

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	Estimação de Densidade - Escalonamento Multidimensional
3	Classificação I
4	Classificação II
5	Regressão e Avaliação de modelos
6	Reuniões de acompanhamento projetos
7	Agrupamento de dados I
8	Agrupamento de dados II
9	Comitês
10	Seleção de atributos
11	Apresentação dos trabalhos
12	Mineração de Textos
13	Recuperação

Carga Horária

- ▶ É estimado um tempo de dedicação de aproximadamente 8h (TPI: 3-1-4) para cada semana:
 - ▶ Aproximadamente 3h30m para acompanhar os vídeos;
 - ▶ Aproximadamente 2h30m para realização de listas de exercícios;
 - ▶ Aproximadamente 2h para desenvolvimento do projeto;

Ferramentas utilizadas

- ▶ As aulas serão disponibilizadas em formato de vídeo no Youtube, com os links indicados no **Moodle**
- ▶ 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

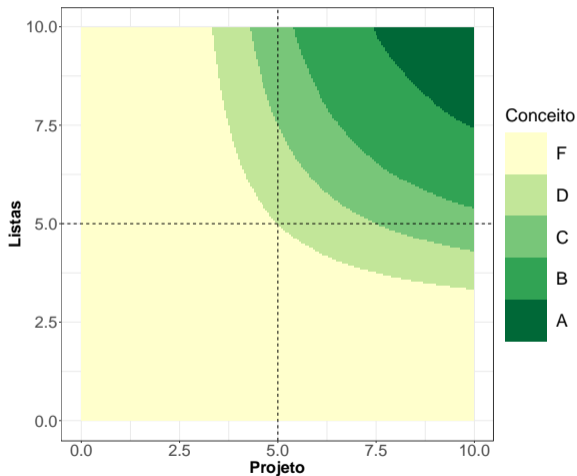
- ▶ Para facilitar aos alunos que tiverem problemas de disponibilidade de horário, as seguintes opções serão dadas neste tipo de plantão:
 - ▶ Segunda à quinta-feira: 18h-19h;

Avaliação

- ▶ Código de Honra
- ▶ Lista de Exercícios para entrega via TIDIA
- ▶ Projeto

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

Média harmônica



Conceitos

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

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

Senão Se $N_{\text{final}} \geq 6 \rightarrow C$

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

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

Recuperação

- ▶ Recuperação (14-19/12) para quem ficar com D (para tentar C) ou F (para tentar D)
- ▶ A atividade de recuperação será baseada nas atividades em que a nota obtida não foi suficiente
- ▶ Ao término do período de aulas (até dia 12/12) será indicada uma lista de todas as atividades que devem ser entregues para a recuperação da nota
 - ▶ O **projeto** pode ser uma dessas atividades, nesse caso o(s) aluno(s) deverão fazer ajustes no mesmo dentro desse período

Projeto

- ▶ Grupos de até 3 pessoas
- ▶ Simulação de consultoria
- ▶ Os integrantes do projeto serão uma equipe e terão que realizar revisão 360º
 - ▶ Todos avaliam todos
- ▶ Relatório - máximo 8 páginas
- ▶ Código para execução dos experimentos computacionais
 - ▶ Qualquer linguagem, desde que todas as dependências sejam gratuitas (isso inclui rodar sem problemas no Linux)

Listas

- ▶ Listas envolvendo exercícios teóricos e de implementação
- ▶ Exercícios teóricos deverão ser entregues manuscritos
 - ▶ Foto legível da folha
- ▶ 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 quando a carteirinha da UFABC para identificação
 - ▶ Só serão avaliadas as listas em que foi possível obter o vídeo
 - ▶ As explicações não devem levar mais do que 5 minutos

Dúvidas

- ▶ Contato via **e-mail**
 - ▶ thiago.covoes@ufabc.edu.br
 - ▶ Prefixo no assunto: [MD2020.QS]
- ▶ **Moodle**:
 - ▶ MD - DAMCZA015-13SA - 2020.QS
- ▶ **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.