



Metodologias de Pré-Processamento de Dados

Tecnologias de Informação - Data & Analytics

Promoção: **Curso com 25% de desconto** Aproveite já!

Live Training (também disponível em presencial)

- **Localidade:** Imprimir Curso
 - **Data:** 19 Jul 2024
 - **Preço:** 1070 € (Os valores apresentados não incluem IVA. Oferta de IVA a particulares e estudantes.)
 - **Horário:** Pós-laboral e Sábados das 6ª feiras 18h45 - 21h45 + sábados 9h30 - 17h00
 - **Nível:**
 - **Duração:** 18h
-

Sobre o curso

Objetivos

- Perceber quais são os tipos de dados.
 - Adquirir as competências em Python para limpar tabelas de dados sujeitas a erros de diversas origens.
 - Perceber quais são as transformações necessárias para preparar os dados para o treino de um algoritmo de classificação supervisionada.
 - Aplicação das técnicas de limpeza e transformação de dados usando as funções dos pacotes Pandas e Scikit learn.
-

Pré-requisitos

Conhecimentos básicos de programação e estruturas de dados. Alguma experiência com Python.

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

- Terminologia
- Tipos de dados
- Características dos dados
- Pré-processamento de dados
- Limpeza de dados
- Integração/enriquecimento dos dados
- Problemas Associados à Integração dos dados
- Transformação dos dados

- Terminologia
 - Dataset... ? O que significa?
 - Variável
 - Variável vs feature
- Tipos de dados
 - Dados Numéricos
 - Dados Discretos vs Contínuos
 - Dados Categóricos
 - Dados Binários vs Nominais vs Ordinais
- Características dos dados
 - Dispersão
 - Monotonicidade
 - Anacronismo
- Pré-processamento de dados

- Paradigma GICO
- Ciclo do Pré-processamento de dados
- Limpeza de dados
 - Natureza dos valores em falta
 - Estratégia para lidar com valores em falta
 - Dados com ruído
- Integração/enriquecimento dos dados
- Problemas Associados à Integração dos dados
- Transformação dos dados
 - Agregações
 - Transformação de variáveis categóricas
 - Discretização
 - Escalamento das variáveis
 - Engenharia/Derivação de atributos