

Metodologias de pré-processamento de dados

- **Formato do curso:** Presencial e Live Training
- **Preço:** 900€
- **Duração:** 18 horas

Pré-requisitos

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

Objectivos

- 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.

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
 - Dataset... ? O que significa?
 - Variável
 - Variável vs feature

- Tipos de dados
 - Dados Numéricos
 - Dados Discretos vs Contínuos
 - Dados Categóricos
 - Dados Binários vs Nominiais 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