísticos para situações reais e a consequente estimação de seus parâmetros. Para algumas dessas distribuições existem tabelas que facilitam o cálculo de probabilidades, em função de seus parâmetros. Nesta seção iremos estudar alguns desses modelos, procurando enfatizar as condições em que eles aparecem, suas funções de probabilidade, parâmetros e como calcular probabilidades. (MORETTIN, PEDRO ALBERTO, 2010)

Um modelo estatístico pode ser usado para descrever o conjunto de distribuição a partir da qual se assume que um determinado conjunto de dados é amostrado. Por exemplo, se assumir que os dados surgem a partir de uma distribuição normal, em seguida, esta assume um modelo gaussiano. (ADÈR, H. J., 2008).

3.2.4 Lei 13.709/2018 - Lei Geral de Proteção de Dados Pessoais (LGPD)

A Lei Geral de Proteção de Dados Pessoais entrou em vigor em 18 de setembro de 2020 e foi aprovada em 2018. A LGPD coloca o Brasil ao lado de mais de 100 países onde há normas específicas para definir limites e condições para coleta, guarda e tratamento de informações pessoais.

A lei no. 13.709 disciplina um conjunto de aspectos: define categorias de dados, circunscreve para quem valem seus ditames, fixa as hipóteses de coleta e tratamento de dados, traz os direitos dos titulares de dados, detalha condições especiais para dados sensíveis e segmentos (como crianças), estabelece obrigações às empresas, institui um regime diferenciado para o Poder Público, coloca sanções em caso de violações e prevê a criação de uma autoridade nacional.

Segundo a norma, dados pessoais são informações que podem identificar alguém. Dentro do conceito, foi criada a “categoria dado sensível”, com informações sobre origem racial ou étnica, convicções religiosas, opiniões políticas, saúde ou vida sexual. Registros como esses

passam a ter nível maior de proteção, para evitar formas de discriminação. (VALENTE, J., 2020).

3.2.5 Mineração de Dados

A mineração de dados é também conhecida pelo termo inglês Data Mining e trata-se do processo de explorar dados à procura de padrões consistentes, como regras de associação ou sequencias temporais, para detectar relacionamentos sistemáticos entre variáveis, desta maneira descobrindo novos padrões nos relacionamentos dos dados, bem como descobrir regras, identificar fatores, descobrir padrões e relacionamentos ocultos em grandes bancos de dados para o auxílio a tomada de decisões. (O' BRIEN, JAMES A., 2005).

Um conjunto de ferramentas e técnicas compõem a mineração de dados, por exemplo, através de algoritmos de aprendizagem ou classificação baseados em redes neurais e estatísticas, são capazes de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento. Esse conhecimento pode ser apresentado por essas ferramentas de diversas formas: agrupamentos, hipóteses, regras, arvores de decisão, grafos ou dendrogramas.

A mineração de dados também tem um caráter multidisciplinar quando é inevitável abordar disciplinas como a estatística, matemática e a computação. Pode ser aplicada tanto para a pesquisa científica, tanto para impulsionar a lucratividade de empresas mantendo altos níveis de inovação e competitividade.

4. Desenvolvimento

O desenvolvimento deste trabalho está condicionado a análise preditiva e mineração dos dados com base nas informações contidas na base de dados cirúrgicos do ProMED. Para tanto, faz-se necessário a identificação dos dados necessários e cabe o questionamento de quais serão os dados que devem estar contidos na análise e identificá-los como tal. Em caso, de informações sensíveis e que estão sob a proteção de leis regulamentadores de proteção de dados, atuar conforme as diretrizes e orientações estabelecendo as tratativas de informações como a anonimização de dados posteriormente.

Cabe a avaliação do cruzamento de dados com outras bases de dados abertos como por exemplo o Data Sus e informações do Instituto de Pesquisas Econômicas Aplicadas (IPEA) ou mesmo do Ministério da Saúde. Esta maneira, será possível propor projetos de melhorias para os sistemas de saúde do Brasil através do apontamento e considerações gerados nos modelos de predição de dados. Garantindo a orientação de novos investimentos e sobretudo contribui para o alívio e atendimento de questões carentes e atendendo aos interesses econômico-financeiros do MS.

Para tanto, pretende-se utilizar as bases dos modelos estatísticos como as distribuições: de Bernoulli, Binomial, Hipergeométrica, de Poisson para casos possíveis. Para as análises massivas de dados as bibliotecas Pandas e Numpy da linguagem Python apresenta grande viabilidade por ser um software gratuito e alta capacidade na manipulação de dados. E, por fim,

interpretar e inspirar o trabalho, bem como os seus relatórios de acordo com as diretrizes de proteção de dados pessoais.

Referências

MCCULLAGH, P., “What is a statistical model?”, *Annals of Statistics*, 30:1225-1310 (2002).

ADÈR, H. J. “Chapter 12: Modelling”. In H.J. Adèr & G. J. Mellenbergh (Eds.) (with contributions by D. J. Hand), *Advinsing on Research Methods: A consultant’s companion* (pp. 271-304). Huizen, The Netherlands: Johannes van Kessel Publishing, 2008

MORETTIN, PEDRO ALBERTO. *Estatística Básica/Pedro A. Morettin, Wilton O. Bussab.* – 6. ed. – São Paulo: Saraiva, 2010.

BENTHAM, J. *An introduction to the principles of morals and legislation.* Kitchener: Batoche Books, 2000.

MQ-NT3D_manual_setorial_da_qualdiade_Rev_28/02/2018

O’ BRIEN, JAMES A. *Sistemas de Informação e as decisões gerenciais na era da Internet 2a. Edição.* São Paulo: Saraiva. p. 143. ISBN 9788502044074, 2005

WHO – WORLD HEALTH ORGANIZATION. *Minutes of the technical preparatory committee for the international health conference held in Paris from 18 March to 5 April 1946.* Paris: United Nations, World Health Organization Interim Commission, 1946a.

NOGUEIRA, ROBERTO P. *Texto para discussão 2598 /os médicos, a saúde como completo bem-estar e a questão do Desenvolvimento / Instituto de Pesquisa Econômica Aplicada.* - Brasília: Rio de Janeiro: Ipea, 2020

AMORIM P., MORAES T., SILVA J., PEDRINI H. *InVesalius: An Interactive Rendering Framework for Health Care Support.* In: *Bebis G. et al. (eds) Advances in Visual Computing.* ISVC 2015.

SANTA BÁRBARA, AILTON. *Processamento de imagens médicas tomográficas para modelagem virtual e física – o software InVesalius / Ailton Santa Bárbara.* --Campinas, SP: [s.n.], 2006.

ALMEIDA, ANDREIA DA SILVA. *Os Sistemas de Gestão da Informação nos Hospitais Públicos Portugueses: Uma Perspectiva Actual.* *Dissertação de Mestrado em Ciências da Documentação e da Informação, apresentada à Faculdade de Letras da Universidade de Lisboa, Lisboa, janeiro de 2012.*

ALVES, LETICIA. *Saiba as principais causas de mortes no Brasil e nos EUA.* Poder360, 2020. Disponível em: <<https://www.poder360.com.br/internacional/saiba-as-principais-causas-de-mortes-no-brasil-e-nos-eua/>>. Acesso em: 18 de outubro de 2020.

VALENTE, JONAS. *Entenda o que muda com a Lei Geral de Proteção de Dados, Agência Brasil, publicação feita em 18 de setembro de 2020 no site <https://agenciabrasil.ebc.com.br>.*