



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS SOCIAIS APLICADAS
Departamento de Ciência e Gestão da Informação

Ficha 2 (variável)

Disciplina: Mineração de Dados		Código: SIN195
Professor responsável: José Marcelo A. P. Cestari (cestari@ufpr.br)		Período da oferta: 17/10/2022 a 25/02/2023.
Natureza: <input checked="" type="checkbox"/> Obrigatória <input type="checkbox"/> Optativa	<input checked="" type="checkbox"/> Semestral <input type="checkbox"/> Anual <input type="checkbox"/> Modular	Vagas: 50.
Pré-requisito: não possui.	Co-requisito: não possui.	
Modalidade: <input type="checkbox"/> Presencial <input type="checkbox"/> EaD <input checked="" type="checkbox"/> Parcialmente EaD: 20% (12h)		
C.H.Total: 60 horas. C. H. Semanal: 4 horas. Prática como Componente Curricular (PCC): 0. Atividade Curricular de Extensão (ACE): 0. Padrão (PD): 30 horas. Laboratório (LB): 30 horas.. Campo (CP): 0. Estágio (ES): 0. Orientada (OR): 0. Prática Específica (PE): 0. Estágio de Formação Pedagógica (EFP): 0.		
6º Período. Recesso entre 26/12/2022 e 14/01/2023 (inclusive). Período do exame final: de 27/02 até 04/03/2023 Exame final da disciplina: 28/02/2023. Feriados e outras datas previstas (conforme PORTARIA ME Nº 14.817, DE 20 DE DEZEMBRO DE 2021): 12/10, 28/10 (ponto facultativo), 02/11, 15/11 e 25/12. Carnaval 2023 previsto para 20, 21 e 22/02/2023 (quarta-feira). Justificativa da oferta 20% EaD: Portaria 280/ PROGRAD, de 13/12/2017 que "Aprova Ajuste Curricular no Curso de Gestão da Informação, aplicável à Resolução 64/13-CEPE" (0690698).		
EMENTA Aspectos teóricos e práticos da descoberta de conhecimento em bancos de dados utilizando tarefas e técnicas de mineração de dados.		
PROGRAMA • Introdução (~16 horas): <ul style="list-style-type: none">◦ Posicionamento.		

- Motivação.
- Conceitos básicos (IA, DW).
- KDD e MD (~3 horas):
 - Etapas do processo de KDD.
 - Metodologia de mineração de dados.
 - Técnicas e tarefas de mineração de dados.
- Tarefas (~31 horas):
 - Classificação (~9 horas + ~4 horas (avaliação)).
 - Associação (~6 horas).
 - Agrupamento (~6 horas + ~4 horas (avaliação)).
- Pós-processamento (~2 horas)
 - Formatos de apresentação de resultados.
 - Avaliação de resultados.
- Trabalho final (~8 horas)
 - Aplicativo Weka.
 - Excel.

OBJETIVO GERAL

O aluno deverá conhecer e ser capaz de aplicar diferentes técnicas estatísticas e computacionais de exploração de dados na identificação de padrões relevantes e não óbvios em bancos de dados.

OBJETIVOS ESPECÍFICOS

O aluno deverá ser capaz de:

1. Discutir os conceitos básicos relacionados à mineração de dados;
2. Aplicar e explicar resultados das principais tarefas de mineração de dados;
3. Justificar a relevância desta disciplina;
4. Identificar e compreender aplicações práticas da mineração de dados.

PROCEDIMENTOS DIDÁTICOS

Aulas, teórico práticas, expositivo dialogadas em encontros presenciais nos quais a docente faz exposição do tema do programa, discute com os alunos os métodos e aplicações e propõe exercícios e atividades em grupo ou individuais, além das atividades avaliativas.

Serão utilizados como recursos: sala de aula no SA, laboratório de computadores, sala no ambiente virtual de aprendizagem – AVA, mais especificamente o UFPR Virtual; o Microsoft Office 365 disponível institucionalmente, o site da Data Science Academy como suporte ao conteúdo e o software Weka. Para as aulas não presenciais o material didático será elaborado observando utilização de linguagem dialógica, apresentação e formato próprios para o ensino à distância. Como suporte ao ensino será utilizada a infraestrutura da UFPR tais como: software (Office 365) e AVA. Em atendimento à Resolução 72/10-CEPE, serão observados os seguintes aspectos:

1. Sistema de comunicação – encontros presenciais, AVA (UFPR Virtual), correios eletrônicos trocados entre coordenação e alunos e professor e alunos;
2. Modelo de tutoria a distância e presencial – não está prevista a tutoria nesta disciplina. As dúvidas serão esclarecidas pela docente e, quando existir, pelos alunos do programa de Monitoria ou alunos de PPGs em estágios docência (quando for o caso);
3. Material didático específico – o material didático será elaborado observando utilização de linguagem dialógica, apresentação e formato próprios para o ensino presencial;
4. Infraestrutura de suporte tecnológico, científico e instrumental à disciplina – o curso conta com Ambiente Virtual de Aprendizagem específico e pacote Office 365, incluindo sala no Teams para eventuais esclarecimentos e interação;
5. Previsão de período de ambientação dos recursos tecnológicos a serem utilizados pelos discentes: os alunos recebem capacitação no Ambiente de Aprendizagem Virtual no primeiro semestre do Curso e esta é uma disciplina de sexto período. Assume-se que os alunos já tenham domínio no uso das tecnologias.
6. As atividades podem ser presenciais ou remotas, conforme definição em sala de aula e rubricas.

IMPORTANTE:

A) WEKA. Recomenda-se que os discentes instalem a ferramenta Weka em suas máquinas pessoais, considerando claro aqueles que tenham possibilidade, disponibilidade e condições. Favor instalar antes do início da disciplina. O link para baixar o Weka é https://waikato.github.io/weka-wiki/downloading_weka/ (visitado em 17/08/2022). Será solicitada a instalação do Weka em algum laboratório para que possamos usá-la.

B) Cursos DSA. Realizar os cursos gratuitos "Inteligência Artificial Fundamentos 2.0" e "Introdução à Ciência de Dados 3.0" da DSA no seguinte link: <https://www.datascienceacademy.com.br/cursosgratuitos> (visitado em 17/08/2022).

C) Minha Biblioteca (Livros on-line - UFPR). Link aqui: <https://minhabiblioteca.ufpr.br/biblioteca/> (visitado em 17/08/2022).

ATIVIDADES NÃO PRESENCIAIS

Sistema de Comunicação: Plataforma Moodle da UFPR, com acesso em www.ufprvirtual.ufpr.br.

Tutoria/Monitoria: As atividades no Moodle terão a supervisão do professor da disciplina, com apoio de monitor do Programa Institucional de Monitoria (PIM) quando este for o caso. A tutoria/monitoria prevê a orientação no cumprimento das tarefas estabelecidas para cada estratégia, verificação de prazos, resposta de dúvidas e identificação de necessidades e problemas na condução da disciplina.

Material didático: No Moodle está disponível o Guia Didático da disciplina, onde consta o cronograma e demais orientações gerais das práticas e atividades. Além deste documento, na plataforma também estão disponíveis demais materiais pertinentes à disciplina, incluindo mas não limitados

à arquivos para leitura (e-books ou arquivos do tipo pdf ou links para acesso a artigos em periódicos); e vídeos (link para visualização de material audiovisual em plataformas como Youtube).

Suporte técnico: O Setor de Ciências Sociais Aplicadas dispõe de laboratório de Informática para os estudantes que necessitarem de acesso a computador e à Internet. O mesmo fica no 1º andar do edifício do Setor, no campus Jardim Botânico. Suporte técnico também é oferecido em relação ao UFPR Virtual pela Coordenação de Integração de Políticas de Educação a Distância (CIPEAD/PROGRAD). Eventuais problemas no Moodle deverão ser comunicados sempre que verificados ao monitor da disciplina ou, então, ao administrador da plataforma, em <https://ufprvirtual.ufpr.br/course/view.php?id=8860>.

Ambientação: o curso de Gestão da Informação prevê a ambientação de seus estudantes no AVA durante o 1º ano, 1o período, a partir da oferta de duas disciplinas 100% EaD, a partir das quais todos os alunos são orientados e experimentam o Moodle e suas funcionalidades. Caso persistam dúvidas sobre a utilização da plataforma, recomenda-se recorrer aos tutoriais disponíveis em <https://ufprvirtual.ufpr.br/course/view.php?id=8737#section-0>.

Frequência: Nos dias previstos para atividades de EaD a frequência é computada a partir da entrega das atividades. A não conclusão das mesmas no prazo é considerada falta.

FORMAS DE AVALIAÇÃO E CONTROLE DE FREQUÊNCIA

A aprovação na disciplina ocorrerá conforme a Resolução nº 37/97-CEPE. Ela dependerá do resultado das avaliações realizadas ao longo do período letivo (atividades), segundo o plano de ensino e cronograma divulgado no início do semestre, sendo o resultado global expresso de zero a cem. Será aprovado por média a e o discente que alcançar, no total do período letivo, frequência mínima de 75% da carga horária inerente à disciplina e obtiver, no mínimo, grau numérico 70 de média aritmética no conjunto de provas e outras tarefas propostas. Discente que não obtiver a média prevista deverá prestar exame final, desde que alcance a frequência mínima exigida e média não inferior a 40. No exame final a aprovação na disciplina dependerá da obtenção de grau numérico igual ou superior a 50 na média aritmética entre o grau do exame final e a média do conjunto das avaliações realizadas.

Os critérios para avaliação da aprendizagem envolverão a participação ativa nas atividades propostas; a posição crítica em relação às leituras e debates; o envolvimento no processo de ensino-aprendizagem; a evolução do aprendizado ao longo da disciplina; e o desenvolvimento de habilidades e competências em relação ao tema abordado.

As etapas avaliativas poderão ser assíncronas e/ou presenciais. Os critérios das atividades e avaliações estão definidos no Guia Didático da disciplina e serão publicados no ambiente virtual. As duas notas principais para composição da média final são:

- primeira nota: média aritmética das atividades (AT) realizadas individualmente ou em equipe (conforme solicitado);
- segunda nota: média aritmética das notas das avaliações (AV) individuais;

A nota final da disciplina será composta resumidamente conforme a seguinte equação:

$$\text{Nota} = \frac{\frac{\sum_{i=1}^x AT_i}{x} + \frac{\sum_{i=1}^n AV_n}{n}}{2}$$

Onde i é um contador incremental que começa valendo 1, x e n são a quantidade de atividades e avaliações (respectivamente) da disciplina. AT é sigla para Atividade e AV é sigla para Avaliação.

De maneira geral as atividades e avaliações irão considerar os seguintes aspectos:

- profundidade na abordagem do tema;
- qualidade do levantamento bibliográfico;
- contextualização do tema na área de interesse;
- explicação do funcionamento do algoritmo em questão;
- exemplificações práticas e contextualizadas do tema;
- análises dos resultados obtidos;
- criatividade da apresentação;
- uso de normas para elaboração de trabalhos de pesquisa (normas ABNT);
- desempenho individual e coletivo dos membros da equipe durante a apresentação.

As frequências serão controladas por chamada em sala de aula e/ou por entrega das atividades (no caso dos encontros remotos). O limite de ausências permitido é 25% da carga horária da disciplina.

Para mais detalhes sobre as datas previstas para as avaliações favor consultar o cronograma dentro do Guia Didático. Eventuais necessidades de ajustes nas datas, atividades e afins serão combinadas em acordo com a turma.

Exame final da disciplina: 28/02/2023.

BIBLIOGRAFIA BÁSICA (mínimo 03 títulos)

Castro, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Editora Saraiva, 2016. 978-85-472-0100-5. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/978-85-472-0100-5/>. Acesso em 17/08/2022.

NG, Andrew. **AI for Everyone**. Disponível em <https://www.deeplearning.ai/machine-learning-yearning/>. Acesso em 17/08/2022.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados - com aplicações em R**. Rio de Janeiro: Grupo GEN, 2016. 9788595155473. Disponível em <https://integrada.minhabiblioteca.com.br/#/books/9788595155473>. Acesso em 17/08/2022.

BIBLIOGRAFIA COMPLEMENTAR (mínimo 05 títulos)

BURKOV, Andriy. **The hundred-page machine learning book**. Disponível em: <http://themlbook.com/wiki/doku.php>. Acesso em 17/08/2022.

DEISENROTH, Marc Peter; FAISAL, A. Aldo e ONG, Cheng Soon. **Mathematics for machine learning**. Disponível em: <https://mml-book.github.io>. Acesso em: 17/08/2022.

GOOGLE. **Machine learning crash course**. Disponível em: <https://developers.google.com/machine-learning/crash-course/>. Acesso em 17/08/2022.

POOLE, David L.; MACKWORTH, Alan K. **Artificial intelligence: foundations of computational agents**. 2ª ed. Disponível em: <https://artint.info/2e/html/ArtInt2e.html>. Acesso em 17/08/2022.

SHALEV-SHWARTZ, Shai; BEN-DAVID, Shai. **Understanding machine learning: From Theory to Algorithms**. Disponível em: <https://www.cse.huji.ac.il/~shais/UnderstandingMachineLearning/> Acesso em 17/08/2022.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement learning: an introduction**. 2ª ed. Disponível em: <http://incompleteideas.net/book/RLbook2018.pdf>. Acesso em 17/08/2022.

Vídeos

RAMESH, Raj. **What is artificial intelligence?** In 5 minutes. Disponível em <https://www.youtube.com/watch?v=2eP9rue1Ao&t=155s> . Acesso em 17/08/2022.

Ventury city. **Amazon's city of the future**. Disponível em <https://www.youtube.com/watch?v=ZjraUKJRIt8>. Acesso em 17/08/2022.

Demais materiais serão disponibilizados no UFPR Virtual conforme a necessidade



Documento assinado eletronicamente por **JOSE MARCELO ALMEIDA PRADO CESTARI, PROFESSOR ENS BASICO TECNOLÓGICO**, em 19/09/2022, às 14:09, conforme art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **RODRIGO EDUARDO BOTELHO FRANCISCO, CHEFE DO DEPARTAMENTO DE CIENCIA E GESTAO DA INFORMACAO - SA**, em 21/09/2022, às 15:39, conforme art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida [aqui](#) informando o código verificador **4795567** e o código CRC **E80DA342**.