



## PLANO DE ENSINO

### 1. IDENTIFICAÇÃO

Componente Curricular:	MINERAÇÃO DE DADOS						
Unidade Ofertante:	FACOM						
Código:	FACOM39803	Período/Série:	8º		Turma:		
Carga Horária:				Natureza:			
Teórica:	36 h/a	Prática:	36 h/a	Total:	72 h/a	Obrigatória( ):	Optativa( )
Professor(A):	LEANDRO NOGUEIRA COUTO				Ano/Semestre:	2023/2	
Observações:							

### 2. EMENTA

Introdução e Motivação ao Processo de Descoberta de Conhecimento em Bases de Dados (KDD). Etapas do Processo de KDD. Técnicas de Pre-processamento dos Dados. Tarefas, Algoritmos e Paradigmas de Mineração de Dados: Associações, Classificação, Agrupamentos, Detecção de Outliers. Pós-processamento dos Resultados: Análise, Interpretação e Visualização. Ferramentas de Mineração de Dados.

### 3. JUSTIFICATIVA

A disciplina é essencial para o Bacharel em Gestão da Informação, apresentando as principais tarefas e métodos de mineração de dados, com foco na aplicação prática dos conceitos aprendidos em conjuntos de dados reais.

### 4. OBJETIVO

#### **Objetivo Geral**

O objetivo da disciplina é capacitar o discente a analisar, manipular, reconhecer padrões e extrair conhecimento computacionalmente de conjuntos de dados.

#### **Objetivos Específicos**

- Introduzir o aluno às principais tarefas e técnicas de Mineração de Dados.
- Habilitar o aluno a aplicar ferramentas de Mineração de Dados em problemas práticos.
- Habilitar o aluno a implementar suas próprias ferramentas de Mineração de Dados.

### 5. PROGRAMA

#### **1. Introdução**

1.1. O que é Mineração de Dados - o que é Descoberta de Conhecimento (KDD).

1.2. As fases do processo de KDD : pré-processamento dos dados. (limpeza, transformação, seleção de

atributos), mineração dos dados, pós-processamento dos resultados (análise, interpretação, visualização).

1.3. Principais Tarefas de Mineração de Dados.

## **2. Preparação dos Dados**

2.1. Sumarização dos dados: medindo a tendência central e a dispersão dos dados.

2.2. Limpeza dos dados: valores ausentes, tratamento de ruídos.

2.3. Integração e Transformação dos dados.

2.4. Redução dos dados: seleção de atributos, redução de dimensionalidade.

2.5. Discretização, Normalização.

## **3. Associação**

3.1. Mineração de Regras de Associação – Algoritmo Apriori e variantes.

3.2. Mineração de Sequências – Algoritmos GSP e Prefix-Span.

## **4. Classificação**

4.1. O que é um classificador?

4.2. Árvore de Decisão.

4.3. Classificadores baseados nos vizinhos mais próximos (KNN).

4.4. Classificadores baseados em Redes Bayesianas de Crença.

4.5. Avaliação de Performance: Método Holdout, Cross-Validation, Bootstrap.

## **5. Agrupamentos (Clusters)**

5.1. Diferentes tipos de clusters : bem separados, baseados em centróides, baseados em grafos, baseados em densidade.

5.2. Diferentes tipos de clusterização : por particionamento, hierárquico, exclusivos, com justaposição, completas, parciais.

5.3. Método K-Means e K-Medóides – análise de performance, complexidade.

5.4. Método hierárquico aglomerativo – análise de performance, complexidade.

## **6. Análise de Clusters**

6.1. Medidas: coesão, separação, SSE, coeficiente de silhueta.

6.2. Técnicas para determinar o número correto de clusters.

6.3. Técnicas para determinar a tendência de clusters nos dados.

## **7. Detecção de Anomalias (Outliers)**

7.1. Introdução: causas de anomalias.

7.2. Técnicas para detecção de anomalias: estatísticas e baseadas em proximidade.

## **8. Pós-Processamento: Análise, Interpretação e Visualização**

### **6. METODOLOGIA**

O conteúdo será desenvolvido presencialmente em sala de aula, com aulas expositivas, material didático apresentado em projetor multimídia e lousa. A carga horária será complementada na forma de atividades de Trabalho Docente Efetivo (TDE). Além disso, haverá práticas de laboratório baseadas no conteúdo teórico, e atividades opcionais de aprofundamento.

#### **1. Aulas presenciais (64 horas aula)**

Aulas presenciais, divididas igualmente entre aulas teóricas (terça-feira de 19:00 a

20:40) e práticas de laboratório (terça-feira de 20:50 a 22:40).

## **2. Atividades TDE (8 horas-aula)**

Listas de exercícios disponibilizadas por Google Drive e plataforma Microsoft Teams, para estudo e aprofundamento no conteúdo. O código de acesso à plataforma Teams será repassado em aula e também enviado por e-mail na primeira semana de aulas do semestre.

## **3. Atendimento ao aluno**

O atendimento aos alunos ocorrerá de forma presencial na quinta-feira das 14:00 às 16:00, na sala do docente, no bloco 1B, ou em outro horário mediante agendamento. O atendimento ao aluno poderá ser feito também de forma assíncrona através de contato por e-mail, ou por mensagem no Microsoft Teams.

## **7. AVALIAÇÃO**

### **1. Avaliação**

Haverá duas provas teóricas:

1ª prova (20/02/2024): Valor total 25 pontos.

2ª prova (16/04/2024): Valor total 25 pontos.

O resto da nota será referente a atividades práticas, no valor total 50 de pontos, avaliadas mediante entrega no tempo estabelecido e correteude.

A soma dos valores das atividades de avaliação totaliza 100 pontos, sendo que a nota final do aluno será composta pelo somatório das notas das duas provas e das atividades práticas.

**Avaliação de recuperação de aprendizagem:** Será realizada uma avaliação de recuperação para os alunos que não tenham atingido 60 pontos na disciplina e tenham obtido a frequência mínima de 75%. A recuperação terá forma de uma prova que engloba todo o conteúdo do semestre e substitui a menor nota das 2 provas teóricas anteriores.

### **2. Validação da assiduidade**

A assiduidade será computada com base na presença do aluno nas aulas presenciais. O aluno deverá ter no mínimo 75% de frequência para aprovação.

## **8. BIBLIOGRAFIA**

### **Básica**

1. TAN, Pang-Ning. **Introdução ao Data Mining: mineração de dados**. Rio de Janeiro: Ciência Moderna, 2009.
2. J. Han, M. Kamber: **Data Mining: Concepts and Techniques**, 2a. Ed., Morgan Kaufmann, 2006.
3. I. H. Witten, E. Frank: **Data Mining: Practical Machine Learning Tools and Techniques**, Morgan Kaufmann, 2005.

### **Complementar**

1. CHAKRABARTI, Soumen. **Mining the Web: discovering knowledge from hypertext data**. San Francisco: Morgan Kaufman, c2003.
2. KANTARDZIC, Mehmed. **Data mining: concepts, models, methods, and algorithms**. 2nd ed. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, c2011. E-Book.
3. LAROSE, Daniel T. **Data mining methods and models**. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, c2006. E-Book.
4. JENSEN, Finn V. **Bayesian networks and decision graphs**. 2nd ed. New York:

Springer, c2007.

5. CHERKASSKY, Vladimir S. **Learning from data: concepts, theory, and methods**. 2nd ed. Hoboken; Piscataway: John Wiley & Sons: IEEE Press, c2007. E-Book.

## 9. **APROVAÇÃO**

Aprovado em reunião do Colegiado realizada em: \_\_\_/\_\_\_/\_\_\_

Coordenação do Curso de Graduação: \_\_\_\_\_



Documento assinado eletronicamente por **Leandro Nogueira Couto**, **Professor(a) do Magistério Superior**, em 29/01/2024, às 18:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://www.sei.ufu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **5142848** e o código CRC **5B18E8F0**.

**Referência:** Processo nº 23117.003020/2024-14

SEI nº 5142848