



PLANO DE ENSINO – PPGICS – DISCIPLINAS

() VERÃO () INVERNO (X) ELETIVA () OBRIGATÓRIA

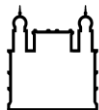
IDENTIFICAÇÃO			
Disciplina: : Ciência de Dados aplicada à Saúde			
Código: PPGICS-DM006	Créditos: 04	Carga Horária: 60h	Período Início: 17/03/2023 Término: 30/06/2023 Dia da semana: Sexta-feira Horário: das 9h às 13h
Código: ICS-DM0924	Créditos: 04	Carga Horária: 120	
Coordenador da Disciplina: Marcel Pedroso Professores: Alexandre Chiavegatto (FSP/USP), Fabio Porto (LNCC), Jefferson Lima (PCDaS), Raphael Saldanha (PCDaS), Eduardo Ogasawara (Cefet e PCDaS) e Rebecca Pontes Salles (Cefet e PCDaS).			
Curso: (X) Mestrado (X) Doutorado Núcleo Comum ()			
Linha 1 ()		Linha 2 ()	Linha 3 (X)

EMENTA DA DISCIPLINA (PREENCHER SOMENTE SE FOR ELETIVA REGULAR)

Introdução à Ciência de Dados. Sistemas de Informação em Saúde (SIS). Introdução a linguagem de programação Python. Análise exploratória de dados usando Python. Análise avançada de dados: Aprendizagem de Máquina (Machine Learning). Tecnologias e ferramentas livres em Ciência de Dados. Ética, transparência e interpretabilidade em modelos de inteligência artificial. Atividades práticas integradas. Apresentação do projeto de trabalho final utilizando a Plataforma de Ciência de Dados aplicada à Saúde.

EMENTA DO CURSO

Introdução à Ciência de Dados. Sistemas de Informação em Saúde (SIS). Introdução a linguagem de programação Python. Análise exploratória de dados usando Python. Análise avançada de dados: Aprendizagem de Máquina (Machine Learning). Tecnologias e ferramentas livres em Ciência de Dados. Ética, transparência e interpretabilidade em modelos de inteligência artificial. Atividades práticas integradas (monitoria). Apresentação do projeto de trabalho final utilizando a Plataforma de Ciência de Dados aplicada à Saúde. Python é uma linguagem de programação de uso geral que está se tornando cada vez mais popular na ciência de dados. Este curso propõe uma introdução ao Python com o foco em análise de dados. Durante o curso serão abordados os principais conceitos e métodos de armazenamento; manipulação e ferramentas úteis para análise de dados de saúde e seus determinantes socioambientais. A disciplina integra as atividades e objetivos do projeto de pesquisa e desenvolvimento tecnológico “Plataforma de Ciência de Dados aplicada à Saúde” coordenado pelo Laboratório de Informação em Saúde do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz (Lis/Icict/Fiocruz) em parceria com o Laboratório Nacional de Computação Científica (LNCC). <https://pcdas.icict.fiocruz.br/>.



OBJETIVOS

Geral: Promover interface entre aspectos teóricos e práticos sobre Ciência de Dados, Mineração de Dados, Aprendizagem de Máquina, Análise Preditiva e Análise Visual de grandes ou complexas bases de dados do setor saúde e de seus determinantes socioambientais.

Específicos: Introdução aos principais conceitos e métodos de armazenamento; manipulação e ferramentas úteis para análise de dados em linguagem de programação Python. Apresentação e manuseio dos principais sistemas de informação em saúde e de interesse para a saúde. Capacitação dos alunos em abordagens teóricas e metodológicas para a análise de grandes quantidades de dados em diferentes formatos por meio de estratégias e técnicas relacionadas a Ciência de Dados aplicada à Saúde. Discutir os principais conceitos sobre ética, transparência e interpretabilidade em modelos de inteligência artificial aplicados ao SUS. Fomentar a utilização da Plataforma de Ciência de Dados aplicada à Saúde do LIS/Icict promovendo inovação tecnológica e aprendizagem colaborativa.

BIBLIOGRAFIA BÁSICA

AGARWAL, Ritu; DHAR, Vasant. Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. **Information Systems Research**, [S. l.], v. 25, n. 3, p.443-448, Sep. 2014.

ALPAYDIN, Ethen. **Introdução ao Machine Learning**. Cambridge: MIT Press, 2010.

BOHON, S. A. Demography in the Big Data Revolution: changing the culture to forge new frontiers. **Population Research and Policy Review**, [S. l.], v. 37, n. 3, p. 323-341, 28 jun. 2018.

CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados**. Conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

CHIAVEGATTO FILHO, A. D. P. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, Brasília, v. 24, p. 325-332, 2015.

DOS SANTOS, H. G.; DO NASCIMENTO, C. F.; IZBICKI, R.; DUARTE, Y. A. O.; CHIAVEGATTO FILHO, A. D. P. Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 35, p. 1, 2019.

EDITORIAL LANCET. GBD 2015: from big data to meaningful change. **The Lancet**, [S. l.], v. 388, Oct. 08, 2016, DOI: [https://doi.org/10.1016/S0140-6736\(16\)31790-1](https://doi.org/10.1016/S0140-6736(16)31790-1) .

GLOBAL GRAND CHALLENGES. **Grand Challenges Explorations** – Brazil: Data Science Approaches to Improve Maternal and Child Health in Brazil [Internet]. 2019 [citado em 2019 ago. 15]. Disponível em: <https://gcgh.grandchallenges.org/challenge/grand-challenges-explorationsbrazil-data-science-approachesimprove-maternal-and-child> . Acesso em: dez. 2022.

INSTITUTO DE COMUNICAÇÃO E INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE. **Plataforma de Ciência de Dados aplicada à Saúde (PCDaS)**. In: Instituto de Comunicação e Informação Científica e Tecnológica em Saúde – Laboratório de Informação em Saúde (LIS). [Rio de Janeiro]: Fundação Oswaldo Cruz. Disponível em: <https://pcdas.icict.fiocruz.br/>. Acesso em: dez. 2022.



KUMAR, Vipin; STEINBACH, Michael; TAN, Pang-Ning. **Introdução ao DATAMINING Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2009.

MATSUDAIRA, Kate. The Science of Managing Data Science. Association for Computing Machinery. **Communications of the ACM**, [S. l.], v. 58, n. 6, p.44, Jun. 2015.

MATTMANN, Chris A. A vision for data science: to get the best out of big data, funding agencies should develop shared tools for optimizing discovery and train a new breed of researchers. **Nature**, [S. l.], v. 493, n. 7.433, p.473-483, Jan. 24 2013.

MITCHELL, Tom. **Machine Learning**. New York: McGraw-Hill, 1997.

OBERMEYER, Z.; POWERS, B.; VOGELI, C.; MULLAINATHAN, S. Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, [S. l.], v. 366, n. 6464, p. 447-453, 2019. <https://doi.org/10.1126/science.aax2342>.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, [S. l.], v. 12, p. 2.825-2.830, 2011.

SALDANHA, R. F.; BARCELLOS C.; PEDROSO, M. M. Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde?. **Cad Saúde Colet.**, Rio de Janeiro, n. 29 (esp.), p. 51-58, 2021. <https://doi.org/10.1590/1414-462X202199010305>

SHMUELI, G. To Explain or To Predict? **Statistical Science**, Berkley, v. 25, n. 3, p. 289-310, May 2010. <http://dx.doi.org/10.2139/ssrn.1351252>.

SIMON, H. A. **The sciences of the artificial**. Massachusetts: MIT Press, 1981.

BIBLIOGRAFIA COMPLEMENTAR (opcional)

CRITÉRIOS DE AVALIAÇÃO

Participação nas atividades presenciais (1,0 ponto).

Aprovação no curso online e gratuito “Introduction to Python” na plataforma DataCamp (1,0 ponto).

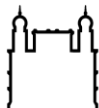
Resolução de exercícios práticos (EP) (4,0 pontos, sendo 1,0 por EP entregue via Google Colaboratory). Ambiente de notebooks Jupyter (não requer configuração e é executado na nuvem gratuita).

Participar de competição acadêmica de machine learning na plataforma Kaggle (2,0 pontos para 1º colocado e 1,0 ponto para cada participante com score acima do mínimo estabelecido).

Entrega de trabalho final utilizando a Plataforma de Ciência de Dados aplicada à Saúde compartilhado por meio da ferramenta Jupyter Notebook da disciplina (2,0 pontos).

CRONOGRAMA DAS AULAS (POR ENCONTRO)

Aula 1	Boas-vindas e pactuação da dinâmica do curso (Avaliação, Exercícios Práticos, comunicação assíncrona)
---------------	--



17/03	
Aula 2 24/03	Introdução à Ciência de Dados aplicada à Saúde e ambientação Google Colab
Aula 3 31/03	Análise exploratória de dados usando Python Introdução e principais conceitos
Aula 4 07/04	Extraindo e manipulando dados
Aula 5 14/04	Análise exploratória de dados
Aula 6 21/04	Análise avançada de dados: machine learning Transformação, padronização e normalização de dados
Aula 7 28/04	Valores faltantes e feature engineering
Aula 8 05/05	Modelos de estimação (regressão)
Aula 9 12/05	Modelos de classificação
Aula 10 19/05	PCA e KMeans
Aula 11 26/05	Ética e interpretabilidade em machine learning
Aula 12 02/09	Atividades práticas integradas (laboratório e monitoria)
Aula 13 16/06	Apresentação (em grupo) do projeto de trabalho final utilizando a PCDaS
Aula 14 23/06	Apresentação (em grupo) do projeto de trabalho final utilizando a PCDaS
Aula 15 30/06	Atividades práticas integradas (laboratório e monitoria)

Rio de Janeiro, 30 de dezembro de 2022.