

DATA MINING COM PYTHON - CALENDÁRIO DE AULAS 2020.1

Prof. Victor Machado da Silva, M.Sc.

victor.silva@ibmec.edu.br

Sextas, 18:30 às 22:50

Objetivos do curso:

Possuir uma visão geral sobre como aplicar técnicas de mineração de dados para obter informações valiosas para a tomada de decisões em negócios. O uso da linguagem Python para o desenvolvimento de ferramentas de ciência de dados contribui para o ensino de uma disciplina voltada para demandas atuais do mercado.

Ementa do curso:

Conceitos de Python: Configuração de ambiente, tipos de dados, operadores, funções, estruturas de fluxo e de decisão, classes, pacotes de análise de dados (Numpy, Pandas, Matplotlib). Obtendo, limpando e transformando dados. Técnicas de mineração de dados aplicadas: Regressão Linear e Múltipla, Árvores de Decisão, Agrupamento, Processamento de Linguagem Natural.

Previsão de programação

Data	Tópico
14/02 sex	Introdução à disciplina
21/02 sex	Python: Configurando o ambiente e iniciando as primeiras aplicações
28/02 sex	Python: Tipos de dados, operadores e funções
06/03 sex	Python: Estruturas de fluxo e de decisão
13/03 sex	Python: Classes (1)
20/03 sex	Python: Classes (2)
27/03 sex	Python: Pacotes de análises de dados (1)
03/04 sex	Python: Pacotes de análises de dados (2)
10/04 sex	SEM AULA (Sexta-feira Santa)
17/04 sex	P1
24/04 sex	Construindo um ETL (1)
01/05 sex	SEM AULA (Dia do Trabalhador)
08/05 sex	Construindo um ETL (2)
15/05 sex	Construindo um ETL (3)
22/05 sex	Técnicas de Mineração de Dados: Regressão Linear e Múltipla
29/05 sex	Técnicas de Mineração de Dados: Árvores de Decisão
05/06 sex	Técnicas de Mineração de Dados: Agrupamento
12/06 sex	SEM AULA (Corpus Christi)
19/06 sex	Técnicas de Mineração de Dados: Processamento de Linguagem Natural
26/06 sex	P2
03/07 sex	Lançamento das notas de P2
10/07 sex	PS
17/07 sex	Prazo para lançamento de notas finais

Metodologia: aulas expositivas com apresentação de exemplos e desenvolvimento de casos práticos. Utilização de vídeos sobre temas relevantes à disciplina. Utilização de aplicações de software para instrumentalizar conceitos.

Avaliação: dois trabalhos de longo prazo realizados em grupos de até quatro alunos desenvolvidos durante o curso, que comporão as notas de P1 e P2. Os trabalhos possuirão entregas parciais, cujos pesos serão definidos posteriormente. As entregas dos trabalhos deverão ser realizadas através do Integrees. Não serão aceitos envios alternativos ou fora do prazo. Caso o aluno não obtenha aproveitamento, um novo trabalho (individual) será incluído como forma de compor a nota de PS. Além disso, serão avaliados exercícios feitos em sala, individuais ou em dupla, para serem entregues na mesma aula ou até a véspera da aula seguinte. A média de trabalhos será composta pela média aritmética simples desses exercícios.

Presença: será tirada falta ao final de cada aula, sendo aplicadas as regras de frequência do IBMEC/RJ. Alunos que não atendam aos critérios mínimos serão reprovados por falta, independente dos resultados das provas ou do trabalho desenvolvido.

Bibliografia:

- Slides de aula e listas de exercício, disponibilizados via Integrees
- Livro texto: Grus, J. **Data Science do Zero**, Alta Books, 2016.
- Livro texto: Wazlawick, R. **Introdução a Algoritmos e Programação com Python: Uma Abordagem Dirigida por Testes**, Elsevier, 2018.
- Livro texto: Silva, L. A. da, Peres, S. M., Boscarioli, C. **Introdução à Mineração de Dados com Aplicações em R**, Elsevier, 2016.