



**TÉCNICO+**  
FORMAÇÃO AVANÇADA

Programa de Especialização

# **Data Science para Engenheiros (DaSh)**

## **Aplicações de Inteligência Artificial e Machine Learning a Grandes Quantidades de Dados**

**Data Science**, aplicação paradigmática das técnicas mais promissoras da Inteligência Artificial e *Machine Learning*, destaca-se como a tecnologia chave para a transformação digital, tendo a própria profissão de *Data Scientist* sido considerada pela *Harvard Business Review* (2012) como *The Sexiest Job of the 21st Century*. Mesmo neste novo contexto de pandemia, o *Data Science* mantém e reforça o seu papel vital nas organizações, ao transformar dados em informação. É esta transformação que permite a descoberta de **modelos capazes de prever e antecipar comportamentos, tanto de clientes como de competidores, e que dá às organizações o suporte necessário à tomada informada de decisões**, de modo a melhor preparar a sua atividade presente e futura.

Esta pós-graduação em *data science* **destina-se a todos os engenheiros que entendam os fundamentos básicos de linguagens de programação, como o python, permitindo-lhes dar os primeiros passos no vasto mundo do data science**. Através de três módulos fundamentais e um módulo avançado em *deep learning* e acompanhado de uma forte componente prática, o curso promove uma aprendizagem eficaz e capacitadora para entrar numa nova profissão.

### OBJETIVOS

- Identificar visualmente perfis de dados;
- Conhecer e aplicar as técnicas mais recentes de classificação, *clustering*, descoberta de padrões e deteção de anomalias, em particular de *deep learning*;
- Escolher e aplicar as abordagens mais adequadas de preparação de dados;
- Avaliar a qualidade dos modelos descobertos.

### DESTINATÁRIOS

Licenciados ou Mestrados em Eng. Informática, Eng. Eletrotécnica, Matemática Aplicada e Computação, ou com experiência equivalente, sem formação prévia em Ciência de Dados.



Carga Horária Presencial  
de 140h



3 mai a 24 jul 2021 e  
11 set a 16 out 2021



Online



2ª e 4ª-feira, 18h30/21h30  
sábado, 09h00/13h00\*



5200€ (condições especiais  
para particulares)



10 ECTS

**tecnicomais.pt**

\*a partir de setembro, o horário das aulas é apenas aos sábados, das 09h00 às 12h00 e das 13h30 às 15h30

## COORDENAÇÃO



### **Cláudia Antunes** | IST

Professora Associada de Sistemas de Informação no Instituto Superior Técnico, tendo lecionado disciplinas, ao nível da graduação, pós-graduação e formação profissional, na área das ciências de dados, aprendizagem automática e programação. O seu doutoramento foi um dos primeiros em Portugal, na área da Ciência de Dados, área na qual tem publicados mais de setenta artigos em conferências e revistas da especialidade. Desde então, orientou dezenas de alunos no âmbito dos seus doutoramentos e mestrados, na mesma área.

## FORMADORES



### **Francisco Saraiva de Melo** | IST

Professor no Departamento de Engenharia Informática do Instituto Superior Técnico e investigador sénior no GAIPS, INESC-ID. A sua investigação abrange problemas de decisão inteligente e aprendizagem automática, particularmente aprendizagem por reforço, planeamento sob incerteza, sistemas multi-agente e multi-robot, e aplicações na robótica e interação humano-robot. Tem participado em vários projetos nacionais e europeus envolvendo aprendizagem automática, redes de sensores e interação humano-robot.



### **Rui Henriques** | IST

Professor Auxiliar no Departamento de Engenharia Informática e investigador associado no INESC-ID, onde se dedica à aprendizagem de modelos descritivos e preditivos a partir de dados estruturados. Neste contexto, os seus principais interesses são a descoberta avançada de padrões em espaços multi-dimensionais, e o enquadramento estatístico da aprendizagem para garantir a significância estatística e confiabilidade dos modelos. É licenciado e doutorado pelo Técnico, ex-McKinsey (2011) e coordenou em 2017 a área de Inteligência Artificial da TEKEVER.

## PLANO DE ESTUDOS

### **1. Módulo I - Introdução à Ciência de Dados e Análise Exploratória** (18h)

*Prof. Cláudia Antunes*

Este módulo apresenta os conceitos básicos da área, e descreve os processos e princípios fundamentais da análise de dados e descoberta de informação. São ainda desmistificadas as relações entre Inteligência Artificial, Aprendizagem e Ciência de Dados, e descritas as diferentes perspetivas de análise exploratória de dados (data profiling). Nomeadamente, são abordadas a dimensionalidade, distribuição, dispersão e granularidade dos dados, dando particular importância à exploração dos diferentes elementos gráficos para representação e sumariação dos mesmos. Serão usados os pacotes pandas e matplotlib.

### **2. Módulo II - Classificação** (60h)

*Prof. Cláudia Antunes*

Este módulo conjuga as várias técnicas de treino de modelos de classificação sobre dados tabulares, com as diferentes técnicas de preparação e transformação de dados mais adequadas. Para além da apresentação dos cinco paradigmas de classificação (Baesianos, analogistas, simbolistas, conexionistas e evolucionistas), são descritas as abordagens mais atuais de combinação de modelos (*ensembles*) e discutidas as medidas e estratégias de avaliação dos modelos encontrados. Em termos de preparação de dados, são aplicadas as operações de escala (*normalization* e *standardization*), imputação de *missing values*, a dumificação de variáveis, a seleção, extração e reengenharia de variáveis (*feature selection, feature extraction and feature engineering*), assim como as técnicas de balanceamento de dados. Será usado o pacote pandas, matplotlib e sklearn.

### Módulo III - Análise Não-Supervisionada (30h)

Prof. Rui Henriques

Este módulo aborda as técnicas fundamentais de *clustering*, descoberta de padrões e deteção de anomalias, acompanhada pela discussão das medidas de avaliação mais adequadas a cada caso. Serão usados os pacotes *pandas*, *matplotlib*, *sklearn* *apyori* e *prefixspan*.

### Módulo IV - Deep Learning (30h)

Prof. Francisco Saraiva de Melo

Este módulo introduz a aprendizagem baseada em redes neuronais profundas, abordando as questões da regularização, treino e otimização das redes. Introduzindo ainda os dois modelos principais das redes profundas – as redes convulsionais (CNNs) e as redes recorrentes (RNNs), discutindo as diferentes aplicações de sucesso das mesmas, assim como a explicabilidade daquelas redes.

## PRÉ-REQUISITOS

Os pré-requisitos para a realização deste curso são os seguintes: fluência na escrita de programas básicos em Python e noções de programação orientada a objetos; noções básicas ao nível da utilização de bases de dados; funções de distribuição e densidade de probabilidades, estimadores elementares e testes de significância; e matrizes, operações sobre matrizes, determinantes e valores próprios.

## ESTATÍSTICAS



The global **market size of Big Data and Business Analytics was estimated at USD 171.39 billion** in 2018 and is projected to reach USD 512.04 billion by 2026, rising at a CAGR of 14.80 percent between 2019 and 2026.

PR Newswire, 2020

## AVALIAÇÃO

A avaliação consiste em projetos nos módulos I, II e III e em laboratórios de *Deep Learning*. Os projectos consistirão na exploração de dois conjuntos de dados, tendo como objetivo a descoberta da informação possível. Serão desenvolvidos em Python, com os pacotes de *data science* (*pandas*, *matplotlib*, *seaborn* e *sklearn*).

VAMOS COMEÇAR UM PROJECTO JUNTOS?

**Fale connosco!**



**Técnico+**

+351 218 417 010 / [Info@tecnicomais.pt](mailto:Info@tecnicomais.pt)