

Disciplina: Data Science (DS)

Professor: Fabrício J. Barth

Objetivo da Disciplina

A prática de Data Science é baseada em um conjunto de metodologias, processos, conceitos e ferramentas. É objetivo deste curso discutir e exercitar tais metodologias, processos, conceitos e ferramentas para que os participantes do curso possam fazer uso da prática de Data Science no seu dia-a-dia. Ao compreender melhor a prática de Data Science, espera-se que os participantes do curso possam: (i) identificar potenciais projetos de Data Science; (ii) avaliar a complexidade de cada projeto; (iii) acompanhar a execução dos projetos; (iv) identificar skills relevantes para a formação de equipes de Data Science, e; (v) interagir com profissionais técnicos da área. Este curso não tem como objetivo formar engenheiros de machine learning, engenheiros de dados, cientista de dados ou qualquer outro profissional que compreenda em profundidade os algoritmos e ferramentas normalmente empregados no desenvolvimento de modelos preditivos ou descritivos. No entanto, é também objetivo deste curso discutir os conceitos e cenários de utilização dos principais algoritmos da prática de Data Science.

Mini CV Professor

Fabrício J. Barth. Líder Técnico de Data & AI na IBM. Formado em Ciências da Computação e com Doutorado em Engenharia Elétrica pela Escola Politécnica da USP. Tem desenvolvido diversos projetos relacionados com Aprendizagem de Máquina e Inteligência Artificial nas áreas financeira, Internet, segurança pública, mineração e recursos humanos desde 2003. Possui alguns artigos publicados, principalmente sobre: recuperação de informação, aprendizagem de máquina, mineração de textos e sistemas de recomendação. Atua como professor em cursos de graduação e pós-graduação desde 2002. Alguns projetos desenvolvidos foram: Bradesco Inteligência Artificial (BIA), A Voz da Arte no Museu da Pinacoteca e IRIS+ no Museu do Amanhã.

Plano de Aula

Horário	Tema e conteúdo
8:00 - 8:15	Apresentação da turma
8:15 - 9:00	Processos para projetos de Data Science e Introdução à Aprendizagem de Máquina
9:00 - 10:00	Desenvolvendo um modelo preditivo do tipo regressão
10:00 - 10:20	<i>Coffee-Break</i>
10:20 - 12:00	Desenvolvendo um modelo preditivo do tipo classificação
12:00 - 14:00	Almoço
14:00 - 15:40	Desenvolvendo um modelo descritivo do tipo clustering
15:40 - 16:00	<i>Coffee-Break</i>
16:00- 17:00	Projetos de Data Science e impactos nos negócios
17:00- 18:00	Conversa com Armen Pischdotchian e Fechamento

Bibliografia Básica

- [1] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230> [<https://github.com/fbarth/ds-saint-paul/blob/master/references/fay1996.pdf>]
- [2] Mitchell, T. The Discipline of Machine Learning. 2006. <http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf> [<https://github.com/fbarth/ds-saint-paul/blob/master/references/MachineLearning.pdf>]
- [3] WICKHAM, Hadley. Tidy Data. *Journal of Statistical Software*, [S.l.], v. 59, Issue 10, p. 1 - 23, sep. 2014. ISSN 1548-7660. Available at: <<https://www.jstatsoft.org/v059/i10>>. doi:<http://dx.doi.org/10.18637/jss.v059.i10>. [<https://github.com/fbarth/ds-saint-paul/blob/master/references/tidy-data.pdf>]
- [4] *Foundational Methodology for Data Science*. Material que faz parte do curso Data Science Practitioner da plataforma Skills Academy da IBM. <https://skills-academy.mylearnerportal.com/> [https://github.com/fbarth/ds-saint-paul/blob/master/references/foundational_ds_ibm.PDF]

- [5] Vasant Dhar. 2013. Data science and prediction. Commun. ACM 56, 12 (December 2013), 64-73.
DOI:<https://doi.org/10.1145/2500499> [https://github.com/fbarth/ds-saint-paul/blob/master/references/ds_prediction.pdf]

Bibliografia Complementar

- [1] Plataforma Skills Academy da IBM. <https://skills-academy.mylearnerportal.com/>
- [2] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.

Método de Avaliação

Avaliação	Participação	
Pre Work	40%	<p>Questionário a ser preenchido pelo participante do curso sobre casos de uso e sobre a literatura referência do curso.</p> <p>As instruções estão detalhadas no arquivo de Pre Work.</p> <p><i>Deverá ser entregue até às 23h59 do dia 27/Jul/2020, via Blackboard.</i></p>
Homework	60%	<p>Questionário a ser preenchido pelo participante do curso sobre casos de uso discutidos ao longo do curso.</p> <p>As instruções serão detalhadas durante a aula do dia 30/Jul/2020.</p> <p><i>Deverá ser entregue até às 23h59 do dia 27/Ago/2020, via Blackboard.</i></p>