



# MINERAÇÃO DE DADOS

**Prof. Julio Cesar dos Reis**

**[jreis@ic.unicamp.br](mailto:jreis@ic.unicamp.br)**

**[www.ic.unicamp.br/~jreis](http://www.ic.unicamp.br/~jreis)**

CT-0611

# Objetivos

2

- Estudar processo de descoberta de conhecimento a partir de dados.
- Aprender algoritmos, técnicas e ferramentas para o desenvolvimento de modelos para mineração de dados.
- Entender métricas de desempenho e as limitações das abordagens de prospecção da informação.

# Conteúdo

3

- ❑ Processo de Descoberta do Conhecimento e tratamento de dados
- ❑ Técnicas para seleção de atributos e separação de conjunto de dados para treinamento
- ❑ Introdução à Python, Numpy, Pandas
- ❑ Algoritmos para classificação de dados
- ❑ Algoritmos para agrupamentos de dados
- ❑ Métodos e Métricas de avaliação

# Dinâmica de ensino-aprendizagem

4

- Aulas online síncronas
- Resolução de lista de exercícios - questionários
- Resolução de dúvidas online
- Resolução de atividades

# Introdução

5

- Conceitos básicos
- Tratamento de dados
- Introdução à Python, Numpy, Pandas

# Classificação

6

- Introdução à classificação de dados
- Algoritmo de árvores de decisão
- Classificação com python

# Agrupamento

7

- Agrupamentos de dados
- Agrupamento de dados com python

# Aulas online

9

## □ Via MS teams

Sujeito a modificações

	15/03	22/03	29/03
9:00 - 10:50	A0 - Plano de Ensino	A5 - Introdução a Visualização	A8 - Classificação II
	A1 - Conceitos básicos	handson tratamento dados e visualização	handson classificação II
11:10 - 13:00	A2 - Tratamento de dados	A6 - Classificação I	A9 - Agrupamento
		A7 - Métodos e Métricas de Avaliação	
<i>almoço</i>			
14:00 - 15:50	A3 - Introdução a Python	handson classificação I	handson Agrupamento
	handson python	com python	com python
16:10 - 18:00	A4 - Introdução a numpy/pandas	TAREFA 1 (Classificação)	TAREFA 2 (Agrupamento)
	handson numpy/pandas		



# Critérios de avaliação

14

- Questionários [individual]
  - ▣ Peso: 20%
  - ▣ Entrega: datas definidas
  
- Atividade 1 [em dupla]
  - ▣ Peso: 40%
  - ▣ Entrega: 29/03/2025
  
- Atividade 2 [em dupla]
  - ▣ Peso: 40%
  - ▣ Entrega: 06/04/2025

# Critérios de avaliação

15

- Nota parcial = média ponderada das notas das avaliações.
- Nota no intervalo  $[0,10]$ .
- Se nota parcial  $\geq 7.0$ , aluno(a) está aprovado(a) sem exame final.

# Exame

16

- O exame consistirá de uma prova teórica contendo  $n$  questões com o conteúdo do curso todo.
- Nota do exame no intervalo  $[0,10]$ .
- Data a ser combinada com a coordenação do curso.
- Nota final =  $(\text{nota parcial} + \text{nota do exame}) / 2$ .
  - ▣ O(a) aluno(a) estará aprovado(a) caso sua nota final  $\geq 7.0$

# Bibliografia

21

- Charu C. Aggarwal (2015) Data Mining: The Textbook. Springer International Publishing, 1st edition.
- Tan, Steinbach, Kumar. (2006) Introduction to Data Mining. Pearson
- Max Bramer. (2007) Principles of Data Mining. Springer International Publishing, 3<sup>rd</sup> edition.
- Trevor Hastie and Robert Tibshirani (2016) The Elements of Statistical Learning: Data Mining, Inference, and Prediction.