Curso: Pós-Graduação em Data Science

Duração: 162h
Área formativa: Cursos

Sobre o curso

A quantidade de dados gerados pelas organizações tem vindo a crescer exponencialmente, aumentando a necessidade de profissionais especializados em extrair valor desses dados. Quando devidamente explorados, os dados permitem suportar decisões estratégicas, criar produtos e oferecer serviços mais personalizados, entre muitas outras potencialidades. Data Science, portanto, é uma área essencial e transversal a todos os tipos de organizações e setores, reconhecida pelo seu potencial em proporcionar vantagens competitivas.

Esta Pós-Graduação é desenvolvida ao abrigo do protocolo de colaboração entre a **Atlântica** e a **Rumos**. Destina-se a todos aqueles que pretendam tirar partido desta nova capacidade estratégica, dotando-os dos conhecimentos necessários para retirar o máximo valor dos dados, dando uma visão detalhada, teórica e prática, de conceitos e metodologias.

Este programa dota os participantes dos conhecimentos necessários para maximizar o valor dos dados, oferecendo uma visão detalhada, teórica e prática, de conceitos e metodologias. Os alunos que concluam a formação com sucesso estarão capacitados para criar e gerir equipas de Data Science, estruturar projetos, identificar oportunidades, extrair, processar e explorar dados utilizando métodos estatísticos, e desenvolver modelos descritivos e preditivos. Adquirirão conhecimentos amplos, teóricos e práticos, para exercer uma carreira altamente procurada e valorizada no mercado nacional e internacional.

Os alunos que concluam a Pós-graduação com sucesso, ficam habilitados a proceder à criação e gestão de equipas de Data Science na organização, a estruturar projetos de Data Science, serem capazes de identificar oportunidades para uso de Data Science dentro das organizações, saberem extrair, processar e explorar dados aplicando métodos estatísticos para retirar mais informação dos dados e serem capazes de criar modelos descritivos e preditivos. Os alunos vão adquirir nesta formação amplos conhecimentos, teóricos e práticos, para exercer uma carreira que é altamente procurada e valorizada no mercado nacional e internacional.

Diploma de Estudos

Cada unidade curricular tem uma avaliação dos conhecimentos adquiridos que permite a atribuição de um Diploma de Estudos no final do curso a todos os alunos que o terminem com sucesso. A avaliação de conhecimentos é individual, sendo obtida através da participação em trabalhos de grupo e testes. Os alunos que não conseguirem realizar os testes e trabalhos de determinada unidade curricular nas datas estipuladas, impossibilitando assim a sua avaliação, poderão solicitar a realização de exame de recuperação mediante um custo adicional. Estes exames serão realizados na época especial de recuperação, a decorrer no final da Pós-graduação.

Atribuição de ECTS

No âmbito da parceria com a **Atlântica,** Escola Universitária de Ciências Empresariais, Saúde, Tecnologias e Engenharia, esta Pós-graduação atribui **25 créditos ECTS** (European Credit Transfer System).

Knowledge Extension Pack

Pack opcional de apoio à Pós-Graduação: +Conteúdo, +Conhecimento, +Apoio

Para melhorar a experiência formativa dos alunos, e mediante um valor adicional, a Rumos disponibiliza um conjunto de materiais adicionais concebidos com os seguintes objetivos:

:: Complementar o conteúdo transmitido em sala de aula.

:: Reforçar conhecimentos base, ajudando a interiorizar o que é ensinado pelos Docentes.

:: Abordar tópicos complementares que serão uma mais-valia para a adaptação dos Data Scientists a novas realidades.

Estes momentos de autoestudo estão estrategicamente distribuídos ao longo do percurso, permitindo que os alunos avancem ao seu próprio ritmo e explorem os conteúdos e recursos de forma independente, promovendo uma maior autonomia e responsabilidade no seu percurso formativo.

As áreas de estudo complementares incluídas:

:: Introdução à Data Science

:: Introdução ao Python

:: Conceitos Fundamentais de Estatística

:: Fundamentos de T-SQL

:: Ética em Data Science

:: Visualização de Dados e Storytelling

:: Microsoft Azure Fundamentals

:: Ferramentas de Data Science

Este pack complementar inclui o **acesso a 9 horas adicionais** (1 hora por unidade curricular) de contacto com os Docentes, permitindo-lhes partilhar feedback sobre os trabalhos apresentados e esclarecer dúvidas que surjam na elaboração desses trabalhos. Estas horas adicionais, não estão definidas previamente em cronograma e serão agendadas no início de cada unidade, de acordo com a disponibilidade de todos os envolvidos.

Valor adicional: 540€ Caso desejes incluir este pack opcional na sua inscrição, este valor será acrescido ao valor total da Pós-graduação.

Condições Financeiras

:: Taxa de matrícula: 200€, dedutível no valor total.

:: Knowledge Extension Pack: 540€ (Opcional)

:: Possibilidade de pagamento faseado para particulares, até 8 prestações, sem juros.

:: Formandos não residentes no território nacional, terão de efetuar um pagamento de 50% do valor total da propina no momento da inscrição.

:: Os valores apresentados não incluem IVA. Isenção do valor do IVA a particulares.

:: Condições especiais para Alumni de Academias ou Pós-graduações Rumos.

:: Para informações completas sobre os requisitos, descontos e condições financeiras disponíveis, contacta-nos.

> Ao concluíres com aproveitamento esta formação, receberás um badge digital para partilhares com a tua rede profissional online.
>
> Sabe mais sobre os badges digitais aqui.

Objectivos

No final da Pós-graduação, os alunos deverão ser capazes de programar em Python, utilizando bibliotecas como Pandas, NumPy e Seaborn para manipulação e análise de dados. Serão também proficientes em conceitos de inferência estatística e A/B Testing, além de capazes de recolher e integrar dados a partir de ficheiros JSON, APIs, web scraping e bases de dados SQL. Aprenderão a pré-processar e limpar dados, realizar análise exploratória, e selecionar variáveis preditivas utilizando pipelines de Data Science.

Os alunos desenvolverão competências em Machine Learning, incluindo aprendizagem supervisionada e não supervisionada, e operacionalizar modelos utilizando ferramentas como Scikit-learn, MLFlow, Docker e FastAPI. Também adquirirão conhecimentos fundamentais de deep learning aplicados a processamento de linguagem natural e visão computacional, utilizando modelos pré-treinados e abordando questões de ética e justiça nestas áreas.

Metodologia

A Pós-graduação em Data Science é ministrada 100% em Live Training, através de sessões online síncronas em tempo real, e é constituída por 9 unidades formativas.

A metodologia pedagógica está focada no saber fazer, pelo que os conceitos e teoria de base são

fortemente explorados em treino orientado para a colocação dos conhecimentos em prática.

Os alunos devem ter em conta que para além da exigência das aulas, há a exigência de estudo adicional e de desenvolvimento de trabalhos práticos que permitam um cabal desenvolvimento do saber fazer nesta área de conhecimento.

Pré-requisitos

As candidaturas à **Pós-Graduação em Data Science** estão abertas a:

:: Todos os que tenham um grau académico de licenciatura ou superior nas áreas científicas;

:: Todos os profissionais com ou sem grau académico, cuja experiência seja considerada adequada para que o aluno tenha sucesso no curso e as turmas resultem homogéneas.

A seleção será sempre feita mediante análise curricular pela Coordenação Científica do curso, que pode chamar o candidato a uma entrevista presencial. Em qualquer dos casos, a decisão será sempre fundamentada e apresentada por escrito ao candidato. Adicionalmente, os alunos deverão ter:

:: Conhecimentos básicos de Matemática e Estatística;

:: Conhecimentos básicos de utilização de uma qualquer linguagem de programação;

:: Bom nível de capacidade de leitura em língua inglesa.

Destinatários

A **Pós-Graduação em Data Science** destina-se a todos aqueles que queiram adquirir conhecimentos que lhes permitam tirar partido desta nova capacidade estratégica, dotando-os dos conhecimentos necessários para conseguirem retirar o máximo valor dos dados e com isso aportarem valor competitivo para as organizações, nomeadamente profissionais que desempenhem funções diretas de tomada de decisão ou que intervenham em processos de análise de dados para a tomada de decisão sobre os negócios da organização.

Programa

Sessão protocolar de abertura – 2h

- Boas-vindas
- Regras, processos e procedimentos existentes na Pós-graduação

Programação em Python para Data Science – 18h

Nesta unidade curricular o aluno vai aprender a programar em Python com foco em Data Science, desde a instalação e configuração do ambiente, até ao uso de bibliotecas específicas.

- Instalar e configurar Python
- Introdução a Python
    - Tipos básicos de dados
    - Variáveis, expressões e operadores
    - Tipos avançados de dados
    - Controlo de fluxo
    - Ciclos
    - Funções
- Princípios de Design
    - Classes, OOP em python.
- Introdução a Pandas, numpy, seaborn.
    - Ler ficheiros
    - Tipos de dados
    - Slicing e indexamento em pandas
    - GroupBy e agregações em pandas
    - Gráficos em pandas

Inferência Estatística e A/B Testing – 18h

Nesta unidade curricular o aluno vai compreender os conceitos fundamentais da estatística inferencial e aplicar métodos de A/B testing para avaliar hipóteses e tomar decisões informadas.

- Conceitos básicos de Estatística Inferencial
- Distribuições Amostrais e Teorema do Limite Central
- Estimação e Intervalos de Confiança
- Testes de Hipóteses
- Introdução ao A/B Testing
- Métodos e Ferramentas para A/B Testing

Metodologias de Recolha de Dados – 18h

Nesta unidade curricular o aluno vai explorar diversas técnicas de recolha de dados, desde a utilização de APIs e web scraping até ao uso avançado de SQL.

- Como trabalhar com ficheiros JSON
- Introdução a APIs: Autenticação, pedidos HTTP e diferentes utilizações que estas podem ter
- Web scraping: utilização de bibliotecas de Python para executar web scrapers e recolha de informação não dependente de APIs
- Introdução a SQL
- Views em SQL
- Stored Procedures em SQL
- Polybase em SQL Server
- Subqueries em SQL
- Indices (clustered, non-clustered) em SQL
- Leitura e escrita de Ficheiros Parquet em Python

Metodologias de Pré-processamento de Dados – 18h

Nesta unidade curricular o aluno vai aprender a preparar os dados para análise, abordando técnicas de limpeza, transformação e integração de dados para garantir a qualidade e a usabilidade dos mesmos.

- 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

Análise Exploratória de Dados – 18h

Nesta unidade curricular o aluno vai realizar análises exploratórias para identificar padrões, tendências e relações nos dados, utilizando técnicas de visualização e redução dimensional.

- Visualização de dados para Análise Exploratória de Dados
- Sumário estatístico dos dados (Alterar o nome?)
- Seleção de variáveis preditivas
- Redução de dimensão/número de variáveis
- Pipelines de Data Science

Fundamentos de Machine Learning – 18h

Nesta unidade curricular o aluno vai aprender os conceitos básicos de Machine Learning, desde a predição e validação cruzada até aos modelos de regressão e classificação, utilizando o pacote Scikit-learn.

- Predição, erros e validação cruzada
- Aprendizagem Supervisionada e não supervisionada
- Modelos de Regressão
- Modelos de Classificação
- Pacote Scikit-learn

Operacionalização de Machine Learning – 18h

Nesta unidade curricular o aluno vai adquirir competências para operacionalizar modelos de Machine Learning, garantindo a reprodutibilidade e a gestão eficiente de projetos de ML em produção.

- Introdução à operacionalização de Machine Learning
  - Conceitos e importância da operacionalização
  - Visão geral das etapas do pipeline de ML
- Reprodutibilidade e Organização de código
  - Boas práticas de organização de código
  - Ferramentas e técnicas para assegurar reprodutibilidade
- Tracking de Experiências de Machine Learning (MLFlow)
  - Introdução ao MLFlow
  - Implementação de tracking de experiências
- Serviço de Machine Learning (Fastapi)
  - Desenvolvimento de APIs com Fastapi
  - Implementação de serviços de ML em produção
- Conteinerização (Docker)
  - Fundamentos de conteinerização
  - Criação e gestão de containers Docker
- Introdução ao Github
  - Conceitos básicos de Git e Github
  - Fluxo de trabalho colaborativo e controlo de versões
- CICD (Github Actions)
  - Configuração de pipelines de CI/CD
  - Automação de testes e deploys com Github Actions
- Deploy na Cloud e Operacionalização de ML na indústria
  - Introdução aos principais fornecedores de cloud
  - Estratégias de deploy e monitorização de serviços em produção

Fundamentos de Deep Learning para NLP e Computer Vision – 27h

Nesta unidade curricular o aluno vai explorar os fundamentos de Deep Learning aplicados ao processamento de linguagem natural (NLP) e visão computacional, utilizando técnicas e modelos avançados para tarefas específicas.

- Fundamentos Deep Learning
  - O que é Deep Learning
  - Logical Unit, Perceptron e Signmoid Neuron
  - Modelos pré-treinados
- Visão Computacional
  - O que é Visão Computacional
  - Filtering, Feature Detectors e Descriptors
  - Transform e Segmentation
  - Image Classification
  - Ethics and Fairness em Visão Computacional
- Processamento de Linguagem Natural
  - Dados não estruturados vs dados estruturados
  - Técnicas de processamento de linguagem como tokenization, stop words e named entity recognition
  - Modelos supervisionados e não-supervisionados em processamento de linguagem natural
  - Modelos Deep Learning e aplicação na área de processamento de linguagem natural

Implementação de Data Science – 9h

Nesta unidade curricular o aluno vai desenvolver um projeto prático de Data Science, aplicando os conhecimentos adquiridos ao longo da pós-graduação e apresentando os resultados.

- Apresentação do projeto a desenvolver e exposição de exemplos de projetos reais. Boas práticas e metodologias
- Sessão para esclarecimento de dúvidas
- Apresentação dos projetos

Sessão protocolar de encerramento Pós-Graduação – 1h