

## Programa de Mestrado Profissional em Economia

Disciplina: Fundamentos de Ciências de Dados

Professor: Genaro Dueire Lins

### PROPOSTA

Este curso apresenta os principais fundamentos de Ciência de Dados, através do uso intensivo da linguagem de programação R. O objetivo do curso é dar ao aluno ferramental teórico, e prático, para importar, explorar e modelar bases de dados. O conjunto de técnicas exploradas inicia com a revisão das técnicas usuais para modelagem (regressão, log e probit), para focar nas técnicas mais usuais de *Machine Learning*, como árvores classificação, *random forest* e *support vector machine (SVM)*.

### METODOLOGIA E PEDAGOGIA

- Aulas expositivas com exposição teórica.
- Discussões sobre aplicações práticas da teoria apresentada.
- Aulas práticas em laboratório

### PROGRAMA

- Contextualização Ciência de Dados
- Introdução ao R
  - Princípios da linguagem R - RStudio
  - Visualização de dados – ggplot
  - Funções e operações básicas
  - Análise exploratória
  - Ferramentas de importação
  - Manipulação de diversas bases de dados
- Revisão de álgebra linear e estatística.
- Regressão, logit e probit
- Classificadores bayesianos
- Árvores de classificação
- *K-Nearest Neighbors*
- *Classifier ensembles*
- *Random forest*
- *Support vector machine*
- Técnicas de avaliação de modelos

## BIBLIOGRAFIA

- GROLEMUND, G.; WICKHAM, H. R for Data Science. O'Reilly, 2016.
- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. USA: John Wiley & Sons Asia PTE, 2001.
- STRANG, G. Introduction to linear algebra. Fifth Edition. Wellesley - Cambridge Press, 2016.
- WOOLDRIDGE, J. M. Introductory Econometrics: A Modern Approach. [s.l.] Thomson - South Western, 2009.
- ABU-MOSTAFA, Y. S.; MAGDON-ISMAIL, M.; LIN, H.-T. Learning From Data. USA: AM Book, 2012.

## AVALIAÇÃO

- Exercícios ..... 20%
- Avaliação ..... 40%
- Trabalho de Conclusão ..... 40%

PROGRAMAÇÃO AULA-A-AULA

Aula	Data	Tópico	Conteúdo
1		Apresentação	Introdução ao Conceito do Ciência de Dados
2		Introdução ao R	Princípios da linguagem R - RStudio Visualização de Dados – ggplot Funções e Operações Básicas Análise Exploratória Ferramentas de Importação Manipulação de diversas bases de dados
3		Fundamentos Teóricos	Revisão de estatística e álgebra linear Regressão, Logit e Probit Componentes Principais
4			Técnicas de avaliação de modelos
5			Classificadores Bayesianos
6		Laboratório	Árvores de Decisão Exercícios Práticos
7		Fundamentos Teóricos	<i>K-Nearest Neighbors</i>
8			<i>Classifier ensembles</i>
9			<i>Random forest</i>
10		Laboratório	<i>Support vector machine</i> Exercícios práticos
11			Apresentação trabalho final
12			Avaliação (teórica)