



UNIVERSIDADE FEDERAL DE UBERLÂNDIA



## FICHA DE COMPONENTE CURRICULAR

<b>CÓDIGO:</b> GSI055	<b>COMPONENTE CURRICULAR:</b> MINERAÇÃO DE DADOS	
<b>UNIDADE ACADÊMICA OFERTANTE:</b> FACULDADE DE COMPUTAÇÃO		<b>SIGLA:</b> FACOM
<b>CH TOTAL TEÓRICA:</b> 60 horas	<b>CH TOTAL PRÁTICA:</b> 0 horas	<b>CH TOTAL:</b> 60 horas

1. **OBJETIVOS**

Os objetivos da disciplina são:

- 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.

2. **EMENTA**

Introdução e motivação ao processo de descoberta de conhecimento em bases de dados (KDD). Etapas do processo de KDD. Técnicas de pré-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. **PROGRAMA**

## 1. Introdução

- O que é Mineração de Dados - o que é Descoberta de Conhecimento (KDD).
- 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).
- Principais tarefas de mineração de dados.

## 2. Preparação dos Dados

- Sumarização dos dados: medindo a tendência central e a dispersão dos dados.
- Limpeza dos dados: valores ausentes, tratamento de ruídos.
- Integração e transformação dos dados.

- Redução dos dados: seleção de atributos, redução de dimensionalidade.
- Discretização e normalização.

### 3. Associação

- Mineração de regras de associação – algoritmo *Apriori* e variantes.
- Mineração de Sequências – algoritmos GSP e *Prefix-Span*.

### 4. Classificação

- O que é um classificador.
- Árvore de Decisão
- Classificadores baseados nos vizinhos mais próximos (KNN).
- Classificadores baseados em redes bayesianas de crença.
- Classificadores baseados em redes neurais.
- Avaliação de desempenho: métodos *Holdout*, *Cross-Validation* e *Bootstrap*.
- Técnicas de classificação para classes não-balanceadas: métricas alternativas, precisão, recall, curva ROC.

### 5. Agrupamentos (clusters)

- Diferentes tipos de clusters : bem separados, baseados em centróides, baseados em grafos, baseados em densidade.
- Diferentes tipos de agrupamento: por particionamento, hierárquico, exclusivos, *fuzzy*, com justaposição, completas, parciais.
- Método *K-Means* e *K-Medóides* - análise de desempenho e complexidade.
- Método hierárquico aglomerativo - análise de desempenho e complexidade.
- Método DBSCAN - análise de desempenho e complexidade.

### 6. Análise de Clusters

- Medidas: coesão, separação, SSE, coeficiente de silhueta.
- Técnicas para determinar o número correto de clusters.
- Técnicas para determinar a tendência de clusters nos dados.

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

- Introdução: causas de anomalias.
- Técnicas para detecção de anomalias: estatísticas e baseadas em proximidade.

8. Pós-Processamento: análise, interpretação e visualização.
9. Aplicações: utilização de ferramentas de mineração de dados.

#### 4. BIBLIOGRAFIA BÁSICA

- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. **Inteligência artificial**: uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011. 396 p.
- TAN, P. **Introdução ao Data mining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.
- ZAKI, M. J.; MEIRA, W. **Data Mining and Analysis: fundamental concepts and algorithms**. Massachusetts: Cambridge University Press, 2014. 562 p.

#### 5. BIBLIOGRAFIA COMPLEMENTAR

- AGGARWAL C. C.; REDDY C. K. **Data Clustering**: Algorithms and Applications. [S. l.]: CRC Press, 2014. v. 1.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2007. v.1. 738p.
- CHAKRABARTI, S. **Mining the Web**: Analysis of Hypertext and Semi Structured Data, Morgan Kaufmann, 2002.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**, 2a. ed., Morgan Kaufmann, 2006.
- TAN, P.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**, Addison-Wesley, 2005.
- THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 4th ed. Amsterdam: Academic, 2009.
- WITTEN, I. H.; FRANK, E. **Data Mining**: Practical Machine Learning Tools and Techniques, 2nd ed. Boston: Morgan Kaufmann, 2005.

#### 6. APROVAÇÃO

Prof. Dr. Jefferson Rodrigo de Souza  
Coordenador do Curso de Sistemas de Informação

Prof. Dr. Mauricio Cunha Escarpinati  
Diretor da Faculdade de Computação



Documento assinado eletronicamente por **Jefferson Rodrigo de Souza, Presidente**, em 21/12/2021, às 13:35, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Mauricio Cunha Escarpinati, Diretor(a)**, em 01/02/2022, às 14:47, 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 **3123512** e o código CRC **57B98FB5**.

---

**Referência:** Processo nº 23117.019924/2019-96

SEI nº 3123512