

INSTITUTO DE INFORMÁTICA
Departamento de Informática Aplicada

Dados de identificação

Disciplina: TÓPICOS ESPECIAIS EM COMPUTAÇÃO VII			
Período Letivo: 2021/01			
Professor Responsável: Karin Becker			
Sigla: INF01054	Créditos: 4		
Carga Horária: 60h	CH Autônoma:	CH Coletiva:	CH Individual:

Súmula

Discute conceitos, técnicas e tendências relacionados a projetos de Ciência de Dados. Aborda através um projeto de ciência de dados todo o processo de descoberta de conhecimento em grandes volumes de dados. As etapas envolvem a compreensão dos objetivos do negócio e dos dados, coleta e preparação dos dados, aplicação de técnicas analíticas e de mineração, e a avaliação dos resultados. Conhecimentos de Banco de Dados e Aprendizado de Máquina são esperados, e, portanto, recomenda-se que os alunos tenham cursado com aprovação em Fundamentos de Banco de dados (INF01145) e Aprendizado de Máquina (INF01017).

Currículos

Currículos	Etapas	Pré-Requisitos	Natureza
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO	Eletiva	94 créditos obrigatórios	Eletiva

Objetivos

Título: Projetos de Ciência de Dados

Através do desenvolvimento de um projeto de ciência de dados, espera-se que os alunos sejam capazes de:

- a) compreender que projetos de ciência de dados envolvem muitas fases, desde a compreensão do negócio e dos dados, até o desenvolvimento e valiação de modelos;
- b) compreender e aplicar técnicas para as principais classes de problema de mineração de dados;
- c) compreender e aplicar o processo de coleta, avaliação de qualidade, e preparação de grandes volumes de dados;

Por se tratar se disciplina centrada no desenvolvimento de projetos através de mentoria, a disciplina está restrita a um número máximo de 30 alunos.

Conteúdo Programático

Semana	Título	Conteúdo
1	Conceitos Básicos.	Objetivos. Decisões orientadas a dados. Etapas do processo de descoberta de conhecimento a partir de dados. Problemas de mineração e recursos analíticos.
2 - 3	Processo de Descoberta de Conhecimento.	CRISP-DM: compreensão do negócio; compreensão dos Dados; preparação dos Dados; modelagem; avaliação. Discussão de projetos reais.

4 - 5	Compreensão e Preparação de Dados. Seleção do tema do projeto.	Extração de Dados. Características dos Dados, Profiling e Exploração de Dados. Técnicas de avaliação de qualidade, limpeza, integração e transformação.
6	Associação	Definição da técnica. Algoritmo Apriori e Variações, Métricas de Avaliação. Estudo prático.
7	Classificação	Definição da técnica. Algoritmos de classificação conhecidos. Treino, avaliação e teste. Métricas de Avaliação. Estudo Prático.
8	Agrupamento	Definição da Técnica. Similaridade. Agrupamento particional, hierárquico e por densidade. Avaliação. Estudo Prático
9 - 16	Desenvolvimento de projetos de ciência de dados: Apresentação da proposta de trabalho; Apresentação de resultados intermediários; Apresentação de resultados finais.	Desenvolvimento de trabalhos de ciência de dados aplicando as técnicas aprendidas, e de acordo com todas as fases do processo. Apresentação da proposta de Trabalho; Apresentação de Resultados preliminares; Apresentação de resultados finais.

Metodologia

As atividades ocorrerão de forma síncrona ou assíncrona, conforme conteúdo.

As atividades síncronas ocorrerão nos horários regulares da disciplina, em datas especificadas no cronograma. Esses encontros serão gravados e disponibilizados aos alunos. As atividades assíncronas consistem na realização das tarefas propostas pelo professor, através do Moodle da disciplina. Essas atividades deverão ser realizadas até a data prevista. Estão previstos questionários e estudos dirigidos para fixação de conceitos e técnicas. Estudos dirigidos envolvem ferramentas computacionais, e devem ser realizados pelos alunos em seus computadores pessoais. Os alunos deverão desenvolver em grupo um projeto de ciência de dados, apresentando e discutindo os resultados em etapas. Haverá mentoria para desenvolvimento do projeto.

Todas as atividades serão propostas, entregues e avaliadas no Moodle da disciplina, onde constará as instruções a serem seguidas para sua realização.

Além das mentorias programadas, os alunos poderão contar com atendimento individualizado do professor, em horário a ser combinado e realizado de forma remota.

Recursos disponibilizados

As atividades previstas assim como as instruções para sua realização serão disponibilizadas no Moodle do INF. Eventuais componentes externos ao Moodle e necessários para a realização das atividades estarão indicados no próprio Moodle.

Recursos computacionais

Para acompanhar as atividades previstas é necessário ter acesso regular à Internet. As atividades síncronas podem ser acompanhadas através de telefone ou computador, com microfone e câmera. As aulas ocorrerão no MS Teams. Os trabalhos de ciência de dados são desenvolvidos com softwares de domínio público, e à escolha do aluno, exigindo computador.

Informações sobre Direitos Autorais e de Imagem:

Todos os materiais disponibilizados são exclusivamente para fins didáticos, sendo vedada a sua utilização para qualquer outra finalidade, sob as penas legais.

Todos os materiais de terceiros que venham a ser utilizados devem ser referenciados, indicando a autoria, sob pena de plágio.

A liberdade de escolha de exposição da imagem e da voz não isenta o aluno de realizar as atividades originalmente propostas ou alternativas;

Todas as gravações de atividades síncronas devem ser previamente informadas por parte dos professores.

Somente poderão ser gravadas pelos alunos as atividades síncronas propostas mediante concordância prévia dos professores e colegas, sob as penas legais.

É proibido disponibilizar, por quaisquer meios digitais ou físicos, os dados, a imagem e a voz de colegas e do professor, sem autorização específica para a finalidade pretendida.

Os materiais disponibilizados no ambiente virtual possuem licença de uso e distribuição específica, sendo vedada a distribuição do material cuja a licença não permita ou sem a autorização prévia dos professores para o material de sua autoria.

Carga Horária

Teórica: 60 horas

Prática: 0 horas

Experiências de Aprendizagem

Aulas expositivas síncronas/assíncronas. Estudos dirigidos usando pacotes de software e ferramentas de domínio público. Desenvolvimento de projetos de ciência de dados com mentoria, apresentação e discussão de resultados.

Critérios de Avaliação

Para ser aprovado é necessário obter média final (MF) igual ou superior a 6.0.

A Avaliação MF é composta por

(1) Atividades Autônomas (30% da nota final): tarefas, exercícios, questionários desenvolvidos para fins de fixação de conteúdo.

(2) Desenvolvimento do Trabalho Prático (70% da nota final).

A conversão da média numérica para conceito será feita conforme o seguinte critério:

MF \geq 9,0 : Conceito A

9,0 > MF \geq 7,5 : Conceito B

7,5 > MF \geq 6,0 : Conceito C

MF < 6,0: ver Atividades de Recuperação Previstas

De acordo com a Resolução do CEPE sobre o ERE, durante o período em que perdurar o ERE, fica inaplicável a atribuição de conceito FF, prevista no §2º, do Art. 44, da Resolução nº 11/2013 do CEPE.

Para os estudantes matriculados até o final do período e que deixaram de participar da Atividade de Ensino, deverá ser atribuído o registro NI (Não Informado) no campo de conceito do sistema acadêmico.

Para os casos previstos no §1º, a justificativa do registro NI deverá conter a referência ao período de excepcionalidade.

Os casos de não informação de conceito durante o ERE, deverão ser resolvidos até o fim do segundo período letivo, após o fim da situação emergencial de saúde.

Atividades de Recuperação Previstas

O aluno que obtiver conceito final D pode realizar recuperação, a qual consiste no desenvolvimento de um artigo relatando pesquisa sobre temas definidos. Se a nota do artigo for igual ou superior a 7,0, o conceito mudará para C.

Bibliografia

A Bibliografia Básica Essencial deve estar disponível de forma digital.

Básica Essencial

1. PROVOST & FOSTER. Data science para negócios. Alta Books, 2016.

2. TAN, O.N.; STEINBACH, M. KUMAR, V. "Introduction to Data Mining". Pearson Education Inc., 2006.

3. SILVA, PERES, BOSCAROLI. Introdução a mineração de dados – com aplicações em R. Elsevier, 2016.

4. Zaki, M.; Meira, Jr., W. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, 2014.

5. Artigos selecionados sobre projetos de ciência de dados.

Outras Referências

Observações