



Metodologias de Pré-Processamento de Dados

Tecnologias de Informação - Data & Analytics

- **Nível:**
 - **Duração:** 18h
-

Sobre o curso

Aprenda técnicas de pré-processamento de dados essenciais para garantir a qualidade e integridade dos seus dados, cobrindo desde a limpeza e integração até a transformação e engenharia de atributos.

Este curso proporciona uma compreensão aprofundada das metodologias de pré-processamento de dados, essenciais para a análise de dados e machine learning. Começando com a terminologia e tipos de dados, os participantes irão explorar características dos dados como dispersão e monotonicidade.

O curso aborda o ciclo do pré-processamento, incluindo limpeza de dados, gestão de valores em falta, integração e enriquecimento de dados, e transformação de variáveis. Técnicas avançadas como agregações, discretização, escalamento e engenharia de atributos serão também cobertas, preparando os participantes para lidar eficazmente com dados em ambientes reais.

Destinatários

Este curso é ideal para analistas de dados, cientistas de dados, e qualquer profissional interessado em aprender técnicas de pré-processamento de dados para análises e modelos de machine learning.

Objetivos

- **Compreender Terminologia e Tipos de Dados:** Familiarizar-se com conceitos essenciais como datasets, variáveis e tipos de dados numéricos e categóricos.

- **Aplicar Técnicas de Limpeza e Integração de Dados:** Desenvolver competências para lidar com valores em falta, dados com ruído e problemas de integração de dados.
 - **Transformar e Enriquecer Dados:** Utilizar técnicas de agregação, transformação de variáveis, escalamento e engenharia de atributos para melhorar a qualidade dos dados.
-

Pré-requisitos

Os participantes devem ter conhecimentos básicos de estatística e programação, preferencialmente em Python. Familiaridade com conceitos de bases de dados será vantajosa.

Metodologia

Sessões expositivas devidamente complementadas com trabalhos práticos. Existirão casos de estudo com debate entre os alunos e orientados pelo 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 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