Dado o momento de pandemia de Covid-19 e reconhecendo que diversas formas de conhecimento podem auxiliar no entendimento dos impactos socioeconômicos e ambientais no Estado do Amazonas, a Secretaria de Estado de Desenvolvimento Econômico, Ciência, Tecnologia e Inovação (SEDECTI), por meio da Secretaria Executiva de Ciência, Tecnologia e Inovação (SECTI), lançou uma chamada para pesquisadores de todas as áreas do conhecimento com a seguinte pergunta: Na perspectiva de sua área de estudo e com a literatura existente, de que forma poderia contribuir com a sociedade e Governo para se organizarem para o agora e para o futuro? As notas técnicas que apresentamos são frutos desta iniciativa.

Saiba mais em www.sedecti.am.gov.br.

Esforços da Ciência de Dados Contra a COVID-19

Recursos, Chamados à Ação e Desafios

NOTA TÉCNICA
COVID-19
Nº 002
28/04/2020

Sociedade Brasileira para o Progresso da Ciência

Secretaria de **Desenvolvimento Econômico, Ciência, Tecnologia e Inovação**

AMAZONAS GOVERNO DO ESTADO

Elloá B. Guedes *ebgcosta@uea.edu.br*
Carlos Maurício S. Figueiredo *cfigueiredo@uea.edu.br*
Tiago E. de Melo *tmelo@uea.edu.br*

Grupo de Pesquisa em Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas

Nota elaborada e revisada até 09/04/2020.

O conteúdo das notas técnicas é de responsabilidade dos autores. A SEDECTI, em parceria com a Sociedade Brasileira para o Progresso da Ciência (SBPC), apenas selecionou, revisou e publicou o conteúdo aqui apresentado.

Esforços da Ciência de Dados Contra a COVID-19

Recursos, Chamados à Ação e Desafios

1. Contextualização

Entende-se por *Big Data* um fenômeno emergente do grande volume de dados produzidos em vários formatos e armazenados por uma grande quantidade de dispositivos e equipamentos (AMARAL, 2016). Estima-se que, a cada segundo, sejam gerados mais de 30:000 gigabytes de dados, e que esta taxa continue acelerando a cada dia (MARZ; WARREN, 2015). A onipresença de dados na contemporaneidade implica em mudanças, tais como, econômicas, sociais e estratégicas e, em decorrência, tem-se que Big Data é considerada a nova Revolução Industrial, em sua quarta fase (PROVOST; FAWCETT, 2016).

Nesse contexto, a Ciência de Dados (do inglês, *Data Science*) é uma área interdisciplinar que congrega métodos científicos, processos e sistemas capazes de extrair informação, produzindo *insights* a partir de dados (AMARAL, 2016). Suas ferramentas permitem a construção de visualizações, as quais favorecem o entendimento de determinados contextos, a extração de características relevantes e a produção de modelos computacionais capazes de inferir padrões existentes nos dados, a chamada Aprendizagem de Máquina (CADY, 2017). Este corpo de conhecimento colabora, por exemplo, para uma análise de cenários contemporâneos e também para a previsão de eventos futuros ou agrupamento de similaridades, o que pode gerar um diferencial inovador e estratégico (PROVOST; FAWCETT, 2016).

No âmbito da Saúde, tem se tornado cada vez mais evidente que repensar os modelos de cuidado com estratégias da Ciência dos Dados pode resultar em uma contribuição positiva em diversos desafios enfrentados globalmente, visando alcançar melhorias na saúde da população, na redução de custos dos tratamentos, na satisfação da experiência do paciente, e no ambiente de trabalho dos profissionais dessas áreas (CONSOLI; RECUPERO; PETKOVIC, 2019). De maneira geral, o corpo de conhecimento da Ciência de Dados irá oferecer novas oportunidades por meio da abrangência de quatro perspectivas: (i) descritiva, relacionada a responder a respeito de algo que aconteceu; (ii) diagnóstica, com vistas a esclarecer as causas do que aconteceu; (iii) preditiva, com vistas a determinar o que irá acontecer; e (iv) prescritiva, para detectar como é possível fazer algo acontecer (ABEDJAN et al., 2019).

Na Medicina, a Aprendizagem de Máquina e o Aprendizado Profundo (do inglês, Deep Learning) têm figurado como componentes centrais, com diversas contribuições e muitas perspectivas interessantes (LEE et al., 2017). Soluções baseadas em Visão Computacional para a análise de imagens médicas, em especial, têm colaborado na detecção de lesões ou doenças, classificação e diagnóstico, especialmente nos âmbitos da imagiologia cardiotorácica e mamária (GREENSPAN; GINNEKEN; SUMMERS, 2016; LEE et al., 2017) Neste contexto, contribuições já desenvolvidas por este grupo de pesquisa abordam a análise de lesões pigmentadas de pele (LIMA et al., 2019) e a classificação de tuberculose pulmonar a partir de imagens de raio-X torácica (EVANGELISTA; GUEDES, 2020), com resultados compatíveis com o estado da arte.

Em face da pandemia de COVID-19 declarada pela Organização Mundial de Saúde (OMS) em 11 de Março de 2020, decorrente da doença causada por um novo agente do coronavírus descoberto em Dezembro de 2019 após casos registrados na China (OMS, 2020c), esforços de âmbito global têm sido evocados para minimizar os impactos deste mal que, em 8 de Abril 2020, já acometeu mais de 202 localidades e de 1:447:466 indivíduos, causando óbito a 83:471 pessoas (OMS, 2020a). De acordo com o Portal LiveScience, em comparação com a última pandemia, do vírus H1N1 nos anos de 2009 a 2010, houve um registro de 1; 4 bilhões de casos (HICKOK, 2020). Percebe-se, assim, com muito mais gravidade, que estes valores já foram superados pelo COVID-19 em cerca de 4 meses desde o seu surgimento. Nesta força-tarefa, oportunidades da Ciência de Dados para colaborar na

NOTA TÉCNICA
COVID-19
Nº 002
28/04/2020

SBPC Sociedade Brasileira para o Progresso da Ciência

Secretaria de Desenvolvimento Econômico, Ciência, Tecnologia e Inovação

AMAZONAS GOVERNO DO ESTADO

mitigação desta pandemia também têm emergido. Esta nota técnica visa, assim, sintetizar algumas destas oportunidades, motivando estudantes, especialistas, técnicos, docentes e pesquisadores desta área a contribuir com os desafios que se apresentam. Além disso, cabe a este documento também ressaltar os desafios e limitações destas soluções, alertando para que esta área ainda emergente não incorra em prejuízos à vida humana e ao meio ambiente.

2. Dados, Recursos e Competições

Esta seção apresenta bases de dados atualmente disponíveis sobre a pandemia do COVID-19, introduz o conceito e ilustra *dashboards* para a visualização dos efeitos da pandemia, discorre sobre as contribuições de *data analytics* para este cenário e apresenta oportunidades de competições e chamados à ação atualmente disponíveis para a comunidade.

2.1 Bases de Dados

As bases de dados (também chamadas de *datasets*) sobre o COVID-19 tipicamente consistem de informações epidemiológicas atualizadas constantemente a respeito dos índices de casos confirmados, suspeitos, óbitos e recuperação dos pacientes, eventualmente complementadas com dados sobre local, data de ocorrência ou registro e meta-dados dos pacientes, tais como idade, estado, doenças pré-existentes e afins.

Os relatórios de situação divulgados diariamente pela OMS costumam ser a fonte primária de informação acerca dos impactos do COVID-19 em escala global (OMS, 2020b), pois contemplam dados de vigilância epidemiológica da maioria dos países, número de casos, número de novos casos, total de óbitos e classificação do tipo de transmissão incidente no local, podendo esta ser transmissão comunitária, local, apenas casos importados, situação sob investigação ou transmissão interrompida.

Com vistas a diminuir a distância entre os dados disponíveis e o desenvolvimento de aplicações em Ciência de Dados para o COVID-19, a Universidade Johns Hopkins juntamente com o Wordometer, desenvolveram uma biblioteca na linguagem de programação Python, denominada `covid`, capaz de prover os dados disponibilizados pela OMS de maneira mais intuitiva para aqueles que desejem executar algoritmos sobre esses dados (PYTHON, 2020; JHU, 2020; WORLDOMETER, 2020).

Para o desenvolvimento de métodos efificientes, os pesquisadores precisam de um grande volume de dados e, neste sentido, a plataforma Kaggle tem disponibilizado bases de dados bastante detalhadas e atualizadas diariamente, as quais incluem diversas informações, tais como idade dos pacientes, localização, início do surgimento dos sintomas, o momento em que os pacientes foram expostos etc. (KAGGLE, 2020). Aproximadamente, 300 grupos estão utilizando esses dados nas suas pesquisas atualmente.

As mídias sociais, como o Twitter e Instagram, são outras fontes importantes de dados para que os pesquisadores possam entender a propagação da pandemia. Nesse ambiente, a IEEE está disponibilizando um grande volume de tuítes que podem dar uma boa representação do alcance da doença e como as pessoas estão lidando com isso (IEEE, 2020a). O *dataset* disponibilizado possui mais de 13 milhões de tuítes e já está no formato csv, adequado para ser processado pelos cientistas de dados.

Há também no Brasil diversas iniciativas de compartilhamento de dados sobre a pandemia do COVID-19. É o caso, por exemplo, do projeto Brasil.IO que foi criado com o objetivo de ser referência para quem procura ou quer procurar dados abertos sobre o Brasil de maneira estruturada e pronta para ser usada por pesquisadores. Em relação ao COVID-19, o projeto disponibiliza boletins informativos e casos do coronavírus por municípios por dia (BRASIL.IO, 2020). As secretarias de saúde estaduais são as fontes de dados.

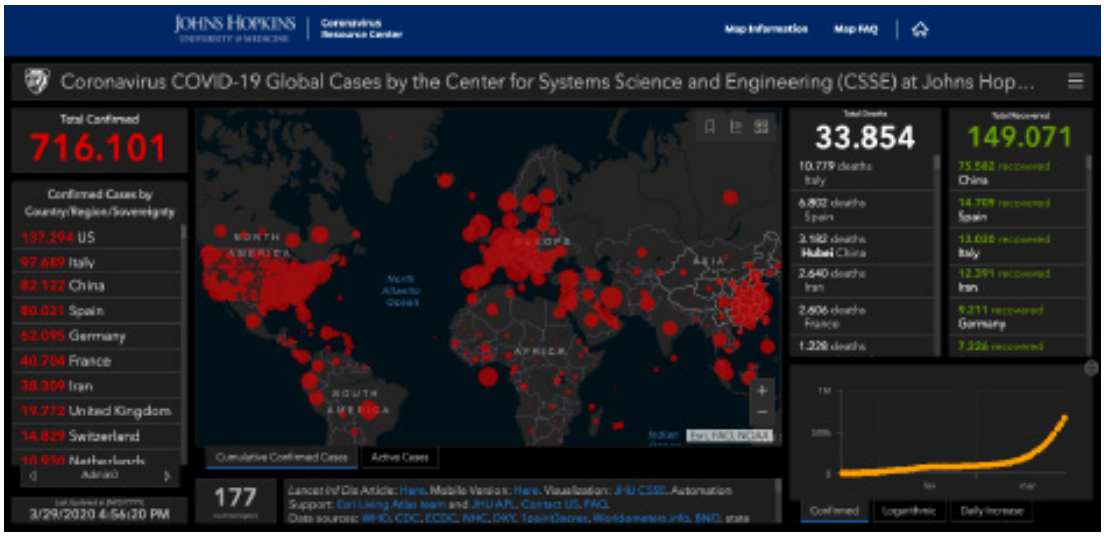
2.2 *Dashboards* e *Data Analytics*

Visualização de dados é uma área de trabalho da Ciência de Dados que foca na comunicação de informações para auxílio à tomada de decisão ou, simplesmente, para fins de monitoramento do fenômeno estudado. Normalmente, a comunicação de dados acontece pelo uso de ferramentas visuais, e um instrumento muito comum é o de representação gráfica de dados (AMARAL, 2016).

Dashboards consistem de visualizações que reúnem vários dados e informações relacionadas a um fenômeno em um único local, como por exemplo, uma página Web. Seu principal objetivo é facilitar o monitoramento do fenômeno e a obtenção de respostas rápidas para perguntas sobre o fenômeno, tais como: Como este fenômeno evolui? Existe algum padrão? Qual a sua tendência? É possível observar alguma métrica como sendo crítica?

Em relação ao coronavírus, desde os primeiros casos, tem-se o exemplo do *dashboard* COVID-19 *Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. Nele é possível observar como o fenômeno se espalha pelo mundo, com quantidades de casos detectados e mortes registradas em decorrência da doença. Gráficos em forma de mapa e curvas mostram as tendências de disseminação geográfica e crescimento de casos por país. Este *dashboard* é dinâmico e pode ser acessado no link: <https://coronavirus.jhu.edu/map.html>. Uma captura de tela do mesmo é mostrada na Figura 1.

Figura 1: *Dashboard* para o COVID-19.



Fonte: <https://coronavirus.jhu.edu/map.html>

De forma similar, vê-se que canais de notícias costumam apresentar seus próprios *dashboards*, facilitando a divulgação de informações junto à população. Dentre centenas de exemplos existentes pelo mundo, citam-se: Monitor Estadão[1], Infográfico O Globo[2], Mapa Coronavírus G1[3], Microsoft Covid Dashboard[4], e The New York Times Dashboard[5]. Desenvolvedores e entusiastas também tem colaborado individualmente na produção de visualizações deste tipo, tais como ilustrado em: <https://www.kaggle.com/elloaguedes/panorama-do-covid-19-no-brasil/>.

Tarefas de *analytics* se dividem em duas categorias principais: análise descritiva e análise preditiva. A análise descritiva consiste em entender os dados do fenômeno em questão, e a análise preditiva, por sua vez, consiste em projetar para o futuro as métricas do fenômeno baseando-se em padrões identificados nos dados existentes.

[1] <https://arte.estadao.com.br/ciencia/novo-coronavirus/monitor-pandemia/>
[2] <https://infograficos.oglobo.globo.com/sociedade/numeros-do-coronavirus-no-brasil-e-no-mundo.html>
[3] <http://especiais.g1.globo.com/bemestar/coronavirus/mapa-coronavirus/#/>
[4] <https://www.bing.com/covid>
[5] <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html>

Atualmente, os dados da pandemia do COVID-19 consistem majoritariamente do número de casos detectados e número de mortes. Nesse caso, as tarefas de análise preditiva consistem principalmente na identificação da curva de crescimento de casos, a qual vem sendo modelada como uma curva de crescimento exponencial. Uma vez modelada essa curva, tem-se, de forma natural, a possibilidade de projetar o número de casos esperados para dias seguintes.

De maneira geral, tem-se pouca informação agregada, neste momento, para um entendimento mais amplo do fenômeno COVID-19. Ao observar as diferenças significativas de casos em diferentes países, como por exemplo entre Itália e Alemanha, há a indicação de que variáveis ocultas são importantes para a construção de modelos mais precisos. Relacionado à métrica de números de casos, indícios apontam que diferentes estratégias de isolamento social impactam na suavização da tendência de crescimento exponencial. Como poucos dados ainda estão em análise, espera-se que não se tenha relevância estatística para conclusões com maiores níveis de confiança.

2.3 Competições e Chamados à Ação

As competições no âmbito da Ciência de Dados são um chamado aos esforços dos desenvolvedores, pesquisadores e entusiastas da área de Ciência de Dados em tarefas de alta demanda que podem colaborar na mitigação dos impactos do COVID-19. Estas competições são de natureza diversa, algumas resultando, inclusive, em recompensa financeira para as melhores soluções. A seguir, há uma síntese de competições vigentes e uma breve descrição dos dados disponíveis e tarefas elencadas:

- ***COVID-19 Open Research Dataset Challenge (CORD-19).*** Promovida pela Casa Branca em uma coalizão com grupos de pesquisa de liderança mundial, esta competição dispõe uma base de dados de 45 mil artigos científicos sobre a COVID-19, SARS-CoV-2 e coronavírus relacionados em cerca de 4 GB de dados. Nesta base de dados pública e gratuita, os competidores devem aplicar as mais recentes técnicas de Processamento de Linguagem Natural e Inteligência Artificial para gerarem insights que apoiem a luta vigente contra esta doença infecciosa, auxiliando a comunidade de pesquisa médica a concentrar-se na literatura relevante e simplificar a triagem de informações essenciais. Esta competição contempla dez tarefas que visam buscar respostas sobre fatores de risco, genética, origem, evolução, transmissão e incubação do vírus, dentre outras. Link: <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks>;

- ***COVID19 Global Forecasting.*** Também promovida pela Casa Branca em colaboração com outras instituições, é uma competição semanal que visa promover a busca por previsões acuradas sobre os casos e óbitos do COVID-19. Os autores ressaltam que este não é o objetivo final único, mas sim promover a curadoria de outras fontes de dados com vistas a identificar novas variáveis que impactam na taxa de transmissão do vírus e que também devam ser monitoradas. Link: <https://www.kaggle.com/c/covid19-global-forecasting-week-2>;

- ***AI4D Predict the Global Spread of COVID-19.*** Uma outra competição recente vem da África. Zindi é uma organização social cujo objetivo é construir um ecossistema de Ciência de Dados na África. Recentemente, a organização lançou um desafio de produzir um modelo capaz de prever como o COVID-19 irá se espalhar pelo mundo nos próximos meses. As soluções poderão ser submetidas até o dia 19 de Abril de 2020 e serão avaliadas com dados obtidos no futuro. A competição distribuirá uma premiação às equipes vencedoras no valor de 5.000 dólares. Link: <https://zindi.africa/competitions/predict-the-global-spread-of-covid-19>

A Sociedade Norte-Americana de Radiologia (RSNA, do inglês *Radiological Society of North America*) também têm evocado esforços de pesquisadores da Inteligência Artificial, em especial da área de Aprendizado Profundo aplicado à Visão Computacional, para colaborar no desenvolvimento de métodos de diagnóstico automatizados (RSNA, 2020a). Primeiramente,

estão convocando profissionais da área médica para prover imagens de pacientes e colaborar na produção de rótulos confiáveis nas demais imagens ali reunidas para, por exemplo, guiar tarefas de Aprendizado Supervisionado (RSNA, 2020b). Ademais, agrega publicações recentes, de Março de 2020, com progressos decorrentes desses esforços, em que destaca-se o trabalho de Li et al. (2020) em que redes neurais convolucionais apresentam uma boa acurácia (AUC ROC igual a 0.96) na distinção de imagens radiológicas pulmonares com COVID-19 e outras doenças pulmonares a partir de uma base de dados composta de 4.356 imagens de 3.322 pacientes (LI et al., 2020).

A empresa Nvidia, uma das líderes globais na produção de recursos de hardware para acelerar tarefas de Inteligência Artificial, está disponibilizando gratuitamente, por 90 dias e para pesquisadores de todo o mundo, a infraestrutura Parabricks, composta por unidades de processamento gráfico (GPUs) capazes de acelerar em até 50 vezes a análise do sequenciamento genético do novo coronavírus (NVIDIA, 2020). O entendimento desta questão, segundo a empresa, pode auxiliar na minimização da disseminação da doença e proteger os mais vulneráveis.

O projeto Folding@Home de computação distribuída voltado para realizar simulações computacionais da estrutura de proteínas tem convocado voluntários de todo o mundo a dispor de seus recursos computacionais ociosos para auxiliar na tarefa de simulação da dinâmica das proteínas do COVID-19, visando ao desenvolvimento de novas oportunidades terapêuticas (FOLDING@HOME, 2020). Para tanto, qualquer pessoa que disponha de um computador conectado à internet pode colaborar, baixando e instalando uma aplicação cliente disponível no site do projeto, no seguinte link: <https://foldingathome.org/start-folding/>

3 Desafios e Limitações

Em se tratando de epidemias, as ferramentas da Ciência de Dados para lidar com *Big Data* já resultaram em um caso peculiar. No ano de 2008 foi apresentada uma metodologia para analisar um grande volume de consultas a um mecanismo de busca com vistas a rastrear gripe (*influenza*) e doenças análogas em uma população. Segundo os autores, como a frequência relativa de certos termos de busca estava altamente correlacionada ao número de visitas a médicos para reportar sintomas análogos à gripe, era possível estimar com boa acurácia a incidência de gripe em cada região dos Estados Unidos (GINSBERG et al., 2009). Porém, em 2013, os modelos produzidos em decorrência desta metodologia superestimaram o número de casos de gripe como sendo o dobro do que realmente ocorreu. Embora tenha se mostrado como uma adversidade do uso de tais tecnologias, este fato atraiu a atenção para que resultados dessa natureza devem atuar de maneira complementar, mas não como substituta às redes tradicionais de vigilância epidemiológica (BUTLER, 2013). Este estudo de caso trouxe à tona lições críticas a serem aprendidas à medida que se avança na era do Big Data, em especial no tocante às preocupações com transparência e replicabilidade (LAZER et al., 2014).

Apesar das limitações, alguns autores argumentam que a vigilância digital, por meio do monitoriamento de postagens em redes sociais e de palavras-chaves utilizadas em mecanismos de busca, por exemplo, pode ser útil para obter conhecimento de questões de saúde pública por meio da análise de dados no domínio digital. Em Novembro de 2002, a Rede Global de Inteligência em Saúde Pública (do inglês, *Global Public Health Intelligence Network*), desenvolvida pela Agência de Saúde Pública do Canadá, gerou um alerta para ocorrência de casos da Síndrome Respiratória Aguda Grave (SARS, do inglês *Severe Acute Respiratory Syndrome*) na província de Guangdong na China, o que foi confirmado três meses depois pela OMS (MYKHALOVSKIY; WEIR, 2006). Acredita-se que resultados análogos poderiam ser obtidos para antever epidemias como a do Ebola e de outras doenças infecciosas emergentes, com impactos positivos para além da escala local (MILINOVICH; MAGALHãES; HU, 2015).

No tocante ao uso de Aprendizagem de Máquina e afins na Saúde, existem desafios éticos a lidar para que a sociedade possa então colher os benefícios dessas novas soluções. Os aspectos do sigilo médico precisam ser preservados, os viéses introduzidos ou reproduzidos pelos algoritmos precisam ser criteriosamente investigados e minimizados, as decisões produzidas precisam ser acompanhadas e asseguradas por especialistas humanos visando a diminuir os danos (CHAR; SHAHAND; MAGNUS, 2018) e, sempre que possível, o processo de tomada de decisão precisa ser claro, muito embora isto seja um desafio diante de tantos modelos que funcionam como caixas-preta. Ademais, precisa haver uma avaliação e validação rigorosa de tais modelos.

A ausência de regulamentação do uso de Inteligência Artificial também representa um obstáculo. No âmbito nacional, a Sociedade Brasileira de Computação (SBC) acredita que o Brasil tem o enorme desafio de encontrar uma forma equilibrada de conciliar esses avanços científicos e tecnológicos com avanços sociais e econômicos, tendo como resultado a melhoria da qualidade de vida da população (MACÊDO, 2020). No mundo, a Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) tem feito uma força tarefa de ordem global para garantir que todas as partes envolvidas no projeto e desenvolvimento de sistemas autônomos e inteligentes sejam educadas, treinadas e capacitadas para priorizar considerações éticas para que essas tecnologias avancem para o benefício da humanidade (IEEE, 2020b).

4 Considerações Finais

Diante desta pandemia mundial sem precedentes, tanto em termos das novas características do coronavírus quanto da configuração geopolítica mundial, a Ciência de Dados tem muito a contribuir nesta ocasião e nos anos que se seguem, analisando e agregando múltiplas fontes de dados, gerando inferências e analisando similaridades, aprendendo com padrões análogos e colaborando, assim, para fornecer panoramas mais completos que mitiguem as dificuldades enfrentadas e que colaborem no diagnóstico e no entendimento dos agravos à saúde, de maneira a complementar às técnicas, tecnologias, metodologias e conhecimentos já existentes e que serão desenvolvidos.

Aviso Legal. As informações contidas neste documento devem servir aos interessados como uma primeira orientação. Os autores não possuem qualquer relação com os organizadores das competições nem sobre os valores monetários por elas fornecidos aos seus participantes.

Agradecimentos

Os autores agradecem o apoio financeiro e material provido pela FAPEAM/CNPq por meio do Edital PPP 04/2017. Os autores também agradecem todo o apoio provido pela Escola Superior de Tecnologia (EST) da Universidade do Estado do Amazonas (UEA) às atividades do Laboratório de Sistemas Inteligentes (LSI).

Referências

ABEDJAN, Z. et al. Data science in healthcare: Benefits, challenges and opportunities. In: SERGIO CONSOLI AND DIEGO REFORGIATO RECUPERO AND MILAN PETKOVIC. *Data Science for Healthcare – Methodologies and Applications*. Suíça: Springer, 2019. p. 3–38.

AMARAL, F. *Introdução à Ciência de Dados – Mineração de Dados e Big Data*. Rio de Janeiro: Alta Books, 2016. 320 p.

BRASIL.IO. *COVID-19. 2020*. Disponível em <https://brasil.io/dataset/covid19/caso>. Acessado em 31 de março de 2020.

BUTLER, D. *Detecting influenza epidemics using search engine query data*. Nature, n. 494, p. 155–156, 2013.

CADY, F. *The Data Science Handbook*. Estados Unidos: Wiley, 2017.

CHAR, D. S.; SHAHAND, N. H.; MAGNUS, D. *Implementing machine learning in health care –addressing ethical challenges*. N Engl J Med, v. 378, n. 11, 2018.

CONSOLI, S.; RECUPERO, D. R.; PETKOVIC, M. *Data Science for Healthcare – Methodologies and Applications*. Suíça: Springer, 2019. 367 p.

EVANGELISTA, L.; GUEDES, E. B. Ensembles of convolutional neural networks on computer-aided pulmonary tuberculosis detection. *IEEE Latin American Transactions*, v. 17, n. 12, p. 1954–1963, 2020.

FOLDING@HOME. *Coronavirus - what we're doing and how you can help in simple terms*. 2020. Disponível em <https://foldingathome.org/covid19/>. Acessado em 31 de março de 2020.

GINSBERG, J. et al. D*etecting influenza epidemics using search engine query data*. Nature, n. 12, p. 1012–1014, 2009.

GREENSPAN, H.; GINNEKEN, B. van; SUMMERS, R. M. Guest editorial Deep Learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, v. 35, n. 5, p. 1153–1159, 2016.

IEEE. *Corona Virus (COVID-19) Tweets Dataset. 2020*. Disponível em <https://ieee-dataport.org/open-access/corona-virus-covid-19-tweets-dataset>. Acessado em 31 de março de 2020.

IEEE. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. 2020. Disponível em <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>. Acessado em 31 de março de 2020.

JHU. *Coronavirus Resource Center*. 2020. Disponível em <https://coronavirus.jhu.edu/map.html>. Acessado em 31 de março de 2020.

KAGGLE. *Novel Corona Virus 2019 Dataset*. 2020. Disponível em <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>. Acessado em 31 de março de 2020.

LAZER, D. et al. The parable of google flu: Traps in big data analysis. *Science*, v. 6176, n. 343, p. 1203–1205, 2014.

LEE, J. G. et al. Deep learning in medical imaging: General overview. *Korean J. Radiol.*, v. 18, n. 4, p. 570–584, 2017.

LI, L. et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*, v. 3, p. 200905, 2020.

LIMA, J. et al. Pigmented dermatological lesions classification using convolutional neural networks ensemble mediated by multilayer perceptron network. *IEEE Latin American Transactions*, v. 17, n. 11, p. 1902–1908, 2019.

MACÊDO, R. J. de A. *SBC se manifesta sobre regulação de IA.* 2020. Disponível em <https://www.sbc.org.br/noticias/10-slideshow-noticias/2181-sbc-se-manifesta-sobre-regulacao-de-ia>. Acessado em 31 de março de 2020.

MARZ, N.; WARREN, J. *Big Data – Principles and best practices of scalable real-time data systems*. Shelter Island: Manning Publications, 2015. 330 p.

MILINOVICH, G. J.; MAGALHãES, R. J. S.; HU, W. Role of Big Data in the early detection of Ebola and other emerging infectious diseases – correspondence. *The Lancet Global Health*, v. 3, n. 1, p. e20–e21, 2015.

MYKHALOVSKIY, E.; WEIR, L. The global public health intelligence network and early warning outbreak detection – a canadian contribution to global public health. *Can J Public Health*, v. 97, n. 1, p. 42–44, 2006.

NVIDIA. *NVIDIA Gives COVID-19 Researchers Free Access to Parabricks – Genome-sequencing software for GPUs accelerates the journey to understanding the novel coronavirus.* 2020. Disponível em <https://blogs.nvidia.com/blog/2020/03/19/coronavirus-research-parabricks/>. Acessado em 31 de março de 2020.

OMS. *Coronavirus disease (COVID-19) Pandemic*. 2020. Disponível em <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Acessado em 31 de março de 2020.

OMS. *Situation Reports.* 2020. Disponível em <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. Acessado em 31 de março de 2020.

OMS. *WHO Director-General's opening remarks at the media briefing on COVID-19*. 2020. Disponível em <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>. Acessado em 31 de março de 2020.

PROVOST, F.; FAWCETT, T. *Data Science para Negócios*. Rio de Janeiro: Alta Books, 2016. 408 p.

PYTHON. *covid SDK*. 2020. Disponível em <https://pypi.org/project/covid/>. Acessado em 31 de março de 2020.

RSNA. *COVID-19 Resources.* 2020. Disponível em <https://www.rsna.org/covid-19>. Acessado em 31 de março de 2020.

RSNA. *RSNA Coronavirus cases*. 2020. Disponível em <https://cases.rsna.org/coronavirus>. Acessado em 31 de março de 2020.

WORLDOMETER. *COVID-19 coronavirus pandemic.* 2020. Disponível em <https://www.worldometers.info/coronavirus/>. Acessado em 31 de março de 2020.

Sobre os Autores

Elloá B. Guedes. Doutora em Ciência da Computação pela Universidade Federal de Campina Grande. Atua como professora adjunta da Escola Superior de Tecnologia da Universidade do Estado do Amazonas desde 2013, onde lidera o Grupo de Pesquisas em Sistemas Inteligentes e co-fundou o Laboratório de Sistemas Inteligentes. Seus interesses de pesquisa concentram-se em Aprendizado de Máquina, Aprendizado Profundo, Ciência de Dados e Inteligência Artificial.
Currículo Lattes: <http://lattes.cnpq.br/6466781778573760>
E-mail: ebgcosta@uea.edu.br
Site pessoal: <www.elloaguedes.com>.

Sociedade Brasileira para o Progresso da Ciência

Secretaria de Desenvolvimento Econômico, Ciência, Tecnologia e Inovação

AMAZONAS
GOVERNO DO ESTADO

Carlos Maurício Seródio Figueiredo. Professor Adjunto da Escola Superior de Tec-nologia da Universidade do Estado do Amazonas (UEA), possui graduação em Eng. Elétrica pela Universidade Federal do Amazonas, mestrado e doutorado em Ciência da Computação pela Universidade Federal de Minas Gerais. É pesquisador co-fundador do Lab. de Sistemas Inteligentes, atuando também no Samsung Ocean, em áreas relacionadas a Internet das Coisas e Aprendizado de Máquina Aplicado.
Currículo Lattes: <http://lattes.cnpq.br/9060002746939878>
E-mail: cfigueiredo@uea.edu.br
Site pessoal: <https://br.linkedin.com/in/cmsfigueiredo>

Tiago Eugenio de Melo. Possui graduação em Processamento de Dados e Direito pela Universidade Federal do Amazonas (UFAM), especialização em Geoprocessamento pela Universidade Federal do Amazonas (UFAM), especialização em Direito do Trabalho e Processo do Trabalho pela Universidade Anhanguera, Mestrado em Ciência da Computação pela Universidade Estadual de Campinas (UNICAMP) e Doutorado em Informática pela Universidade Federal do Amazonas (UFAM). Atualmente é professor concursado da Universidade do Estado do Amazonas (UEA) e pesquisador-bolsista do Samsung Ocean Manaus. Tem atuado principalmente nos seguintes temas: análise de sentimentos, processamento de linguagem natural, recuperação de informação e aprendizagem de máquina.
Currículo Lattes: <http://lattes.cnpq.br/9912454472927669>
E-mail: tmelo@uea.edu.br
Site pessoal: <http://tiagodemelo.info>.

NOTA TÉCNICA
COVID-19
Nº 002
28/04/2020

Sociedade Brasileira para o Progresso da Ciência

Secretaria de **Desenvolvimento Econômico, Ciência, Tecnologia e Inovação**

AMAZONAS
GOVERNO DO ESTADO