



Componente Curricular:	ESTAT0109 – MINERAÇÃO DE DADOS EM ESTATÍSTICA		
Carga horária:	60 horas (4 créditos)	Horário:	Terças – 17h00 às 18h30 Quintas – 17h00 às 18h30
Unidade Responsável:	Departamento de Estatística e Ciências Atuariais		
Docente:	Sadraque E. F. Lucena		
Tipo de Componente:	Disciplina	Quantidade de Avaliações:	3
Período Letivo:	2025-2		

Ementa: Análise estatística em grandes bancos de dados. Tratamento de dados para processos de Data Mining. Principais funcionalidades, técnicas e algoritmos. Análise de associações. Classificação de dados. Árvores de decisão. Regressão Logística. Redes Neurais. Segmentação e Análise de Cluster. Estudo de casos.

Objetivos: Capacitar os(as) alunos(as) a aplicar técnicas estatísticas e computacionais para extrair padrões e conhecimentos úteis de conjuntos de dados.

Metodologia: Serão ministradas aulas teóricas expositivas; utilizados recursos visuais; resolução de problemas práticos; solicitação de atividades extraclasse.

Habilidades e Competências: Ao final do curso, o estudante será capaz de preparar, modelar e avaliar conjuntos de dados complexos, utilizando os principais algoritmos de mineração de dados para resolver problemas práticos. O cronograma de aulas encontra-se no Quadro 1.

Avaliação: A avaliação do aprendizado será realizada por meio de um portfólio de projetos práticos, composto por:

- Mini-Projetos: Ao longo do semestre, serão propostos estudos de caso práticos para a aplicação das técnicas de mineração de dados vistas em aula. Cada entrega consistirá em um relatório técnico sucinto, documentando a preparação dos dados, a implementação do algoritmo e a interpretação dos resultados.
- Projeto Final Integrador: Ao final da disciplina, os estudantes deverão compilar suas análises em um relatório final robusto. Este documento deverá realizar uma análise comparativa crítica entre os modelos desenvolvidos, avaliando suas métricas de desempenho e discutindo a adequação de cada técnica para o problema analisado, culminando na recomendação do modelo de melhor performance.

Hora-trabalho: A carga horária extraclasse da disciplina será cumprida por meio do desenvolvimento contínuo dos mini-projetos e do projeto final. Recomenda-se que os estudantes mantenham um ritmo de trabalho semanal para aprofundar os estudos teóricos e avançar nas análises práticas exigidas nos relatórios.

Conteúdo:

1. Fundamentos da Mineração de Dados
2. Pré-processamento de Dados
 - 2.1. Exploração
 - 2.2. Limpeza
 - 2.3. Transformação
 - 2.4. Redução
3. Aprendizado Não Supervisionado
 - 3.1. Regras de Associação
 - 3.2. k -means
4. Aprendizado Supervisionado
 - 4.1. Regressão
 - 4.1.1. Regressão linear
 - 4.1.2. Regressão logística
 - 4.2. k -Nearest Neighbors
 - 4.3. Naïve Bayes
 - 4.4. Árvores de Decisão
 - 4.5. Florestas aleatórias
 - 4.6. Avaliação de desempenho
 - 4.6.1. Validação Cruzada
 - 4.6.2. Amostragem bootstrap
 - 4.6.3. Acurácia
 - 4.6.4. Kappa
 - 4.6.5. Precisão e revocação
 - 4.6.6. Sensibilidade e especificidade
 - 4.7. Ajuste de parâmetros
 - 4.8. Métodos de conjunto (ensemble methods)
 - 4.8.1. Bagging
 - 4.8.2. Boosting
 - 4.8.3. Stacking

Bibliografia:

Básica:

- PANG-NING, Tan; STEINBACH, Michael, KARPATNE, Anuj; KUMAR, Vipin. Introduction to data mining. 2. ed. New York: Pearson, 2019.
- BRAMER, Max. Practical Machine Learning in R. Indianapolis, Indiana: John Wiley & Sons, 2020.
- AGGARWAL, Charu C. et al. Data mining: the textbook. New York: springer, 2015.
- HASTIE, T.; TIBSHIRANI, R; FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. 2. ed. New York, NY: Springer, 2009.
- DUNHAM, Margaret H. Data Mining: Introductory and Advanced Topics. Pearson, 2020.

Complementar:

- CASTRO, Leandro N.; FERRARI, Daniel G. Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações. São Paulo : Saraiva, 2016.
- BRAMER, Max. Principles of Data Mining. 4. ed. Springer, 2020.
- LAROSE, Daniel T.; LAROSE, Chantal D. . Discovering knowledge in data: an introduction to data mining. 2. ed. John Wiley & Sons, 2014.

Quadro 1: Cronograma de aulas de ESTAT0109 – Mineração de Dados em Estatística para 2025-2.

Data	Dia da Semana	Aula	Assunto Previsto
07/10/25	Terça	1	VI Encontro de Estatística e Ciências Atuariais da UFS
09/10/25	Quinta	2	Apresentação da disciplina. Introdução à mineração de dados.
14/10/25	Terça	3	Introdução ao Quarto.
16/10/25	Quinta	4	Pré-processamento de dados.
21/10/25	Terça	5	Pré-processamento de dados.
23/10/25	Quinta	6	Algoritmos de agrupamento.
28/10/25	Terça	7	Algoritmos de agrupamento.
30/10/25	Quinta	8	Algoritmos de agrupamento.
04/11/25	Terça	9	Introdução à Classificação de dados. Validação cruzada e avaliação de desempenho.
06/11/25	Quinta	10	Regressão Logística.
11/11/25	Terça	11	Regressão Logística.
13/11/25	Quinta	12	k-Nearest Neighbors (k-NN).
18/11/25	Terça	13	k-Nearest Neighbors (k-NN).
20/11/25	Quinta	–	Dia Nacional de Zumbi e da Consciência Negra (feriado nacional)
25/11/25	Terça	14	XI SEMAC
27/11/25	Quinta	15	XI SEMAC
02/12/25	Terça	16	Naive Bayes.
04/12/25	Quinta	17	Naive Bayes.
09/12/25	Terça	18	Árvores de decisão.
11/12/25	Quinta	19	Árvores de decisão.
16/12/25	Terça	20	Floresta Aleatória (Random Forest).
18/12/25	Quinta	21	Floresta Aleatória (Random Forest).
23/12/25	Terça	–	Recesso acadêmico
25/12/25	Quinta	–	Recesso acadêmico
30/12/25	Terça	–	Recesso acadêmico
01/01/26	Quinta	–	Confraternização Universal (feriado nacional)
06/01/26	Terça	–	Férias coletivas para docentes
08/01/26	Quinta	–	Férias coletivas para docentes
13/01/26	Terça	22	Support Vector Machine
15/01/26	Quinta	23	Support Vector Machine
20/01/26	Terça	24	Regressão Linear
22/01/26	Quinta	25	Regressão Linear
27/01/26	Terça	26	Regras de Associação
29/01/26	Quinta	27	Regras de Associação
03/02/26	Terça	28	Projeto Final
05/02/26	Quinta	29	Projeto Final
10/02/26	Terça	30	Apresentação do Projeto Final