



DADOS DE IDENTIFICAÇÃO E ATRIBUTOS¹

CÓDIGO							NOME						DEPARTAMENTO OU EQUIVALENTE					
MATF52							Introdução à Mineração de Dados						Departamento de Estatística					
CARGA HORÁRIA (estudante)							MODALIDADE/ SUBMODALIDADE						PRÉ-REQUISITO (POR CURSO)					
T	T/P	P	PP	Ext	E	TOTAL	Disciplina/Teórico-prática						MAT229 – Análise de Regressão					
34	0	34	0	0	0	68												
CARGA HORÁRIA (docente/turma)							MÓDULO ²						SEMESTRE DE INÍCIO DA VIGÊNCIA					
T	T/P	P	PP	Ext	E	TOTAL	T	T/P	P	PP	Ext	E	08/09/2020 a 18/12/2020					
34	0	34	0	0	0	68	30	0	30	0	0	0						

EMENTA

Introdução à mineração de dados. Análise estatística de grandes bancos de dados. O processo de descoberta de conhecimento em bancos de dados. Principais funcionalidades e técnicas de mineração de dados. Classificação de dados. Métodos de classificação supervisionada. Combinação de classificadores. Segmentação. Estudos de caso.

OBJETIVOS

OBJETIVO GERAL

Conhecer de forma prática e teórica os tópicos do programa.

OBJETIVOS ESPECÍFICOS

Aplicar métodos de análise estatística e de mineração de dados², bem como conhecer e utilizar as ferramentas necessárias.

CONTEÚDO PROGRAMÁTICO

- Introdução à mineração de dados.
 - O que é mineração de dados?
 - A interdisciplinaridade da técnica.
 - Problemas típicos e potenciais aplicações de mineração de dados.
- Mineração de dados e análise estatística.
 - Entendendo as diferenças.
 - Questionando estratégias.
- Importância dos bancos de dados.
 - Características em um banco de dados.
 - As ideias de armazém de dados (Data Warehouse) e processamento analítico em tempo real (OLAP).
- Visão geral sobre as fases do processo de descoberta de conhecimento em bancos de dados (Knowledge Discovery in Databases).

¹ Os “dados de identificação e atributos” devem estar registrados conforme especificado no Programa do Componente Curricular e disponível no site da Superintendência Acadêmica (SUPAC). O único campo a ser preenchido nesse tópico do formulário é o que diz respeito ao módulo de vagas ofertadas.

² Conforme Resolução CONSUNI 01/2020 e CAE 01/2020, é possível flexibilizar o disposto na Resolução CONSEPE 02/2009.

5. Classificação supervisionada.

5.1. Métodos.

5.1.1. Árvores de decisão.

5.1.2. Regressão logística.

5.1.3. Naive Bayes.

5.1.4. K vizinhos mais próximos.

5.2. Avaliação dos modelos de classificação supervisionada.

5.2.1. Medidas de capacidade preditiva.

5.2.1.1. Sensibilidade.

5.2.1.2. Especificidade.

5.2.1.3. Valores de predição positivo e negativo.

5.2.1.4. Acurácia.

5.2.2. Curva ROC.

6. Combinação de classificadores.

6.1. Bagging de modelos.

6.2. Métodos de combinação.

6.2.1. Combinação via voto.

6.2.2. Combinação via regressão logística.

7. Procedimentos de segmentação de base.

7.1. Segmentação por benefícios.

7.2. Segmentação demográfica.

7.3. Segmentação por nível de utilização de um produto.

7.4. Segmentação por estilo de vida.

7.5. Segmentação com base no valor do cliente.

METODOLOGIA DE ENSINO-APRENDIZAGEM

1. Web conferências e seminários síncronos.
2. Material complementar a ser lido pelo aluno.
3. Vídeo aulas a serem disponibilizadas sobre os tópicos.
4. Fórum de discussão.
5. Pesquisa e trabalhos em grupos.
6. Resolução de exercícios.

AVALIAÇÃO DA APRENDIZAGEM

Avaliação contínua mediante a evolução dos tópicos do conteúdo programático. Em suma, uma avaliação ao fim de cada módulo (o aluno deverá desenvolver a cada 3 ou 4 semanas). Todas as avaliações serão compostas por duas partes (síncrona e assíncrona). A parte síncrona se baseará em questões conceituais (50%) e a parte assíncrona a trabalhos práticos e/ou de pesquisa (50%). A nota final da disciplina, será composta pela média aritmética das quatro avaliações.

REFERÊNCIAS

REFERÊNCIAS BÁSICAS

Impressos

JOHNSON, R. A.; WICHERN, D. W. Applied Multivariate Statistical Analysis. New Jersey: Prentice Hall, 4a edição, 1998, 816 p.

SANTOS, L. A. N. Contribuição da mineração de dados e da otimização heurística para a interpretação dos dados da produção científica brasileira. 2011. 114 f. Dissertação (Mestrado em Ciência da Informação) – Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador. 2011.

Ebooks

BRAMER, M. Principles of Data Mining. London: Springer-Verlag, 2a edição, 2013, 440 p.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. An Introduction to Statistical Learning: with Applications in R. New York: Springer-Verlag, 2013, 426 p.

KUHN, M.; JOHNSON, K. Applied Predictive Modeling. New York: Springer-Verlag, 2013, 600 p.

REFERÊNCIAS COMPLEMENTARES

Impressos

KLEINBAUM, D. G. Logistic regression: a self-learning text. Statistics in the health sciences. New York: Springer-Verlag, 1994, 282 p.

MICHIE, D.; SPIEGELHALTER, D.J.; TAYLOR, C.C. Machine Learning, Neural and Statistical Classification. New York: Ellis Horwood, *Ebooks*
CUZZOCREA, A.; DAYAL, U. (eds.). Data Warehousing and Knowledge Discovery. 14th International Conference, DaWaK 2012, Vienna, Austria, September 3-6, 2012, Proceedings. Berlin/Heidelberg: Springer-Verlag, 454 p.
DOUGHERTY, G. Pattern Recognition and Classification: An Introduction. New York: Springer-Verlag, 2013, 196 p.
GABER, M.M. Journeys to Data Mining: Experiences from 15 Renowned Researchers. Berlin/Heidelberg: Springer-Verlag, 2012, 244 p.
GIUSTI, A.; RITTER, G.; VICHI, M. (eds.). Classification and Data Mining. Studies in Classification, Data Analysis, and Knowledge Organization. Berlin/Heidelberg: Springer-Verlag, 2013, 286 p.

Docente(s) Responsável(is) à época da aprovação do Plano de ensino-aprendizagem:

Nome: _____ Assinatura: _____

Nome: _____ Assinatura: _____

Aprovado em reunião de Departamento (ou equivalente): _____ **em** ____/____/____ _____
Assinatura do Chefe

ANEXO

CRONOGRAMA³

Código e nome do componente:	MATF52 – Introdução a Mineração de Dados
Nome do/s docente/s:	Anderson Luiz Ara Souza
Período:	08/09/2020 a 18/12/2020

Data ou período de realização	Unidade Temática ou Conteúdo	Técnicas ou estratégias⁴ de ensino previstas	Atividade/ Recurso⁵	CH Docente⁶	CH Discente⁷
Semana 1 e Semana 2	1. Introdução à mineração de dados. 1.1. O que é mineração de dados? 1.2. A interdisciplinaridade da técnica. 1.3. Problemas típicos e potenciais aplicações de mineração de dados.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	5:00h	5:00h
Semana 2 e Semana 3	2. Mineração de dados e análise estatística. 2.1. Entendendo as diferenças. 2.2. Questionando estratégias.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	5:00h	5:00h

³ Esta é uma sugestão de cronograma. A sua adoção é facultativa, sendo possível, a critério do(s) professor(es), adotar outra forma de expressar aspectos temporais e de uso de dispositivos tecnológicos. Para o SLS, recomenda-se pensar a organização do componente em unidades ou temáticas amplas, considerando períodos equivalentes à carga horária de uma ou mais semanas

4 Possibilidades de técnicas e estratégias de ensino-aprendizagem:

Síncronas: Aula dialogada (ao vivo) pelos professores em interatividade com os estudantes; Apresentação de artigos ou temas pelos estudantes com mediação dos professores); Aula invertida (*chat* a partir de texto ou vídeo com mediação dos professores); Chats com pequenos grupos.

Assíncronas: Aula expositiva (preleções feitas pelos professores e gravadas como videoaulas); Aula invertida (fórum de discussão a partir de texto ou vídeo) com mediação dos professores; Discussão de tema (problematizado) com X postagem dos estudantes e mediação dos professores; Cocriação de textos colaborativos pelos estudantes com mediação dos professores; Desenvolvimento de atividades/tarefas pelos estudantes: resenha, confecção de vídeos, modelos, questionários, peças jurídicas, roteiros, guias de estudo, produções artísticas com mediação dos professores.

⁵ As palavras **Atividade** e **Recursos** aqui acompanham a classificação do Moodle. As atividades podem ser: Fórum, chat, wiki, tarefas, jogos, escolha, glossários, base de dados, pesquisa, questionário etc. Os recursos podem ser: arquivo, URL, livro, pasta, rótulo etc.

⁶ Indicar carga horária também de elaboração e realização.

⁷ Indicar o tempo previsto para que o estudante realize a atividade/tarefa.

Semana 4 a Semana 5	3. Importância dos bancos de dados. 3.1. Características em um banco de dados. 3.2. As ideias de armazém de dados (Data Warehouse) e processamento analítico em tempo real (OLAP).	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	5:00h	5:00h
Semana 6	Avaliação 1	Avaliação 1	Prova	4:00h	4:00h
Semanas 6 e 7	5. Classificação supervisionada. 5.1. Métodos. 5.1.1. Árvores de decisão.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	4:00h	4:00h
Semana 8	5.1.2. Regressão logística. .	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	4:00h	4:00h
Semana 9	5.1.3. Naive Bayes.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	4:00h	4:00h
Semana 10	5.1.4. K vizinhos mais próximos.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	4:00h	4:00h
Semanas 10 a 11	5.2. Avaliação dos modelos de classificação supervisionada. 5.2.1. Medidas de capacidade preditiva.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online)	AVA e Google for Education, Fórum, Textos e Artigos	5:00h	5:00h

	5.2.1.1. Sensibilidade. 5.2.1.2. Especificidade. 5.2.1.3. Valores de predição positivo e negativo. 5.2.1.4. Acurácia.	Atividade assíncrona ao decorrer da semana			
Semana 11	5.2.2. Curva ROC.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	4:00h	4:00h
Semana 12	6. Combinação de classificadores. 6.1. Bagging de modelos. 6.2. Métodos de combinação. 6.2.1. Combinação via voto. 6.2.2. Combinação via regressão logística.	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos e Artigos	4:00h	4:00h
Semana 13	Avaliação 2	Avaliação 2	Prova	5:00h	5:00h
Semanas 13 a 14	7. Procedimentos de segmentação de base. 7.1. Segmentação por benefícios. 7.2. Segmentação demográfica. 7.3. Segmentação por nível de utilização de um produto. 7.4. Segmentação por estilo de vida. 7.5. Segmentação com base no valor do cliente	Terça das: 14h50 às 16h00 Atividade síncrona (Explicações e dúvidas online) Atividade assíncrona ao decorrer da semana	AVA e Google for Education, Fórum, Textos, Artigos, Vídeos e Plataformas hands-on	5:00h	5:00h
Semana 14	Avaliação 3	Avaliação 3	Prova	5:00h	5:00h
Semana 15	Avaliação Final	Avaliação	Atividade de Pesquisa e seminário	5:00h	5:00h