

Prof. Carlos Andrés Ferrero

Mineração de Dados

Introdução à Disciplina e Plano de Ensino

- **Nome do professor:** Carlos Andres Ferrero
- **Avaliações:**
 - Duas provas, um trabalho
 - Atividades em sala de aula e estudos dirigidos.
- **Como serão as aulas:**
 - Aulas práticas em laboratório
 - No início de cada aula conversamos sobre conteúdo anterior e introduzimos novos conteúdos. E no restante da aula praticamos exercícios.
- A disciplina exige:
 - Dedicação, compromisso e atenção em sala de aula
 - Procurar o professor quando estiverem com dúvidas:
 - No horário de atendimento
 - **Email:** andres.ferrero@ifsc.edu.br
 - Chegar no horário correto da aula. Se tiver qualquer situação específica, entrar em contato com o professor.

Conteúdo Programático

1. Introdução à Mineração de Dados
2. Pré-processamento de dados
3. Paradigma simbólico
4. Paradigma baseado em exemplos
5. Aprendizado não supervisionado
6. Paradigmas probabilístico e conexionista

1. Introdução a Mineração de Dados

- Processo de Descoberta de Conhecimento
- Aprendizado Supervisionado e Não Supervisionado
- Paradigmas de Aprendizado
- Ferramentas: R, python e Jupyter
- Visualização de dados

2. Pré-processamento

- Normalização e discretização
- Redução de dimensionalidade
- Medidas de qualidade e seleção de atributos.

3. Paradigma Simbólico

- Algoritmos de árvores de decisão e regras de classificação
- Técnicas de avaliação de modelos
- Análise univariada e multivariada

4. Paradigma Baseado em Exemplos

- Algoritmo dos vizinhos mais próximos
- Funções de predição
- Funções de similaridade/distância
- Funções de similaridade/distância para dados não convencionais: sequências, texto e séries temporais.

5. Aprendizado não-supervisionado

- Algoritmos de agrupamento
- Regras de associação com *Apriori*
- Regras de associação sequenciais com *AprioriAll*

6. Paradigmas Probabilístico e Conexionista

- Aprendizado Bayesiano
- Redes Neurais Artificiais
- Máquinas Vetores de Suporte

Avaliação

Serão realizadas as seguintes avaliações:

- $P1$ - Prova escrita abordando Unidades 1, 2 e 3
- $P2$ - Prova escrita abordando Unidades 4, 5 e 6
- $T1$ - Trabalho abordando as Unidades 1 a 7
- E - Exercícios em sala de aula

O valor das notas $P1$, $P2$, $T1$ e E serão em valores inteiros. Em caso de notas quebradas (ex. 8,3 ou 8,6) usa-se o arredondamento matemático (ex. 8,3 para 8 e 8,6 para 9).

A Nota Final (NF) será calculada de acordo com a seguinte fórmula:

$$NF = \frac{(P1 + P2 + T1 + E)}{4}$$

O valor da nota NF também será arredondado para um número inteiro.

Recuperação

São previstas as recuperações:

- *RP1* - Recuperação prova 1 das Unidades 1, 2 e 3
- *RP2* - Recuperação prova 2 das Unidades 4, 5 e 6
- *RT1* - Recuperação trabalho 1 das Unidades 1 a 6

Estas recuperações serão feitas dentro da carga horária da disciplina.

Bibliografia

Básica

- SILVA, L. A.; PERES, S. M; BOSCARIOLI, C. **Introdução à mineração de dados: com aplicações em R** Rio de Janeiro: Elsevier, 2016.
- CARVALHO, Luís Alfredo Vidal de. **Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração**. Rio de Janeiro: Ciência Moderna, 2005.
- GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data mining: conceitos, técnicas, algoritmos, orientações e - aplicações**. 2. ed. Rio de Janeiro: Elsevier, 2015.

Complementar

- ATTWELL, P. A.; MONAGHAN, D. B. **Data mining for the social sciences: an introduction**. Oakland: University of California Press, 2015.
- AZZALINI, A.; WALTON, G.; SCARPA, B. **Data analysis and data mining: an introduction**. Oxford: Oxford University Press, 2012.
- LAWRENCE, K. D.; KLIMBERG, R. K. **Contemporary perspectives in data mining**. Charlotte: Information Age Publishing, 2012.
- MA, X.; CAPRI, H. L. **Data mining: principles, applications and emerging challenges**. New York : Nova Science Publishers, 2014.
- O'BYRNE, S. H. **Data Security, Data mining, and data management: technologies and challenges**. New York: Nova Science Publishers, 2013.