



Aplicação de Data Science

Tecnologias de Informação - Data & Analytics

- **Nível:** Avançado
 - **Duração:** 18h
-

Sobre o curso

O aluno vai trabalhar com ferramentas de data science, nas quais vai desenvolver desafios que envolvem técnicas de manipulação e tratamento de dados já conhecidas pela comunidade de data scientists.

Este curso vai permitir ao aluno desenvolver, em teoria e na prática, a sua capacidade analítica. Será dada ênfase à importância de permitir que o trabalho realizado seja replicado por outros, o que representa um problema atual, bem como trabalhar na qualidade dos dados e identificar quais as métricas a escolher para avaliar e suportar a solução escolhida.

Objetivos:

Desenvolver conhecimento e capacidade de comunicação dos métodos a diferentes audiências. Em complemento, pretende-se desenvolver problemas de data science. Com base nestes dois objetivos mais abrangentes serão abordados os seguintes subtemas:

Saber:

- Dados que não são perfeitos: valores em falta e duplicados, ruído, outliers, dimensionalidade, bias;
- Escalabilidade: computação distribuída, big data;
- Análise de dados: modelação preditiva, analisar e interpretar resultados;
- Reproducibility, causalidade versus correlação, testes A/B e manutenção de modelos;

Fazer:

- Análise de dados;
- Implementar modelos preditivos, analisar e interpretar resultados;
- Implementar testes A/B;

Soft-Skills:

- Consciencializar para a importância da qualidade de dados;

- Desenvolver capacidades analíticas em análise de dados;
 - Trabalhar em equipa com o desenvolvimento de trabalho prático e escrita de relatório cumprindo prazos definidos;
-

Destinatários

Todos os que pretendam tirar partido dos benefícios da Ciência de Dados no suporte às decisões estratégicas de negócio.

Pré-requisitos

- Conhecimentos da língua inglesa
 - Não sendo um pré-requisito, conhecimentos em Matemática, Estatística e Algoritmia ajudam a compreender melhor os conteúdos.
-

Metodologia

- Aulas expositivas devidamente complementadas com trabalhos práticos.
 - Existirão casos de estudo com debate entre os alunos e orientados pela docente. Estes casos consistem na apresentação da situação de uma organização (real ou ficcionada) onde são descritos os desafios que se lhe colocam. Os alunos deverão propor uma solução para o problema de acordo com a matéria exposta e o seu estudo.
-

Programa

- Comparação do cenário ideal versus o cenário real
- Qualidade dos dados fonte
- Fatores que afetam os resultados
- Inferência estatística versus predição
- Dimensão dos dados
- Interpretação de resultados
- Escalabilidade
- Reprodutibilidade
- Causalidade versus Correlação

- A/B testing
- Manutenção dos modelos