

## PLANO DE ENSINO – PPGICS - 2022.1

IDENTIFICAÇÃO			
<b>NOME DA DISCIPLINA:</b> Ciência de Dados aplicada à Saúde			
<b>CÓDIGO:</b> ICS-DM092	<b>CRÉDITOS:</b> 04	<b>CARGA HORÁRIA:</b> 60 horas	<b>LINHA:</b> 03
<b>DOCENTE RESPONSÁVEL:</b> Marcel Pedroso			
<b>DOCENTE(S) ASSISTENTE(S) / AUXILIAR(ES) / CONVIDADO(A)(S):</b>  Alexandre Chiavegatto (FSP/USP), Fabio Porto (LNCC), Jefferson Lima (PPGICS e PCDaS), Raphael Saldanha (PPGICS e PCDaS), Igor da Silva Moraes (CEFET e PCDaS) e Rebecca Pontes Salles (CEFET e PCDaS).			
<b>INÍCIO:</b>  06/04/2022	<b>TÉRMINO:</b>  13/07/2022	<b>DIA DA SEMANA:</b>  Quarta-feira	<b>HORÁRIO:</b>  9h00 à 13h00

### OBJETIVOS

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/>

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;
- Fomentar a utilização da Plataforma de Ciência de Dados aplicada à Saúde do LIS/ICICT promovendo inovação tecnológica e aprendizagem colaborativa.

## PLANO DE ENSINO – PPGICS - 2022.1

### CRITÉRIOS DE AVALIAÇÃO

- Participação nas atividades síncronas (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 que 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).

### ESTRUTURA CURRICULAR RESUMIDA

1. Introdução à Ciência de Dados
2. Sistemas de Informação em Saúde (SIS)
3. Introdução a linguagem de programação Python
4. Análise exploratória de dados usando Python
5. Análise avançada de dados: Aprendizagem de Máquina (*Machine Learning*)
6. Tecnologias e ferramentas livres em Ciência de Dados
7. Ética e interpretabilidade em *Machine Learning*
8. Atividades práticas integradas (monitoria)
9. Apresentação do projeto de trabalho final utilizando a Plataforma de Ciência de Dados aplicada à Saúde

**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMAÇÃO E  
COMUNICAÇÃO EM SAÚDE  
PPGICS/ICICT/FIOCRUZ**

**Rio de Janeiro, 15 de fevereiro de 2022.**



## BIBLIOGRAFIA BÁSICA

- Agarwal, Ritu, Dhar, Vasant. Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research, Sep 2014, Vol.25(3), pp.443-448.
- Alpaydin, Ethen. Introdução ao Machine Learning. MIT Press, 2010.
- Bohon, S. A. Demography in the Big Data Revolution: Changing the Culture to Forge New Frontiers. Population Research and Policy Review, 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, v. 24, p. 325-332, 2015.
- dos Santos, H. G. ; do Nascimento, CF ; Izbicki, R. ; Duarte, Y.A.O. ; Chiavegatto Filho, A.D.P. . Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. CADERNOS DE SAÚDE PÚBLICA, v. 35, p. 1, 2019.big
- Editorial Lancet. GBD 2015: from big data to meaningful change. The Lancet. Vol 388, October 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-explorations-brazil-data-science-approachesimprove-maternal-and-child>
- 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, Jan 24, 2013, Vol.493(7433), p.473(3).
- Matsudaira, Kate. The Science of Managing Data Science. Association for Computing Machinery. Communications of the ACM, Jun 2015, Vol.58(6), p.44
- Mitchell, Tom. Machine Learning. 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, vol. 366, no. 6464, p. 447–453, 2019. <https://doi.org/10.1126/science.aax2342>.
- Pedregosa, et al. Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830, 2011.
- PCDaS. Plataforma de Ciência de Dados aplicada à Saúde. Laboratório de Informação em Saúde (Lis). Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict). Fundação Oswaldo Cruz (Fiocruz). Disponível em: <https://pcdas.icict.fiocruz.br/> . Acesso em janeiro de 2022.
- Saldanha RF, Barcellos C, Pedroso MM. Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde?. Cad Saúde Colet, 2021; Ahead of Print. <https://doi.org/10.1590/1414-462X202199010305>
- Simon, H. A. The sciences of the artificial. Massachusetts: MIT Press, 1969 [1981].
- Shmueli, Galit, To Explain or To Predict? (May 24, 2010). Statistical Science. Available at SSRN: <https://ssrn.com/abstract=1351252> or <http://dx.doi.org/10.2139/ssrn.1351252>
- Vipin Kumar, Michael Steinbach, Pang-Ning Tan. Introdução ao DATAMINING Mineração de Dados. Rio de Janeiro: Editora Ciência Moderna Ltda, 2009.