

Plano de Ensino - 2020.QS  
Mineração de dados

Thiago Ferreira Covões

# Objetivo

- ▶ Assimilar conceitos de mineração de dados e descoberta de conhecimento em bases dados
- ▶ Apresentar/experimentar técnicas de mineração de dados e descoberta de conhecimento de bases de dados para a modelagem de problemas em áreas científicas e tecnológicas

# Conteúdo

Semana	Conteúdo
1	Introdução - Análise Exploratória
2	Classificação I
3	Classificação II
4	Regressão e Avaliação de Modelos
5	Reuniões de acompanhamento dos projetos
6	Agrupamento de dados I
7	Agrupamento de dados II
8	Comitês
9	Seleção e extração de atributos
10	Regras de associação
11	Apresentação dos trabalhos
12	Séries Temporais

# Carga Horária

- ▶ É estimado um tempo de dedicação de aproximadamente 8h para cada semana:
  - ▶ Aproximadamente 3h30m para acompanhar os vídeos;
  - ▶ Aproximadamente 2h para realização de listas de exercícios;
  - ▶ Aproximadamente 2h30m para desenvolvimento dos projetos.

## Ferramentas utilizadas

- ▶ As aulas serão disponibilizadas em formato de vídeo no Youtube, com os links indicados no SIGAA e na página do curso
- ▶ As listas de exercícios devem ser entregues via Github Classroom (detalhes na página do curso).

## Ferramentas utilizadas

- ▶ Os plantões de dúvidas serão feitos utilizando:
  - ▶ **Discord**: para sessões síncronas, possivelmente com compartilhamento de tela, com dúvidas sendo apresentadas via voz ou chat (plantão no formato síncrono);
  - ▶ **Discord**: canal com todos os alunos para dúvidas via chat (plantão no formato assíncrono);
  - ▶ Via e-mail para os alunos que preferem não utilizar as ferramentas mencionadas.

# Plantões síncronos

- ▶ Para facilitar aos alunos que tiverem problemas de disponibilidade de horário, as seguintes opções serão dadas neste tipo de plantão:
  - ▶ Segunda à quinta-feira: 18h-19h;

# Avaliação

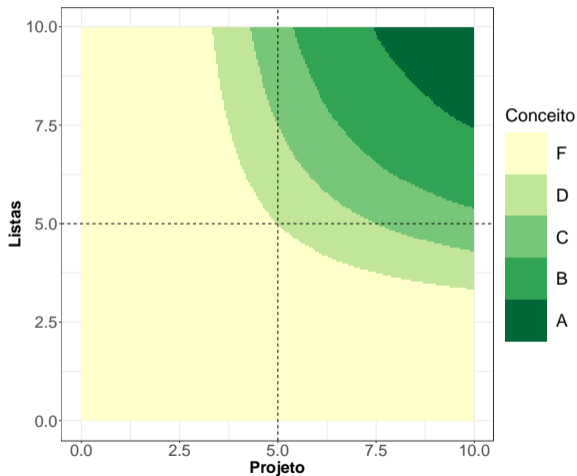
- ▶ Código de Honra
- ▶ Listas de Exercícios
- ▶ 2 Projetos

$$N_{\text{projeto}} = \frac{2}{\frac{1}{\max(0.01, N_{\text{Proj1}})} + \frac{1}{\max(0.01, N_{\text{Proj2}})}}$$

$$N_{\text{final}} = \frac{2}{\frac{1}{\max(0.01, N_{\text{lista}})} + \frac{1}{\max(0.01, N_{\text{projeto}})}}$$



# Média harmônica



# Conceitos

Se  $N_{\text{final}} \geq 8.5 \rightarrow A$

Senão Se  $N_{\text{final}} \geq 7 \rightarrow B$

Senão Se  $N_{\text{final}} \geq 5 \rightarrow C$

Senão  $N_{\text{final}} \rightarrow F$

# Projeto

- ▶ Grupos de até 2 pessoas
- ▶ 2 Seminários
  - ▶ Algoritmos
  - ▶ Aplicações
- ▶ Relatório e Vídeo

# Listas

- ▶ Listas envolvendo exercícios teóricos e de implementação
- ▶ Entregues em Markdown/PDF
- ▶ Em cada lista terá um conjunto de exercícios selecionados para explicação em vídeo
  - ▶ Disponibilizar o vídeo da explicação da resolução em link (para Google Drive/One Drive/Youtube/Dropbox) na entrega
  - ▶ No início do vídeo deve aparecer tanto o rosto quanto documento para identificação
  - ▶ Só serão avaliadas as listas em que foi possível obter o vídeo
  - ▶ O vídeo deve ter duração máxima de 5 minutos

# Dúvidas

- ▶ Contato via **e-mail**
  - ▶ `thiago.covoes@ufabc.edu.br`
  - ▶ Prefixo no assunto: `[POSMD2020.QS]`
- ▶ **Página do curso**
- ▶ **Servidor Discord**

# Bibliografia

1. Tan, M. Steinbach e V. Kumar, Introduction to Data Mining.
2. Hand, H. Manilla e P. Smith. Principles of Data Mining.
3. Alpaydin, Introduction to Machine Learning.
4. Bishop. Pattern Recognition and Machine Learning.
5. Duda, P. Hart e D. Stork. Pattern Classification.